Structured Uncertainty and Model Misspecification*

Lars Peter Hansen† Thomas J.Sargent‡

December 18, 2018

Abstract

An ambiguity averse decision maker evaluates plans under a restricted family of what we call *structured* models and *unstructured* alternatives that are statistically close to them. The structured models can 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, Tim 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. 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. 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 a set of *structured models* plus mixtures of them.[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 the structured models from suspicions that the structured models are misspecified. Our decision maker expresses his specification suspicions about the structured models by evaluating plans under unstructured alternatives that approximate the structured models in terms of their statistical fits.

Thus, we construct a set of probability models in two steps. First, we specify a set of *structured* probability models that have either fixed or time-varying parameters. Second, we add *unstructured* models that are statistically near one of the structured models. The *unstructured* models can be nonparametric and described incompletely in the sense that they are required only to reside within a statistical neighborhood of the set of structured models. We use relative entropy to measure statistical proximity. 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 tradition of either the Cowles commission or rational expectations 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 Gilboa's conjectured 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

We formalize aversions to the two components of uncertainty – uncertainty about a prior over structured models and uncertainty about the specification of the structured models themselves – in terms of a variational extension of max-min preferences that was conceived by Maccheroni et al. (2006a,b). Our use of dynamic variational preferences here constitutes a substantial extension of Hansen and Sargent (2001) and Anderson et al. (2003). Macro-finance applications of our new framework bring new sources of variations in resource allocations and valuations. As we describe 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 the decision maker's set of probability specifications. Working in continuous time environments with Brownian motion information structures gives us a convenient way to represent positive martingales, as we review in section 3. We use martingales twice: first, when we form a set of structured models with a recursive structure suggested by Chen and Epstein (2002); and second, when we add probabilities associated with unstructured models that are hard to distinguish from the structured models by applying statistical methods to limited data.

To represent and assess potential misspecifications, we draw on insights from robust control theory and from some of our earlier work. This leads us to use relative entropy measures of statistical neighborhoods both to construct families of structured models and to explore misspecifications of those structured models, as we describe in section 4.

To be concrete, we have added inessential auxiliary assumptions that we find helpful and enlightening. Important aspects of our analysis extend to more general settings. Thus, direct extensions of the framework presented here relax the Brownian information structure and do not use relative entropy to constrain the family of structured models.

2 Decision theory components

Our formal approach to decision making under uncertainty presents a practical response to unavoidable tensions among three attractive properties to impose on preferences, namely, (i) dynamic consistency, (ii) the statistical decision-theoretic notion of *admissibility*, and (iii) concerns about model misspecifications. Since we are interested in intertemporal decision problems, we like preferences with recursive structures that impose dynamic consistency. In

methods that would be required to construct better statistical approximations that might be provided by an unstructured model.

addition, to judge the plausibility of his quantitative models, we want our decision maker to verify their admissibility by inspecting implied worst-case probabilities. Within the confines of the max-min formulation of Gilboa and Schmeidler (1989), we describe settings in which a tension between dynamic consistency and admissibility emerges and also provide substantively interesting situations in which it does not. Concerns about misspecification lead us to explore a rich range of model uncertainties that we want our decision maker to fear. We display interesting situations in which a decision maker's preferences can't be dynamically consistent except in degenerate and special cases if we adhere to the max-min utility formulation of Gilboa and Schmeidler (1989). Motivated by this unfortunate outcome, we demonstrate how to construct preferences that reconcile dynamic consistency and admissibility by using a version of the variational prefererences of Maccheroni et al. (2006a,b). The following paragraphs tell the logic that takes us to 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 wants to evaluate *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 are free to interpret the mixing distribution as a set of possible priors over the structured models. An admissible decision rule is one that cannot be weakly dominated by another decision rule for all $\theta \in \Theta$ and be strictly dominated for some $\theta \in \Theta$.

