Structured Uncertainty and Model Misspecification*

Lars Peter Hansen† Thomas J.Sargent‡

April 19, 2019

Abstract

An ambiguity averse decision maker evaluates plans under a restricted family of *structured* models and *unstructured* alternatives that are statistically close to them. Structured models include parametric models in which parameter values vary over time in ways that the decision maker cannot describe probabilistically. Because he suspects that all parametric models are misspecified, the decision maker also evaluates plans under alternative probability distributions with much less structure.

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

---

*We thank Ben Brooks, Xiaohong Chen, Timothy Christensen, Yiran Fan, Itzhak Gilboa, Doron Ravid, Marcinano Siniscalchi, Bálint Szőke, and John Wilson for critical comments on earlier drafts. We acknowledge and thank the Alfred P. Sloan Foundation Grant G-2018-11113 for support.

†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. Irving J. Good (1952)

Now it would be very remarkable if any system existing in the real world could be exactly represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations. George Box (1979)

1 Introduction

To create a set of probability distributions for a cautious decision maker of a type described by Wald (1950) and axiomatized in different ways by Gilboa and Schmeidler (1989) and Maccheroni et al. (2006a,b), we start with mixtures of *structured models*.[1] Alternative mixture weights are possible Bayesian priors. We then add doubts about each structured model and represent a decision maker's aversion to uncertainty in a way that distinguishes ambiguity about a prior over structured models from suspicions that structured models are misspecified. Our decision maker responds to specification suspicions about structured models by evaluating plans under unstructured alternatives that approximate structured models well in terms of their statistical fits.

Thus, the decision maker constructs a set of probability models in two steps. First, she specifies a set of *structured* probability models that have either fixed or time-varying parameters. Second, she adds *unstructured* models that are statistically near a structured model. *Unstructured* models can be non parametric and are described incompletely in the sense that they are required only to reside within a statistical neighborhood of the set of structured models as measured by relative entropy. The decision maker thus acknowledges approximation concerns like those expressed by Box in the above quotation.[2]

[1]By "structured" we don't mean what econometricians in the Cowles commission and rational expectations traditions call "structural" models. We simply mean more or less tightly parameterized statistical models.

[2]Itzhak Gilboa suggested to us that there is a connection between our distinction between structured and unstructured models and the contrast that Gilboa and Schmeidler (2001) draw between rule-based and case-based reasoning. We find that possible connection intriguing but defer formalizing it to subsequent research. We suspect that our structured models could express Gilboa and Schmeidler's notion of rule-based reasoning, while our unstructured models resemble their case-based reasoning. But our approach here differs from theirs because we proceed by modifying an approach from robust control theory that seeks to acknowledge misspecifications of structured models while avoiding the flexible estimation methods that

We use a dynamic variational extension of max-min preferences that was conceived by Maccheroni et al. (2006a,b) to represent aversions to two components of uncertainty – uncertainty about a prior over structured models and uncertainty about specifications of the structured models themselves. Our use of dynamic variational preferences here substantially extends Hansen and Sargent (2001) and Anderson et al. (2003). Hansen and Sargent (2019) show that macro-finance applications of our new framework bring new sources of variations in resource allocations and valuations. In section 2, we employ a statistical decision theoretic concept called admissibility that allows us to implement the suggestion of Good cited above that is a prominent element of robust Bayesian analysis.

We use positive martingales to represent a decision maker's set of probability specifications. Working in continuous time and with Brownian motion information structures gives us a convenient way to represent positive martingales, as we describe in section 3. We use martingales twice: first, when we form a set of structured models having a recursive structure suggested by Chen and Epstein (2002); and second, when we add probabilities associated with unstructured models that are difficult to distinguish from structured models by applying statistical methods to limited data.

To represent and assess potential misspecifications, we use relative entropy measures of statistical neighborhoods to construct families of structured models and also to explore misspecifications of those structured models, as we describe in section 4.

While important aspects of our analysis apply in more general settings, to be concrete, we have added inessential auxiliary assumptions that we find helpful and enlightening. Extensions of the framework presented here relax the Brownian information structure and do not use relative entropy to constrain a decision maker's family of structured models.

2 Decision theory components

Our model of decision making under uncertainty strikes a balance among three attractive but potentially incompatible preference properties, namely, (i) dynamic consistency, (ii) a statistical decision-theoretic concept called *admissibility*, and (iii) concerns about model misspecifications. Since we are interested in intertemporal decision problems, we like preferences with recursive structures that exhibit dynamic consistency. In addition, to judge the plausibility of the modeling inputs, we want our decision maker to verify admissibility and

would be required to construct better statistical approximations that might be provided by unstructured models.

to inspect the implied worst-case probabilities. Within the confines of the max-min formulation of Gilboa and Schmeidler (1989), we describe situations in which a conflict between dynamic consistency and admissibility exists and also substantively interesting situations in which it does not. Within the max-min utility formulation of Gilboa and Schmeidler (1989), we describe interesting situations in which a decision maker's preferences cannot be dynamically consistent except in degenerate and special cases. Those situations motivate us to leave the Gilboa and Schmeidler (1989) formulation and instead to use a version of the variational preferences of Maccheroni et al. (2006a,b) that reconcile dynamic consistency with admissibility. The following paragraphs tell the logic that led us to adopt our version of variational preferences.

Let $\mathfrak{F} = \{\mathfrak{F}_t : t \geqslant 0\}$ be a filtration that describes information available at each $t \geqslant 0$. A decision maker evaluates *plans* or *decision processes* that are restricted to be progressively measurable with respect to $\mathfrak{F}$. Each structured model indexed by, say, $\theta \in \Theta$ assigns probabilities to $\mathfrak{F}$, as do mixtures of these models. We can interpret alternative mixing distributions as possible priors over structured models. An admissible decision rule is one that cannot be weakly dominated by another decision rule for all $\theta \in \Theta$ and that can be strictly dominated for some $\theta \in \Theta$.

A Bayesian decision maker completes her probability specification by choosing a unique prior over the set of structured models. Temporarily, suppose that for each possible probability specification over $\mathfrak{F}$ implied by a prior over the set of structured models, the decision problem has a recursive structure with the following properties: (i) a plan solves the time 0 decision problem; (ii) for each $t > 0$, the time t continuation of the plan for the time 0 problem is the unique solution of a time t continuation problem. A plan with properties (i) and (ii) is said to be dynamically consistent. The Bayesian decision maker's (dynamically consistent) optimal plan typically depends on the choice prior.

A "robust Bayesian" evaluates plans under a nontrivial set of priors. A popular representation of ambiguity aversion is max-min decision theory, where minimization is over mathematical expectations of utilities of plans associated with the alternative priors. Application of the max-min decision theory axiomatized by Gilboa and Schmeidler (1989) can produce a decision process that is supported by a Bayesian prior and is therefore admissible. This can be established by verifying applicability of the Minimax Theorem that justifies exchanging the order of maximization and minimization.[3] In particular, after exchanging orders of extremization, the outcome of the outer minimization is a worst-case prior for

[3]See Fan (1952).

which the decision process choice is "optimal" in a Bayesian sense. Good (1952) refers to such a worst-case prior in the above quote. Admissibility and dynamic consistency under this worst-case prior follow, for reasons discussed in previous paragraphs. Computing and assessing the plausibility of a worst-case prior are important parts of a robust Bayesian analysis.

Epstein and Schneider (2003) stress that a max-min decision rule with multiple priors may not be dynamically consistent. To address this, Epstein and Schneider extended an initial set of probability models in a recursive manner order to render the resulting max-min preferences dynamically consistent. It is useful to analyze how Epstein and Schneider proceeded. Temporarily consider a discrete-time setting in which $\epsilon > 0$ is the time increment. Start with a family of probabilities associated with a set of priors over the set of structured models. At date t, consider all possible probability assignments implied by possible choices of time 0 priors for events on $\mathfrak{F}_{t+\epsilon}$ conditioned on $\mathfrak{F}_t$. Define preferences recursively in terms of continuation values. For a date $t+\epsilon$ continuation value that is $\mathfrak{F}_{t+\epsilon}$ measurable, minimize over all possible mathematical expectations conditioned on $\mathfrak{F}_t$ and construct a date t continuation value that is $\mathfrak{F}_t$ measurable. When working backwards, notice how at each step this construction incorporates subjective discounting of the future as well as a contribution from current period utility. By adding probabilities constructed from all possible t to $t+\epsilon$ conditional probabilities, *including ones that come from distinct date zero priors*, this construction creates a new set of probabilities for which max-min preferences are dynamically consistent. Following Epstein and Schneider (2003), we refer to the enlarged set of probability distributions as a *rectangular* embedding of the decision maker's original set of probabilities.

In recommending this expanded set of probabilities as an appropriate object to use within a max-min decision theory, 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.

Thus, 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 create an enlarged set that is rectangular (i.e., their P). In this way, the recursive probability augmentation procedure constructs dynamically consistent preferences. But it does so by adding possibly implausible probability measures to the set of probabilities. By so doing, it can induce an inadmissible decision process with respect to the set of structured models that are of initial interest to the decision maker. Thus, applying the Minimax Theorem to the rectangular embedding of an initial subjectively specified set of probabilities can yield a worst-case probability that may or may not be one that is associated with a single prior over the decision maker's original family of structured models. When it is, admissibility of the max-min decision rule and applicability of Good's plausibility assessment procedure still prevail. This desirable outcome can happen, for instance, when the original specification of (priors over) structured models already implies a rectangular set of probabilities. But more generally, the worst-case probability can be in the expanded set P that Epstein and Schneider use to achieve a rectangular embedding and not be a member of the set of models $Prob$ that the decision maker thinks are possible. In that case, the plausibility criterion advocated by Good (1952) fails because the worst-case model is not among the possible models that interest the decision maker. This situation presents an applied economic model builder with a difficult conflict between dynamic consistency and admissibility.

Our paper studies two classes of examples that explore aspects of the tension between dynamic consistency and admissibility. In one class, a rectangular specification is justified on subjective grounds by how it captures time variation in parameters. In a continuous-time model that can be viewed as a limit of a discrete-time model attained by driving a time interval ϵ to zero, we draw on a representation provided by Chen and Epstein (2002) to verify rectangularity. Here admissibility and dynamic consistency coexist. In the other class, the two concepts cannot coexist because so many additional models must be added to construct a rectangular set that it renders inapplicable Good's plausibility assessment procedure. In particular, we show that expanding the set of structured models to include relative entropy neighborhoods and then adding the additional models to construct a rectangular set of probabilities requires adding a vast number of statistically implausible models that need only satisfy some weak absolute continuity restrictions over finite intervals of time. The vastness of the set of models can make the worst-case probability from this expanded set completely uninteresting from the standpoint of Good's test. We show that this second class of environments includes ones in which the decision maker is concerned about model misspecification.

In applications, we want to work with this second class of environments in which concerns about misspecification appear. Because we want preferences that are dynamically consistent and also entertain a large set of alternative models, we abandon the Gilboa and

Schmeidler (1989) way of representing ambiguity aversion in favor of the more general class of variational preferences conceived and justified by Maccheroni et al. (2006a). We employ the dynamic recursive formulation of Maccheroni et al. (2006b). To express Box's view that models are flawed but useful simplifications, we apply relative entropy to construct statistical neighborhoods around each of the structured models. These neighborhoods include what we call "unstructured models" that express the decision maker's concern about misspecifications of the structured models. Variational preferences allow us to implement a penalty on relative entropy in the context of preferences that are dynamically consistent while avoiding the extremely statistically implausible worst-case models associated with the embedding procedure of Epstein and Schneider (2003).

Thus, to explore misspecification, our decision maker enlarges the set of potential models beyond the structured ones, but for a different reason and in a different way than is done in the rectangular embedding procedure of Epstein and Schneider (2003). Indeed, our application of variational preferences allows a decision maker to explore decision-relevant consequences of a potentially large set of models represented in terms of a penalty on entropy. With this large set of models, it might at first be thought that admissibility is a less interesting concept because the decision maker articulates no details about the alternative specifications, saying only that he is interested in wanting to explore expected valuations under a vast collection of statistically nearby models of unknown forms. However, in section 7 we show that within our new framework, Good's way of judging the plausibility of a worst-case model remains workable and interesting. We do this by computing a worst-case structured model and a worst-case unstructured adjustment to that model and showing that both are statistically interesting to the decision maker. We also compute an implied statistical discrepancy that tells how concerned the decision maker is about misspecification under alternative values of the penalty parameter.

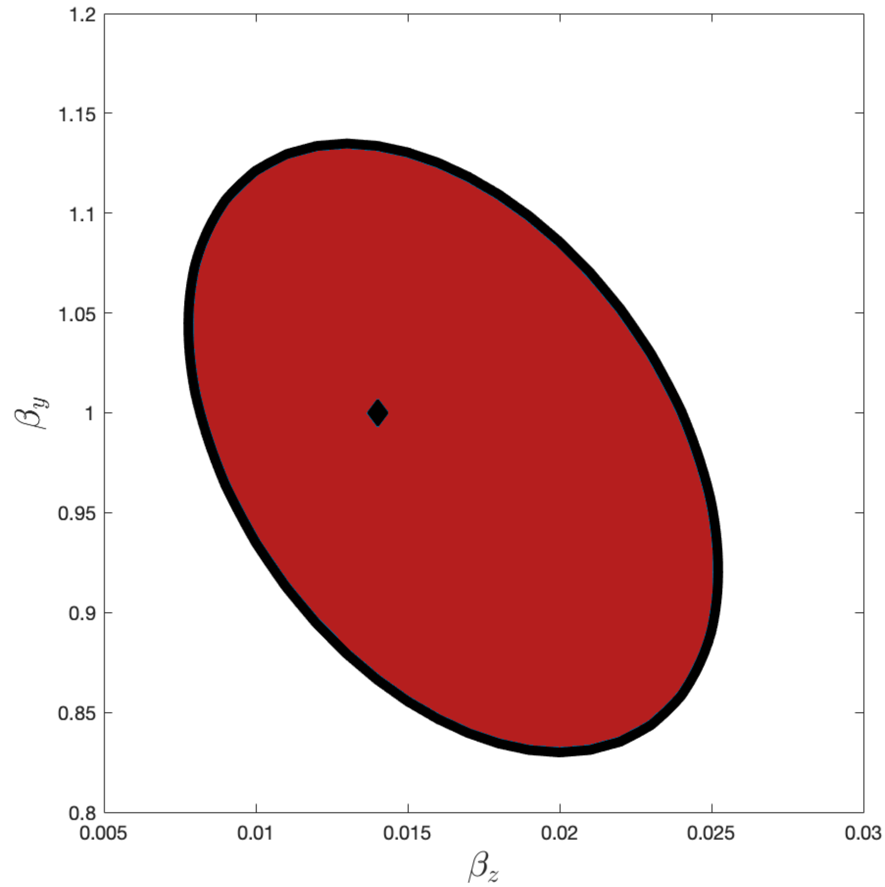

Figure 1: Ambiguity contour for (β_y, β_z) satisfying relative entropy constraint described in section 5.4. The boundary is computed with $\mathsf{q} = .05$, where $\frac{\mathsf{q}^2}{2}$ measures relative entropy. The black triangle depicts the baseline model abstracting from model ambiguity.