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 unique plan solves the time 0 decision problem; (ii) for each $t > 0$, the time t continuation of the optimal plan for the time 0 problem is the unique solution of the time t continuation problem. An optimal plan with properties (i) and (ii) is said to be dynamically consistent. A Bayesian decision maker completes the probability specification by choosing a unique prior over the set of structured models. The choice of prior typically influences the dynamically consistent optimal plan.

A "robust Bayesian" evaluates plans under a nontrivial set of priors. A popular representation of ambiguity aversion is max-min decision theory, where the minimization is over mathematical expectations of utilities of plans associated with the alternative priors. Applications of the max-min decision theory axiomatized by Gilboa and Schmeidler (1989) often produce decision processes that are supported by a Bayesian prior and are admissible. This outcome can be established by verifying that the Minimax Theorem justifies exchanging orders of maximization and minimization.[3] After exchanging orders of extremization, the outcome of the "outer" minimization is a "worst-case prior" for which the decision process choice is "optimal" in a Bayesian sense. Good (1952) refers to this prior in the above quote. Admissibility and dynamic consistency under this prior follow, for reasons discussed in previous paragraphs. Computing and characterizing the plausibility of this worst-case prior are important parts of a robust Bayesian analysis.

While the max-min decision rule is dynamically consistent under the worst-case prior, it may not be dynamically consistent from the perspective of a max-min decision theory with multiple priors, a key finding of Epstein and Schneider (2003). It is useful here to probe aspects of decision problems that can keep max-min preferences dynamically consistent, and aspects that can make them dynamically inconsistent.

Epstein and Schneider (2003) modified the max-min utility framework in the following way in order to make it dynamically consistent. Temporarily consider a discrete-time setting in which ϵ is the gap in time for a continuous-time approximation. Start with a family of probabilities created by imposing a set of priors over the set of structured models. At date t, consider all possible probability assignments for events on $\mathfrak{F}_{t+\epsilon}$ conditioned on $\mathfrak{F}_t$ implied by alternative choices of time 0 priors. Define preferences using continuation values. For a date $t+\epsilon$ continuation value that is $\mathfrak{F}_{t+\epsilon}$ measurable, minimize over all possible expectations conditioned on $\mathfrak{F}_t$ and construct a date t continuation value that is $\mathfrak{F}_t$ measurable. This backward construction incorporates both subjective discounting of the future and adding a contribution from current period utility.

We formalize this construction by expanding the original set of probabilities and applying the max-min utility framework of Gilboa and Schmeidler (1989). The expansion procedure adds probabilities constructed from all possible t to $t+\epsilon$ conditional probabilities, including ones that come from distinct date zero priors. 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 advocating this expanded set of probabilities as an appropriate object to input into 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.

[3] See Fan (1952).

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 make an enlarged set (i.e., their P) that is rectangular. In this way, the recursive probability set augmentation procedure constructs dynamically consistent preferences by adding possibly uninteresting probability measures to the set of probabilities that originally interested the decision maker. By so doing it can induce an inadmissible decision process with respect to the decision maker's original set of structured models, where "original" means "before the set of models has been embedded in a larger, rectangular set". Thus, although the Minimax Theorem may still be applicable, it now applies to the Epstein and Schneider-expanded set of probabilities. The resulting worst-case probability may or may not be one that is associated with a single prior over the 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 in general, the worst-case probability can be in the expanded set used to achieve a rectangular embedding and not be a member of the set of models that interest the decision maker. As a consequence, assessment of the resulting decision process can fail to satisfy the plausibility criterion advocated by Good (1952). In that case, the rectangular embedding 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. This outcome presents an applied economic model builder a difficult choice between dynamic consistency and admissibility.

Our paper studies two classes of examples that explore aspects of the tension between dynamic consistency and admissibiliy. In one class, a rectangular specification is justified on subjective grounds by positing time variation in parameters. In a continuous-time formulation that can be viewed as a limit associated with 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 can't live together because the huge augmentation required to construct the rectangular embedding vitiates admissibility. Expanding the set of structured models to include relative entropy neighborhoods while preserving rectangularity requires adding a vast number of statistically implausible models to the set subject only to some weak absolute continuity restrictions over finite intervals of time. This 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 those in which the decision maker is concerned about model misspecification.