As an illustration, in section 5.4 we consider a model in which an investor is ambiguous about a parameter β_y that captures her exposure to random macroeconomic growth and a parameter β_z that describes persistence of macroeconomic growth rates. Larger values of β_z are associated with less temporal dependence. Figure 1 displays a point associated with baseline parameter values and an ambiguity region corresponding to a restricted specification of relative entropy. A Bayesian might place a prior over the (β_y, β_z) parameter space, while a robust Bayesian might consider a family of priors. Suppose instead that a decision maker's subjective inputs imply that the parameters reside in the convex portrayed in figure 1 and are expressed by the set of all priors over this same region. Choosing a worst-case prior once-and-for-all at date zero leads to a preference ordering that typically will not be dynamically consistent. To acquire dynamic consistency, we could expand the original set of probabilities by allowing a different date zero prior at each instant and using the implied local transition probability for the state dynamics to evaluate conditional expected utilities. The decision plan that emerges from this procedure might be inadmissible if the decision maker cares only about time-invariant parameter models. We shall show that we can acquire admissibility by adding to the decision maker's original set of models structured models that allow parameters to vary over time in a flexible way while continuing to embrace a dynamic version of max-min expected utility. But, as we will also show, when the decision maker also wants to explore model misspecifications by including unstructured models residing within a relative entropy neighborhood of the set of structured models, expanding an original set to make it rectangular as Epstein and Schneider recommend implies a degenerate and uninteresting decision problem. We overcome this problem by abandoning the min-max preferences of Gilboa and Schmeidler (1989) and by instead adopting a version of the variational preferences of Maccheroni et al. (2006b).

In Hansen and Sargent (2019), we apply our approach to a macroeconomic model that we use to explore consequences of a representative investor's ambiguity for equilibrium asset valuations. We show that the investor's worst-case model is statistically plausible and that it has interesting consequences for the investor's behavior and for equilibrium prices. The worst-cast model displays more growth rate persistence than does a baseline model when macroeconomic growth is sluggish and less persistence when macroeconomic growth is fast. This pattern induces new source of nonlinearities in responses of asset valuations to shock exposures.

3 Model perturbations

This section describes nonnegative martingales that we use to perturb a baseline probability model. Section 4 then describes how we use a family of parametric alternatives to a baseline model to form a convex set of martingales representing unstructured models that we shall use to pose robust decision problems.

3.1 Mathematical framework

To fix ideas, we use a specific *baseline* model and in section 4 an associated family of alternatives that we call *structured* models. A decision maker cares about a stochastic process $X \doteq \{X_t : t \geqslant 0\}$ that she approximates with a baseline model[4]

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

where W is a multivariate Brownian motion.[5] A *plan* is a $C = \{C_t : t \geqslant 0\}$ process that is progressively measurable with respect to the filtration $\mathfrak{F} = \{\mathfrak{F}_t : t \geqslant 0\}$ associated with the Brownian motion W augmented by information available at date zero. Progressively measurable means that the date t component C_t is measurable with respect to $\mathfrak{F}_t$. A decision maker cares about plans.

Because he does not fully trust baseline model (1), the decision maker explores utility consequences of other probability models that he obtains by multiplying probabilities associated with (1) by appropriate likelihood ratios. Following Hansen et al. (2006), we represent a likelihood ratio process by a positive martingale M^U with respect to the probability distribution induced by the baseline model (1). The martingale M^U 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 $\mathfrak{F}$. We adopt the convention that M_t^U is zero when $\int_0^t |U_\tau|^2 d\tau$ is infinite. 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 formally defined as a probability limit. Imposing the initial condition $M_0^U = 1$, we express the solution of stochastic

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

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

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

differential equation (2) as the stochastic exponential[7]

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

Definition 3.1. *$\mathcal{M}$ denotes the set of all martingales M^U that can be constructed as stochastic exponentials via representation* (5) *with a U that satisfies* (4) *and are progressively measurable with respect to $\mathfrak{F}$.*

Associated with U are probabilities defined by

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

for any $t \geqslant 0$ and any bounded $\mathfrak{F}_t$-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 successive conditional expectations operators E^U 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, its 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 write model (1) as:

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

[7] M_t^U specified as in (5) is a local martingale, but not necessarily a martingale. It is not convenient here to impose sufficient conditions for the stochastic exponential to be a martingale like Kazamaki's or Novikov's. Instead, we will verify that an extremum of a pertinent optimization problem does indeed result in a martingale.

4 Measuring statistical discrepancies

We use entropy relative to a baseline probability to restrict martingales that represent alternative probabilities.[8] We start with the likelihood ratio process M^U and from it construct ingredients of a notion of relative entropy for the process M^U. To begin, we note that the process $M^U \log M^U$ evolves as an Ito process with date t drift $\frac{1}{2} M_t^U |U_t|^2$ (also called a local mean). Write the conditional mean of $M^U \log M^U$ in terms of a history of local means as[9]

$$E\left[M_t^U \log M_t^U | \mathfrak{F}_0\right] = \frac{1}{2} E\left(\int_0^t M_\tau^U |U_\tau|^2 d\tau | \mathfrak{F}_0\right). \tag{8}$$

Also, let M^S be a martingale defined by a drift distortion process S that is measurable with respect to $\mathfrak{F}$. To construct entropy relative to a probability distribution affiliated with M^S instead of martingale M^U, we use a log likelihood ratio $\log M_t^U - \log M_t^S$ with respect to the M_t^S model to arrive at:

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

A notion of relative entropy appropriate for stochastic processes is

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

provided that these limits exist. The second line is the limit of Abel integral averages, where scaling by δ makes the weights $\delta \exp(-\delta\tau)$ integrate to one. Rather than using undiscounted relative entropy, we find it convenient sometimes to use Abel averages with a discount rate equal to the subjective rate that discounts an expected utility flow. With

[8] Entropy is widely used to measure discrepancies between models in the statistical and machine learning literatures. For example, see Amari (2016) and Nielsen (2014).

[9] A variety of sufficient conditions justify equality (8). When we choose a probability distortion to minimize expected utility, we will use representation (8) 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. See their Claims 6.1 and 6.2.

that in mind, we define a discrepancy between two martingales M^U and M^S as:

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

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

$$\Delta(M^U; 1|\mathfrak{F}_0) = \frac{\delta}{2}\int_0^\infty \exp(-\delta t) E\left(M_t^U |U_t|^2 \Big| \mathfrak{F}_0\right) dt \geqslant 0, \tag{9}$$

where baseline probabilities are represented here by the degenerate $S_t \equiv 0$ drift distortion that is affiliated with a martingale that is identically one. Formula (9) quantifies how a martingale M^U distorts baseline model probabilities.

5 Families of structured models

We construct a family of structured probabilities by forming a set $\mathcal{M}^o$ of martingales M^S with respect to a baseline probability associated with model (1) using a formulation of Chen and Epstein (2002). Formally,

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

where $\Gamma = \{\Gamma_t\}$ is a process of convex sets adapted to the filtration $\mathfrak{F}$.[10]

Hansen and Sargent (2001) and Hansen et al. (2006) started from a unique baseline model and then surrounded it with a relative entropy ball of unstructured models. In this paper, we instead start from a convex set $\mathcal{M}^o$ such that $M^S \in \mathcal{M}^o$ is a set of martingales with respect to a conveniently chosen and unique baseline model. The set $\mathcal{M}^o$ represents a set of *structured* models that in section 6 we shall surround with an entropy ball of unstructured models. This section contains several examples of sets of structured models formed according to particular versions of (10). Subsection 5.1 starts with a parametric family; subsection 5.2 then adds time-varying parameters, while subsection 5.3 uses relative entropy to construct a set of structured models.

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

5.1 Finite number of underlying models

We present two examples that feature a finite number n of structured models of interest, with model j being represented by an S_t^j process that is a time-invariant function of the Markov state X_t for $j = 1, \ldots, n$. The two examples differ in their processes of convex sets $\{\Gamma_t\}$ defining the set of martingales $\mathcal{M}^o$ in (10).

5.1.1 Time-invariant models

Each S^j process represents a probability assignment for all $t \geqslant 0$. Let Π_0 denote a convex set of probability vectors that reside in a subset of the probability simplex in $\mathbb{R}^n$. Alternative $\pi_0 \in \Pi_0$'s are potential initial period priors across models.

To update under a prior $\pi_0 \in \Pi_0$, we 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 prior $\pi_0 \in \Pi_o$ assign probability $\pi_0^j \geqslant 0$ to model S^j, where $\sum_{i=1}^n \pi_0^j = 1$. A martingale

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

characterizes a mixture of S^j models. The mathematical expectation of M_t conditioned on date zero information equals unity for all $t \geqslant 0$. Martingale M evolves as

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

where the date t posterior π_t^j probability assigned to model S^j is

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

and the associated drift distortion of martingale M is

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

It is helpful to frame the potential conflict between admissibility and dynamic consistency in terms of a standard robust Bayesian formulation of a time 0 decision problem. A positive martingale generated by a process S implies a change in probability measure. Consider probability measures generated by the set

$$\Gamma = \left\{ S = \{S_t : t \geqslant 0\} : S_t = \sum_{j=1}^{n} \pi_t^j S_t^j, \ \pi_t^j = \frac{\pi_0^j M_t^{S^j}}{\sum_{\ell=1}^{n} \pi_0^\ell M_t^{S^\ell}}, \ \pi_0 \in \Pi_0 \right\}.$$

This family of probabilities indexed by an initial prior will in general not be rectangular so that max-min preferences with this set of probabilities violate the Epstein and Schneider (2003) dynamic consistency axiom. Nevertheless, think of a max-min utility decision maker who solves a date zero choice problem by minimizing over initial priors $\pi_0 \in \Pi_0$. Standard arguments that invoke the Minimax theorem to justify exchanging the order of maximization and minimization imply that the max-min utility worst-case model can be admissible and thus allow us to apply Good's plausibility test.

We can create a rectangular set of probabilities by adding other probabilities to the family of probabilities associated with the set of martingales Γ. To represent this rectangular set, let Π_t denote the associated set of date t posteriors and form the set:

$$\Gamma_t = \left\{ S_t = \sum_{j=1}^{n} \pi_t^j S_t^j, \ \pi_t \in \Pi_t \right\}.$$

Think of constructing alternative processes S by selecting alternative $S_t \in \Gamma_t$. Notice that here we index conditional probabilities by a process of potential posteriors π_t that no longer need be tied to a single prior $\pi_0 \in \Pi_0$. This means that more probabilities are entertained than were under the preceding robust Bayesian formulation that was based on a single worst-case time 0 prior $\pi_0 \in \Pi_0$. Now admissibility relative to the initial set of models does not necessarily follow because we have *expanded* the set of models to obtain rectangularity.

Thus, these constructions of alternative sets of potential S processes generated by the set Γ, on one hand, and the sets Γ_t, on the other hand, illustrate the tension between admissibility and dynamic consistency within the Gilboa and Schmeidler (1989) max-min utility framework.

5.1.2 Pools of models

Geweke and Amisano (2011) propose a procedure that averages predictions from a finite pool of models. Their suspicion that all models within the pool are misspecified motivates Geweke and Amisano to choose weights over models in the pool that improve forecasting performance. These weights are not posterior probabilities over models in the pool and may not converge to limits that "select" a single model from the pool, in contrast to what often happens when weights over models are Bayesian posterior probabilities. Waggoner and Zha (2012) extend this approach by explicitly modeling time variation in the weights according to a well behaved stochastic process.

In contrast to this approach, our decision maker expresses his specification concerns formally in terms of a set of structured models. An agnostic expression of the decision maker's weighting over models can be represented in terms of the set

$$\Gamma_t = \left\{ S_t = \sum_{j=1}^{n} \pi_t^j S_t^j, \ \pi_t \in \overline{\Pi} \right\},$$

where $\overline{\Pi}$ is a time invariant set of possible model weights that can be taken to be the set of all potential nonnegative weights across models that sum to one. A decision problem can be posed that determines weights that vary over time in ways designed to manage concerns about model misspecifications. To employ Good's 1952 criterion, the decision maker must view a weighted average of models as a plausible specification.[11]

In the next subsection, we shall consider other ways to construct a set $\mathcal{M}^o$ of martingales that determine structured models that allow time variation in parameters.

[11]For some of the examples of Waggoner and Zha that take the form of mixtures of rational expectations models, this requirement could be problematic because mixtures of rational expectations models are not rational expectations models.

5.2 Time-varying parameter models

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

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

Here the unknown parameter vector is $\theta = \begin{bmatrix} \theta^1 & \theta^2 & \ldots & \theta^n \end{bmatrix}' \in \Theta$, a closed convex subset of $\mathbb{R}^n$. We can include time-varying parameter models by changing (11) to:

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

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

$$\Gamma_t = \left\{ S_t = \sum_{j=1}^{n} \bar{\theta}_t^j S_t^j, \ \bar{\theta}_t \in \Theta, \ \bar{\theta}_t \text{ is } \mathcal{F}_t \text{ measurable} \right\}. \tag{13}$$

5.3 Structured models restricted by relative entropy

We can construct a set of martingales $\mathcal{M}^o$ by imposing a constraint on entropy relative to a baseline model that restricts drift distortions as functions of the Markov state. This method has proved useful in applications.

Section 4 defined relative entropy for a stochastic process generated by martingale M^S to be

$$\varepsilon(M^S) = \lim_{t \to \infty} \frac{1}{2t} \int_0^t E\left(M_\tau^S |S_\tau|^2 \middle| \mathfrak{F}_0 \right) d\tau. \tag{14}$$

Evidently $\varepsilon(M^S)$ is the limit as $t \to +\infty$ of a process of mathematical expectations of time series averages

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

under the probability measure implied by martingale M^S. Suppose, for instance, that M^S is defined by drift distortion $S_t = \eta(X_t)$, where X is an ergodic Markov process with transition probabilities that converge to a well-defined and unique stationary distribution Q under the M^S probability. In this case, we can compute relative entropy from

$$\varepsilon(M^S) = \frac{1}{2}\int |\eta|^2 dQ. \tag{15}$$

In what follows, we parameterize relative entropy by $\frac{\mathsf{q}^2}{2}$, where q measures the magnitude of the drift distortion using a mean-square norm.

We want the decision maker's set of structured models to be rectangular in the sense that it satisfies an instant-by-instant constraint $S_t \in \Gamma_t$ for all $t \geqslant 0$ in (10) for a collection of $\mathfrak{F}_t$-measurable convex sets $\{\Gamma_t : t \geqslant 0\}$. We cannot accomplish this by simply specifying an upper bound $\frac{\mathsf{q}^2}{2}$ on relative entropy and then finding all drift distortion S processes that satisfy $\varepsilon(M^S) \leqslant \frac{\mathsf{q}^2}{2}$ because that would produce a family of probabilities that fails to satisfy an instant-by-instant rectangularity constraint of the form in (10) that we want. Furthermore, starting with a set of probabilities that satisfies $\varepsilon(M^S) \leqslant \frac{\mathsf{q}^2}{2}$ and enlarging it to make it rectangular as Epstein and Schneider recommend would yield a set of probabilities that is much much too large for max-min preferences, as we describe in detail in section 7.2.