In applications, we want to work with this second class of environments. Because we want preferences that are both dynamically consistent and entertain a large set of alternative models, we relax 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 counterpart to their formulation provided by Maccheroni et al. (2006b). To express the view that models are flawed but useful simplifications, we apply relative entropy measures of statistical neighborhoods that include what we call "unstructured models" around each of the structured models. We show how to use entropy penalties to capture a practical response to concerns 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).

To explore misspecification, our decision maker enlarges the set of potential models beyond the structured ones. But he does this in a substantially different way than does 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 subjected to a penalty on entropy. With this large set of models, it might be argued that admissibility is a less interesting concept because the decision maker does not articulate any precise details about the potential misspecifications, saying only that there is a vast collection of statistically nearby models. However, in section 7 we show that within our new framework, judging the plausibility of a worst-case model, as proposed by Good (1952), remains attractive and workable. We do this by imputing a worst-case structured model and a worst-case unstructured adjustment to that model, both of which are interesting to the decision maker in terms of the procedure of Good (1952). We also describe and compute an implied statistical discrepancy that tells how concerned the decision maker is about misspecification under alternative values for penalty parameters.

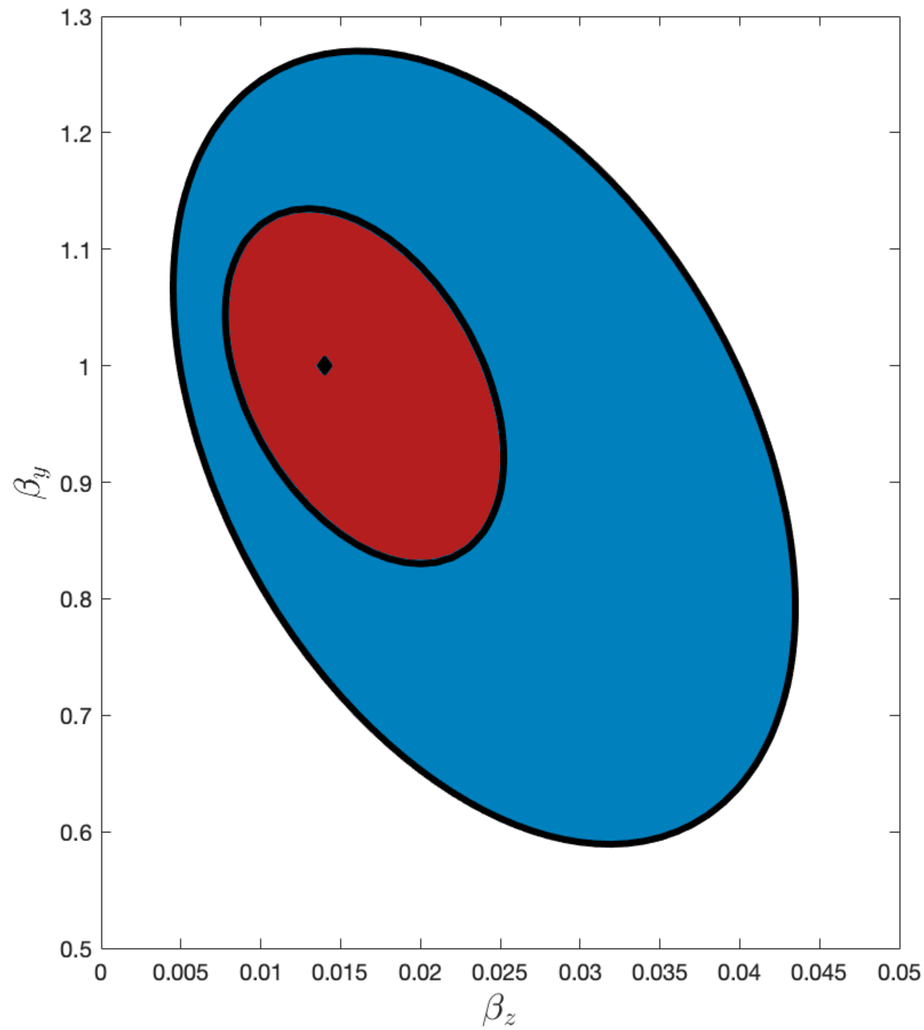

Figure 1: Parameter contours for (β_y, β_z) holding relative entropy and σ_z fixed. The outer curve depicts $\mathsf{q}_{s,0} = .1$ and the inner curve $\mathsf{q}_{s,0} = .05$ where relative entropy equals $\frac{1}{2}\left(\mathsf{q}_{s,0}\right)^2$. The small diamond depicts the baseline model.

As an illustration, in section 5.3 we will consider a model in which an investor is ambiguous about two parameters, namely, a parameter β_y that captures exposure to random macroeconomic growth, and a parameter β_z that captures the persistence in macroeconomic growth rates. Larger values of β_z are associated with less temporal dependence. Figure 1 gives a point associated with baseline parameter values and two regions corresponding to different magnitudes 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 prior information implies that the parameters reside in one of the convex regions determined by relative entropy portrayed in figure 1 and is captured by the set of all priors over one of the convex regions. 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 can expand the original set of priors, in particular, by constructing a rectangular embedding of the original set by entertaining a different date zero prior at each instant and using the implied local transition for the state dynamics to evaluate conditional expected utilities. Unfortunately, the decision plan that emerges from this procedure might be inadmissible if the decision maker targets only time-invariant parameter models. We can get admissibility, however, by adding more structured models to the decision maker's original set of models, including ones that allow parameters to vary over time in a flexible way and embracing a dynamic version of max-min expected utility. But, as we will show, expanding an original set to make it rectangular implies a degenerate and uninteresting decision problem when the decision maker's also explores model misspecification by including unstructured models residing within a relative entropy neighborhood of the set of structured models. We show how to overcome this problem by adapting variational preferences.