To implement an instant-by-instant constraint, we restrain what is essentially a time derivative of relative entropy using logic very similar to that employed in deriving HJB equations. By bounding the time derivative of relative entropy we necessarily make the resulting constraint on structured models more restrictive. To motivate an HJB equation, we start with a low frequency refinement of relative entropy: For $S_t = \eta(X_t)$ consider the log-likelihood-ratio process

$$\begin{aligned} L_t &= \int_0^t \eta(X_\tau) \cdot dW_\tau - \frac{1}{2}\int_0^t |\eta(X_\tau)|^2 d\tau \\ &= \int_0^t \eta(X_\tau) \cdot dW_\tau^S + \frac{1}{2}\int_0^t |\eta(X_\tau)|^2 d\tau. \end{aligned} \tag{16}$$

From (14), relative entropy is the long-horizon limiting average of the expectation of L_t under M^S probability. To refine a characterization of its limiting behavior, we note that a log-likelihood process has an additive structure that allows us to decompose L_t as:

$$L_t = \frac{\mathsf{q}^2}{2} t + D_t + \rho(X_0) - \rho(X_t) \tag{17}$$

where

$$D_t = \int_0^t \left[\left(\frac{\partial \rho}{\partial x}(X_\tau) \right)' \sigma + \eta(X_\tau) \right] \cdot dW_\tau^S.$$

See the discussion in Hansen (2012, Sec. 3). Shortly we will describe how to compute the function ρ.[12] The process $\{D_t :\geqslant 0\}$ is constructed as a stochastic integral and is typically a martingale under the probability measure induced by M^S. Since $D_0 = 0$, $E\left(M_t^S D_t \mid X_0 = 0\right)$ for all $t \geqslant 0$. We use decomposition (17) to describe the behavior of the log-likelihood over long horizons. Notice that coefficient $\frac{\mathsf{q}^2}{2}$ on the trend term in decomposition (17) is relative entropy, as could be anticipated from the definition of relative entropy as a long-run average. Subtracting the time trend and taking date zero conditional expectations under the probability measure induced by M^S gives:

$$\begin{aligned} \lim_{t\to\infty} \left[E\left(M_t^S L_t | X_0 = x\right) - \frac{\mathsf{q}^2}{2} t \right] =& \lim_{t\to\infty} E\left(M_t^S [D_t - \rho(X_t)] \mid X_0 = x\right) + \rho(x) \\ =& \rho(x) - \int \rho dQ, \end{aligned}$$

where Q is the limiting stationary distribution under the M^S probability in the sense that

$$\lim_{t\to\infty} E\left(M_t^S \rho(X_t) | X_0 = x\right) = \int \rho dQ.$$

This limit is valid because X is presumed to be stochastically stable under the S implied probability. Thus, $\rho - \int \rho dQ$ provides a long-horizon first-order refinement of relative entropy.

To compute relative entropy using recursive methods, we provide a functional equation in ρ and q. We start by representing an instantaneous counterpart to a discrete-time one-period transition distribution for a Markov process in terms of an infinitesimal generator that describes how conditional expectations of the Markov state evolve locally. The generator can be derived informally by differentiating the family of conditional expectation operators with respect to the gap of elapsed time. For a diffusion like baseline model (1), the infinitesimal generator $\mathcal{A}^\eta$ of transitions under the M^S probability is the second-order differential operator

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

[12]Section three of Hansen (2012) states a formal result and cites closely related sources.

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

$$\int \mathcal{A}^\eta \rho dQ = 0. \tag{18}$$

Equation (18) can be derived heuristically by applying the Law of Iterated expectations to predicting $\rho(x)$.

By equating the local means implied by equations (16) and (17) we obtain the relation:

$$\frac{|\eta|^2}{2} = \frac{\mathsf{q}^2}{2} - \mathcal{A}^\eta \rho. \tag{19}$$

We can use this equation to solve for the pair (ρ, q); ρ is determined only up to translation by a constant. Integrating (19) with respect to Q, and substituting from equation (18), we verify that $\frac{\mathsf{q}^2}{2}$ is relative entropy.[13]

Instead of pre-specifying η and solving for (q, ρ), we now use (q, ρ) to restrict η and then go one step further by constructing the sequence of $\mathfrak{F}_t$-measurable convex sets that we can use to define the decision maker's set of structured probability models. Rearranging terms in (19) gives

$$\frac{|\eta|^2}{2} + \eta \cdot \left(\sigma' \frac{\partial \rho}{\partial x}\right) = \frac{\mathsf{q}^2}{2} - \mathcal{A}^0 \rho \tag{20}$$

We construct a "boundary" of the set of the set of interest by locating η's that satisfy (20). Each η on this boundary has the same entropy $\frac{\mathsf{q}^2}{2}$ and the same first-order refinement ρ. We think of this boundary as an ambiguity contour. Notice that this contour is constructed in terms of the drift restriction used in computing relative entropy. Using a pre-specified ρ in addition to q, as is done here, limits substantially the η's.

Notice that (20) is quadratic in the function η and is an a sphere for each value of x. The (state dependent) center of this sphere is $-\sigma' \frac{\partial \rho}{\partial x}$ and the radius is $\frac{\mathsf{q}^2}{2} - \mathcal{A}^0 \rho + \left|\sigma' \frac{\partial \rho}{\partial x}\right|^2$. We fill this sphere to construct the convex set of interest used to restrict S_t:

$$\Gamma_t = \left\{ s : \frac{|s|^2}{2} + s \cdot \left[\sigma(X_t)' \frac{\partial \rho}{\partial x}(X_t) \right] \leqslant \frac{\mathsf{q}^2}{2} - \mathcal{A}^0 \rho(X_t) \right\}. \tag{21}$$

[13]This approach to computing relative entropy has direct extensions to Markov jump processes and mixed jump diffusion processes. For diffusion processes, equation (19) is a special case of a Feynman-Kac equation. Had we wanted to compute discounted relative entropy, equation (19) would include a term $-\delta\rho$ on the left-side and the term $\frac{\mathsf{q}^2}{2}$ would be omitted. Discounted relative entropy is state-dependent and given by $\delta\rho(x)$ with ρ satisfying a different Feynman-Kac equation.

To ensure a solution, the function ρ can be constructed by using a candidate $\tilde{\eta}$ on the sphere and solving for the implied (q, ρ).[14]

There exist many S processes that have relative entropy that is less than or equal to $\frac{\mathsf{q}^2}{2}$ but that violate the instant-by-instant inequality on the right-hand side of definition (21). Thus, by using process of sets Γ_t given by (21) to form the set of probabilities defined in (10), we are imposing a refinement of the relative entropy constraint $\int \frac{|s|^2}{2} dQ \leqslant \frac{\mathsf{q}^2}{2}$. in the sense that many processes satisfy the relative entropy constraint but violate the rectangularity constraint incorporated in definition (21).

Remark 5.1. *As an alternative, we could have imposed the restriction:*

$$\frac{|S_t|^2}{2} \leqslant \frac{\mathsf{q}^2}{2}$$

While this would also impose a quadratic refinement on relative entropy and is tractable to implement, the boundary will typically have relative entropies that are strictly less than $\frac{\mathsf{q}^2}{2}$. *Moreover, for some examples that interest us motivated by unknown coefficients, the* S_t*'s are not bounded independently of the Markov state.*

Remark 5.2. *As another alternative, we could impose a state-dependent restriction:*

$$\frac{|S_t|^2}{2} \leqslant \frac{|\tilde{\eta}(X_t)|^2}{2}$$

where $\tilde{\eta}(X_t)$ *is constructed with a particular model in mind, perhaps motivated by uncertain parameters. While this would be tractable and could be of interest for some applications, its connection to relative entropy is less evident. For instance, even if this restriction is satisfied, the relative entropy of the* S *model may exceed that of the* $\{\eta(X_t) : t \geqslant 0\}$ *model because respective relative entropies are computed by taking expectations with different probability specifications.*

In summary, we have shown how to use relative entropy along with an additional refinement to construct a family of structured models. By specifying the function ρ (up to a translation) along with the relative entropy q, we restrict the set of structured models to be rectangular. If we had instead specified only relative entropy q and not the function ρ too, the set of models would cease to be rectangular, as we discuss in detail in section 7.2.