In Hansen and Sargent (2018), we apply our approach to a macroeconomic model that we use to explore consequences of ambiguity for asset valuations. In the spirit of Good (1952), we show that the worst-case model is statistically plausible and that it has interesting behavior consequences. The worst-cast model displays more growth rate persistence than does a baseline model when macroeconomic growth is sluggish and conversely less persistence when macroeconomic growth is fast. This pattern induces novel source of nonlinearities in responses of asset valuations to shock exposures.

3 Models and 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 he 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 likelihood ratios. Following Hansen et al. (2006), we represent a likelihood ratio 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]Applications typically use Markov specifications, but a Markov formulation is not essential. It could be generalized to allow other stochastic processes that can be constructed as functions of a Brownian motion information structure.

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

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 $E^U[\ \cdot\ |X_0]$. The martingale property of the process M^U ensures that successive 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, 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 (also called a local mean) $\frac{1}{2} M_t^U |U_t|^2$. Write the conditional mean of $M^U \log M^U$ in terms of a history of local means[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:

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

A notion of relative entropy appropriate for stochastic processes is

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

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 shall sometimes use Abel averages with a discount rate equal to the subjective rate that discounts expected utility flows. With that in mind, we

[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] There exists a variety of sufficient conditions that 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.

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 of martingales M^S with respect to a baseline probability associated with model (1) using an abstract specification suggested by 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. 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 two examples of sets of structured models formed according to a particular version of (10). Subsection 5.1 starts with a parametric family and then entertains time varying parameters while subsection 5.2 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 three 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 three examples differ in their processes $\{\Gamma_t\}$ in (10).

5.1.1 Invariant models

Each S^j process represents a probability assignment for all $t \geqslant 0$. Linear combinations of S_t^j's generate the following set of time-invariant parameter models:

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

where the unknown parameter vector is $\pi = \begin{bmatrix} \pi^1 & \pi^2 & \ldots & \pi^n \end{bmatrix}' \in \Pi_0$, a closed convex subset of $\mathbb{R}^n$. Alternative π's are potential initial period priors across models.