[14] Had we chosen to use discounted relative entropy we have a corresponding constraint with $\delta\rho$ replacing $\frac{\mathsf{q}^2}{2}$ and ρ satisfying the discounted version of the corresponding Feynman-Kac equation.

If we were modeling a decision maker who is interested only in a set of models defined by (10), we could stop here and use a dynamic version of the max-min preferences of Gilboa and Schmeidler (1989). That way of proceeding could indeed lead to interesting applications and is well worth pursuing in its own right. But because he distrusts all of those models, the decision maker who is the subject of this paper also wants to investigate the utility consequences of models not in the set defined by (10). This will lead us to an approach in section 6 that uses a continuous-time version of the variational preferences that are an extension of max-min preferences. Before doing that, we describe an example of a set of structured models that naturally occur in an application of interest to us.

5.4 Illustration

In this subsection, we offer an example of a set $\mathcal{M}^o$ for structured models that can be constructed by the approach of subsection 5.3. We start with a baseline parametric model for a representative investor's consumption process Y, then form a family of parametric structured probability models. We deduce the pertinent version of the second-order differential equation (19) to be solved for ρ. The baseline model for consumption is

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

We scale by .01 because we want to work with growth rates and Y is typically expressed in logarithms. The mean of Z in the implied stationary distribution is $\bar{z} = \widehat{\alpha}_z / \widehat{\beta}_z$.

Let

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

The decision maker focuses on the following collection of alternative structured parametric models:

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

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

are parameters common to models (22) and (23).

We represent members of the parametric class defined by (23) in terms of our section 3.1 structure with drift distortions S of the form

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

then use (1), (6), and (23) to deduce the following restrictions on η_1:

$$\sigma\eta_1 = \begin{bmatrix} \beta_y - \widehat{\beta}_y \\ \widehat{\beta}_z - \beta_z \end{bmatrix} \tag{24}$$

where

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

Given an η that satisfies these restrictions, we compute a function ρ that is quadratic and depend only on z so that $\rho(x) = \rho^o(z)$. Relative entropy $\frac{\mathsf{q}^2}{2}$ emerges as part of the solution to the following relevant instance of differential equation (19):

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

Under parametric alternatives (23), the solution for ρ is quadratic in $z - \bar{z}$. Write:

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

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

$$\frac{|S_t|^2}{2} + [\rho_1 + \rho_2(Z_t - \bar{z})]\,\sigma_z \cdot S_t \leqslant \frac{|\sigma_z|^2}{2}\rho_2 - \frac{\mathsf{q}^2}{2} - [\rho_1 + \rho_2(Z_t - \bar{z})]\,\widehat{\beta}(\bar{z} - Z_t)$$

Figure 1 portrays an example in which $\rho_1 = 0$ and ρ_2 satisfies:

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

When $S_t = \eta(Z_t)$ is restricted to be $\eta_1(Z_t - \bar{z})$, a given value of q imposes a restriction on

η_1 and, through equation (24), implicitly on (β_y, β_z). Figure 1 plots the $\mathsf{q} = .05$ iso-entropy contour as the boundary of a convex set for (β_y, β_z).[15]

While Figure 1 displays contours of time-invariant parameters with the same relative entropies as the boundary of convex region, our restriction allows parameters(β_y, β_z) to vary over time provided that they remain within the plotted region. Indeed, we use (10) as a convenient way to build a set of structured models that includes ones with time varying parameters that lack probabilistic descriptions of how parameters vary.

If we were to stop here and endow a max-min decision maker with the set of probabilities determined by the set of martingales $\mathcal{M}^o$, we could study max-min preferences associated with this set of probabilities. Restriction (10) on the set of $\mathcal{M}^o$ martingales guarantees that the set of probabilities is rectangular and that therefore these preferences satisfy the dynamic consistency axiom of Epstein and Schneider (2003) that justifies dynamic programming. However, as we emphasize in section 6, our decision maker expands the set of models because he wants to evaluate outcomes under probability models inside relative entropy neighborhoods of structured models. This expanded set is not rectangular and for reasons stated formally in subsection 7.2 can't be made rectangular by following Epstein and Schneider's expansion procedure and still yield a set of models of interest to a decision maker who like ours wants to apply Good's plausibility criterion. That motivates us to penalize relative entropies from the family of structured models in $\mathcal{M}^o$ in order to describe additional potential misspecifications taking the form of unstructured models that reside within a vast collection of models that fit nearly as well as the structured models in $\mathcal{M}^o$. We describe details in section 6. But first we briefly describe alternative approaches.

5.5 Other approaches

In our example so far, we assumed that the structured model probabilities can be represented as martingales with respect to a baseline model. A different approach, invented by Peng (2004), uses a theory of stochastic backward differential equations under a no-

[15]This figure was constructed using the parameter values:

$$\begin{array}{rclcrcl} \widehat{\alpha}_y & = & .484 & \quad & \widehat{\beta}_y & = & 1 \\ \widehat{\alpha}_z & = & 0 & \quad & \widehat{\beta}_z & = & .014 \end{array}$$

$$(\sigma_y)' = \begin{bmatrix} .477 & 0 \end{bmatrix}$$

$$(\sigma_z)' = \begin{bmatrix} .011 & .025 \end{bmatrix}$$

taken from Hansen and Sargent (2019).

tion of ambiguity that is rich enough to allow for uncertainty in conditional volatilities of Brownian increments.[16] Because alternative probability specifications fail to be absolutely continuous (over finite time intervals), standard likelihood ratio analysis does not apply. This approach would push us outside the Chen and Epstein (2002) formulation but would still let us construct a rectangular embedding that we use could use to construct structured models.[17]

6 Including unstructured alternatives

In section 5.1, we described how the decision maker forms a set $\mathcal{M}^o$ of structured models that are parametric alternatives to the baseline model. To represent the unstructured models that also concern the decision maker, we proceed as follows. After constructing $\mathcal{M}^o$, for scalar $\xi > 0$, we define a scaled discrepancy of martingale M^U from a set of martingales $\mathcal{M}^o$ as

$$\begin{aligned} \Xi(M^U|\mathfrak{F}_0) &= \xi \inf_{M^S \in \mathcal{M}^o} \Delta\left(M^U; M^S|\mathcal{F}_0\right) \\ &= \frac{\xi\delta}{2}\int_0^\infty \exp(-\delta t) E\left[M_t^U \gamma_t(U_t)\Big|\mathfrak{F}_0\right] dt. \end{aligned} \tag{25}$$

where

$$\gamma_t(U_t) = \inf_{S_t \in \Gamma_t} |U_t - S_t|^2. \tag{26}$$

Scaled discrepancy $\Xi(M^U|\mathfrak{F}_0)$ equals zero for M^U in $\mathcal{M}^o$ and is positive for M^U not in $\mathcal{M}^o$. We use discrepancy $\Xi(M^U|\mathfrak{F}_0)$ to define a set of unstructured models near $\mathcal{M}^o$ whose utility consequences a decision maker wants to know. When we pose a max-min decision problem, the scaling parameter ξ will be used to measure how the expected utility minimizer is penalized for choosing unstructured models that are statistically farther from the structured models in $\mathcal{M}^o$.

The decision maker doesn't stop with the set of structured models generated by martingales in $\mathcal{M}^o$ because he wants to evaluate the utility consequences not just of the structured models in $\mathcal{M}^o$ but also of unstructured models that statistically are difficult to distinguish from them. For that purpose, he employs the scaled statistical discrepancy measure

[16]See Chen et al. (2005) for a further discussion of Peng's characterizations of a class of nonlinear expectations to Choquet integration used in decision theory in both economics and statistics.

[17]See Epstein and Ji (2014) for an application of the Peng analysis to asset pricing.

$\Xi(M^U|\mathfrak{F}_0)$ defined in (25).[18]

7 Recursive Representation of Preferences

The decision maker uses relative entropy implied by the scaling parameter ξ to restrain the statistical discrepancy between unstructured models and the set of structured models. The decision maker solves a minimization problem in which ξ serves as a penalty parameter that effectively excludes unstructured probabilities that are statistically too far from the set $\mathcal{M}^o$ of structured models. That minimization problem induces a special case of the dynamic variational preference ordering that Maccheroni et al. (2006b) showed are dynamically consistent.

7.1 Continuation values

The decision maker ranks alternative consumption plans with a scalar continuation value stochastic process. Date t continuation values tell a decision maker's date t ranking. Continuation value processes have a recursive structure that makes preferences be dynamically consistent. Thus, for Markovian plans, a Hamilton-Jacobi-Bellman (HJB) equation restricts the evolution of continuation values. In particular, for a consumption plan $\{C_t\}$, a continuation value process $\{V_t\}_{t=0}^{\infty}$ is defined by

$$V_t = \min_{\{U_\tau : t \leqslant \tau < \infty\}} E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M^U_{t+\tau}}{M^U_t}\right)\left[\psi(C_{t+\tau}) + \left(\frac{\xi\delta}{2}\right)\gamma_{t+\tau}(U_{t+\tau})\right] d\tau \mid \mathfrak{F}_t\right) \tag{27}$$

where ψ is an instantaneous utility function. We can use (27) to derive an inequality that describes a sense in which a minimizing process $\{U_\tau : t \leqslant \tau < \infty\}$ isolates a statistical model that is robust. After deriving and discussing this inequality and the associated robustness bound, we shall use (27) to provide a recursive representation of preferences.

Turning to the derived bound, we proceed by applying an inequality familiar from optimization problems subject to penalties. Let U^o be the minimizer for problem (27) and let $S^o = S(U^o)$ be the minimizing S implied by equation (26). The process affiliated with the pair (U^o, S^o) gives a lower bound on discounted expected utility that can be represented in the following way.

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

Bound 7.1. *If* (U, S) *satisfies:*

$$\frac{\delta}{2} E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^U}{M_t^U}\right)|S_{t+\tau} - U_{t+\tau}|^2 d\tau \mid \mathfrak{F}_t\right)$$
$$\leqslant \frac{\delta}{2} E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^{U^o}}{M_t^{U^o}}\right)|S_{t+\tau}^o - U_{t+\tau}^o|^2 d\tau \mid \mathfrak{F}_t\right) \tag{28}$$

then

$$E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^U}{M_t^U}\right)\psi(C_{t+\tau}) d\tau \mid \mathfrak{F}_t\right)$$
$$\geqslant E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^{U^o}}{M_t^{U^o}}\right)\psi(C_{t+\tau}) d\tau \mid \mathfrak{F}_t\right) \tag{29}$$

for all $t \geqslant 0$.

Inequality (29) is a direct implication of minimization problem (27). It gives probability specifications that have date t discounted expected utilities that are at least as large as the one parameterized by U^o. The structured models all satisfy this bound; so do unstructured models that are statistically close to them as measured by the date t conditional counterpart to our discrepancy measure.

Turning next to a recursive representation of preferences, note that equation (27) implies that

$$V_t = \min_{\{U_\tau : t \leqslant \tau < t+\epsilon\}} \left\{ E\left[\int_0^\epsilon \exp(-\delta\tau)\left(\frac{M_{t+\tau}^U}{M_t^U}\right)\left[\psi(C_{t+\tau}) + \left(\frac{\xi\delta}{2}\right)\gamma_{t+\tau}(U_{t+\tau})\right] d\tau \mid \mathfrak{F}_t\right] \right.$$
$$\left. + \exp(-\delta\epsilon) E\left[\left(\frac{M_{t+\epsilon}^U}{M_t^U}\right) V_{t+\epsilon} \mid \mathfrak{F}_t\right]\right\} \tag{30}$$

for $\epsilon > 0$. Heuristically, we can "differentiate" the right side of (30) with respect to ϵ to obtain an instantaneous counterpart to a Bellman equation. Viewing the continuation value process $\{V_t\}$ as an Ito process, write:

$$dV_t = \nu_t dt + \varsigma_t \cdot dW_t.$$

A local counterpart to (30) is then

$$0 = \min_{U_t}\left[\psi(C_t) - \frac{\xi\delta}{2}\gamma_t(U_t) - \delta V_t + U_t \cdot \varsigma_t + \nu_t\right]$$

$$
\begin{aligned}
&= \min_{S_t \in \Gamma_t} \min_{U_t} \left[\psi(C_t) + \frac{\xi\delta}{2} |U_t - S_t|^2 - \delta V_t + U_t \cdot \varsigma_t + \nu_t \right] \\
&= \min_{S_t \in \Gamma_t} \left[\psi(C_t) - \frac{1}{2\xi\delta} \varsigma_t \cdot \varsigma_t - \delta V_t + S_t \cdot \varsigma_t + \nu_t \right] \qquad (31)
\end{aligned}
$$

where the minimizing U_t expressed as a function of S_t satisfies

$$
U_t = S_t - \frac{1}{\delta\xi} \varsigma_t
$$

The term $U_t \cdot \varsigma_t$ on the right side of (31) comes from an Ito adjustment to the local covariance between $\frac{dM_t^U}{M_t^U}$ and dV_t. Equivalently, $U_t \cdot \varsigma_t$ is an adjustment to the drift ν_t of dV_t that is induced by using martingale M^U to change the probability measure. For a continuous-time Markov decision problem, (31) gives rise to an HJB equation for a corresponding value function expressed as a function of a Markov state.

Remark 7.2. *With preferences described by* (31)*, we can still discuss admissibility relative to a set of structured models using the representation on the third line of* (31)*. Recall that the S process parameterizes a structured model. For a given decision process C, solve*

$$
0 = \psi(C_t) - \frac{1}{2\xi\delta} \tilde{\varsigma}_t \cdot \tilde{\varsigma}_t - \delta \tilde{V}_t + S_t \cdot \tilde{\varsigma}_t + \tilde{\nu}_t
$$

where

$$
d\widetilde{V} = \tilde{\nu}_t dt + \tilde{\varsigma}_t \cdot dW_t.
$$

Solving this equation backwards for alternative C processes gives a ranking of them for a given S probability. By posing a Markov decision problem, we can study admissibility by applying a Minimax theorem along with a Bellman-Isaacs condition for a dynamic two-person game. See, for instance, Fleming and Souganidis (1989). Provided that we can exchange orders of maximization and minimization, the implied worst-case structured model process S^* can be used in the fashion recommended by Good (1952) in the quote with which this paper began.

By extending Bound 7.1, the implied adjustment U^* for misspecification of the structured models is also enlightening. Specifically, we can use (U^*, S^*) in place of (U^o, S^o) in inequality (28) *and conclude that a counterpart to inequality* (29) *holds in which we maximize both the right and left sides by choice of a C plan subject to the constraints imposed on the decision problem. Thus, the entropy of U^* relative to S^* tells us over what probabilities we can bound discounted expected utilities.*

Remark 7.3. *It is useful to compare roles of the baseline model here and in the robust decision model based on the multiplier preferences of Hansen and Sargent (2001) and Hansen et al. (2006), another continuous time version of variational preferences.*[19] *Their baseline model is a unique structured model, distrust of which motivates a decision maker to compute a worst-case unstructured model to guide evaluations and decisions. In the present paper, the baseline model is just one of a set of structured models that the decision maker maintains. The baseline model here merely anchors specifications of other members of the set of structured models. The decision maker in this paper distrusts all models in the set of structured models associated with martingales in* $\mathcal{M}^o$.

7.2 With relative entropy and rectangularity, anything goes

In this subsection, we show that if a decision maker starts with a set of unstructured models constrained by relative entropy to be close to the set of structured models, enlarging that set to make it rectangular results in the set of all unstructured models that are absolutely continuous to a structured model over finite intervals. Most of those are statistically very implausible and unworthy of the decision maker's concern.

Our decision maker starts with a set of structured probability models that we have constructed to be rectangular in the sense of Epstein and Schneider. But our decision maker's suspicion that all of these structured models are misspecified motivates him to explore the utility consequences of a larger set that includes unstructured probability models. This larger set is not rectangular, even though as measured by relative entropy, all of the unstructured models are statistically close to models in the rectangular set formed by the structured models.

An alternative approach to formulating the decision maker's problems with the dynamic variational preferences of Maccheroni et al. (2006b) would have been first to construct a set that includes relative entropy neighborhoods of all martingales in $\mathcal{M}^o$. For instance,

[19] Our way of formulating preferences differs from how equation (17) of Maccheroni et al. (2006b) describes Hansen and Sargent (2001) and Hansen et al. (2006)'s "multiplier preferences". The disparity reflects what we regard as a minor blemish in Maccheroni et al. (2006b). The term $\frac{\xi\delta}{2}\gamma_t$ in our analysis is γ_t in Maccheroni et al. (2006b) and our equation (31) is a continuous time counterpart to equation (12) in their paper. In Hansen and Sargent (2001) and Hansen et al. (2006), $\gamma_t = |U_t|^2$ as we define γ_t. We point out this minor error here only because the analysis in the present paper generalizes our earlier work by measuring discrepancy from a non-singleton set $\mathcal{M}^o$ of structured models.

for $\epsilon > 0$, we could have started with a set

$$\overline{\mathcal{M}} = \left\{ M^U \in \mathcal{M} : \Xi(M^U | \mathfrak{F}_0) < \epsilon \right\} \tag{32}$$

that yields a set of implied probabilities that are not rectangular. At this point, why not follow Epstein and Schneider's (2003) recommendation and add just enough martingales to attain a rectangular set of probability measures? A compelling practical reason not to do so is that doing so would include all martingales in $\mathcal{M}$ defined in definition 3.1 – implying a set much too large for an interesting max-min decision analysis.

To show this, it suffices to look at relative entropy neighborhoods of the baseline model.[20] To construct a rectangular set of models that includes the baseline model, for a fixed date τ, consider a random vector $\overline{U}_\tau$ that is observable at τ and that satisfies

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

Form a stochastic process

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

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

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

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

Proposition 7.4. *Any rectangular set of probabilities that contains the probabilities induced by martingales in* (32) *must also contain the probabilities induced by any martingale in* $\mathcal{M}$.

This rectangular set of martingales allows far too much freedom in setting date τ and random vector $\overline{U}_\tau$: all martingales in the set $\mathcal{M}$ isolated in definition 3.1 are included in

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

the smallest rectangular set that embeds the set described by (32). That set is far too big to pose a meaningful max-min decision problem. Here the rectangular expansion that Epstein and Schneider use to acquire dynamic consistency for Gilboa and Schmeidler max-min preferences achieves a Pyrrhic victory by rendering the worst-case model uninteresting to the decision maker and disabling the Good (1952) plausibility recommendation.

8 Conclusion

Our continuous-time formulation (31) exploits mathematically convenient properties of a Brownian information structure. A discrete-time version starts from a baseline model cast in terms of a nonlinear stochastic difference equation. Counterparts to structured and unstructured models play the same roles that they do in the continuous time formulation described in this paper. In the discrete time formulation, preference orderings defined in terms of continuation values are dynamically consistent.

In both the continuous time and discrete time settings, there are compelling reasons for the decision maker not to think that a rectangular set of structured probability models describes the entire set that concerns him. The set of structured models is either *too small* to include potential misspecifications because it excludes statistically nearby unstructured models (again see the quotation above by Box), or it is *too vast* to lead to plausible decision problems in the sense of Good (1952) because it includes models that are statistically very implausible. Therefore, we find it natural for the decision maker to adopt the framework of the present paper to include concerns about unstructured models that satisfy a penalty on entropy relative to the set of structured models, the same type of statistical neighborhood routinely applied to construct probability approximations in computational information geometry.[21]

While we do not explore the issue here, we suspect that the tension between admissibility and dynamic consistency that sometimes emerges in the setup of this paper is also present in other approaches to ambiguity and misspecification, including ones proposed by Hansen and Sargent (2007) and Hansen and Miao (2018).

A purpose of this research is to provide a framework for analyzing the consequences of long-term uncertainties in macroeconomic growth coming from rates of technological progress, climate change, and demographics. Such uncertainties confront private and public decision makers.

[21] See Amari (2016) and Nielsen (2014).

A Computing relative entropy

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 the second-order differential equation (19) appropriately specialized to become:

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

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

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

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

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

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

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

Thus,

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

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

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

Thus,

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

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,[22]

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

[22]We could also have derived this same formula by computing the expectation of $\frac{|\tilde{\eta}(Z_t)|^2}{2}$ under the perturbed distribution.

References

Amari, Shunichi. 2016. *Information Geometry and its Applications*. Japan: Springer.

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

———. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection. *Journal of the European Economic Association* 1 (1):68–123.

Box, Goerge E. P. 1979. Robustness in the Strategy of Scientific Model Building. In *Robustness in Statistics*, edited by Robert L Launer and Graham N. Wilkinson, 2009–236. Academic Press.

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

Chen, Zengjing, Tao Chen, and Matt Davison. 2005. Choquet Expectation and Peng's g -expectation. *Ann. Probab.* 33 (3):1179–1199.

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.

Fan, K. 1952. Fixed Point and Minimax Theorems in Locally Convex Topological Linear Spaces. *Proceedings of the National Academy of Sciences* 38:121–126.

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

Geweke, John and Gianni Amisano. 2011. Optimal prediction pools. *Journal of Econometrics* 164 (1):130 – 141. Annals Issue on Forecasting.

Gilboa, Itzhak and David Schmeidler. 1989. Maxmin Expected Utility with Non-unique Prior. *Journal of Mathematical Economics* 18 (2):141–153.

———. 2001. *A Theory of Case-based Decisions*. Cambridge University Press.

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

Hansen, Lars Peter. 2012. Dynamic Valuation Decomposition Within Stochastic Economies. *Econometrica* 80 (3):911–967.

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

Hansen, Lars Peter and Jianjun Miao. 2018. Aversion to Ambiguity and Model Misspecification in Dynamic Stochastic Environments. *Proceedings of the National Academy of Sciences* 115 (37):9163–9168.

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.

———. 2019. Macroeconomic Uncertainty Prices when Beliefs are Tenuous. SSRN Working Paper No. 2888511.

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.

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.

Nielsen, Frank. 2014. *Geometric Theory of Information*. Heidelberg and New York: Springer.

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.

Waggoner, Daniel F. and Tao Zha. 2012. Confronting model misspecification in macroeconomics. *Journal of Econometrics* 171 (2):167 – 184. Bayesian Models, Methods and Applications.

Wald, A. 1950. *Statisical Decision Functions.* New York: John Wiley and Sons.

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