For updating each of the priors, 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 probability $\pi_0^j \geqslant 0$ be assigned to model S^j, where $\sum_{i=1}^n \pi_0^j = 1$. A martingale

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

characterizes a mixture of S^j models. The mathematical expectation of M conditioned on date zero information equals unity. Martingale M evolves as

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

where π_t^j is the date t posterior

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

and the associated drift distortion of martingale M is

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

Let

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

Recall that there is a positive martingale associated with each process S which implies a change in probability measure. In posing max-min utility the decision maker could solve a date zero minimization problem by choice of the initial prior π_0. The family of probabilities indexed by the choice of initial prior will not be rectangular. On the other hand, the max-min utility may deliver admissibility in the plans by standard arguments and may allow for the application of Good's test.

We can imbed this family of probabilities in a rectangular specification. Let Π_t denote the corresponding set of date t posteriors and form:

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

and constructing alternative processes S by alternative selections of $S_t \in \Gamma_t$. In this construction, we index the conditional probabilities by a process of potential posteriors π_t no longer tied to a single prior π_0. Thus more probabilities are entertained other than ones implied by the initial robust Bayesian formation based on an unknown prior. Admissibility relative to the initial set of models does not necessarily follow. These two constructions of potential S processes illustrate the tension between admissibility and dynamic consistency within the Gilboa and Schmeidler (1989) max-min utility framework.

5.1.2 Time-varying regimes

We now alter the previous set by including a time-varying regime specification with uncertain probabilities. Change $\mathcal{M}^o$ in (11) to:

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

where the time-varying parameter vector $\tilde{\pi}_t = \begin{bmatrix} \tilde{\pi}_t^1 & \tilde{\pi}_t^2 & \dots & \tilde{\pi}_t^n \end{bmatrix}'$ has realizations confined to Π_o, the same convex subset of $\mathbb{R}^n$ in (11). The potential realizations for $\tilde{\pi}_t$ could be coordinate vectors.

The decision maker does not directly observe $\tilde{\pi}_t$. Given a fully specified probabilistic structure for $\{\tilde{\pi}_t : t \geqslant 0\}$ the decision maker would have the incentive to compute the mathematical expectation of $\tilde{\pi}_t$ conditional on date t information, which we denote π_t. For instance, one possibility might be to entertain a Markov regime switching specification where a regime corresponds to one of the potential models. If the probabilistic structure for the regimes is unknown, then the decision maker entertains any positive weighting scheme with nonnegative weights that sum to one. In this setting Γ_t in (10) becomes

$$\Gamma_t = \left\{ S_t = \sum_{j=1}^{n} \pi_t^j S_t^j,\ \pi_t^j \geqslant 0,\ \sum_{i=1}^{n} \pi_t^j = 1,\ \pi_t \text{ is } \mathfrak{F}_t \text{ measurable} \right\}$$

The set of models associated with this specification is rectangular with no need for augmentation. Admissibility can be established under familiar conditions and Good's test can be applied. While this example presumes time varying regimes, similar logic applied to time varying parameter specifications.

5.1.3 Pools of models

Concerned about misspecifications, Geweke and Amisano (2011) propose an approach for using a finite pool of models to predict in which weights over models are not determined by Bayesian model averaging. Instead, a suspicion that all models within the pool are misspecified leads 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 with 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.

Our decision maker shares these authors' concerns about model misspecification but proceeds differently by expressing his concerns formally in terms of sets of unstructured 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 potential model weights. Γ_t can be taken to be the set of all potential nonnegative weights across models that sum to one. A decision problem determines weights that vary over time in a way designed to manage concerns about and aversion to model misspecifications. To employ Good's 1952 criterion, the decision maker must view a weighted average of models as a substantively interesting possible misspecification.[11]

In the next subsection, we shall describe another way to construct a set $\mathcal{M}^o$ of martingales that determine structured models surrounding a baseline model.

5.2 A family of structured models restricted by relative entropy

We now show how to construct the 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 M^S to be

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

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 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 use Q to evaluate relative entropy by computing:

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

[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.

We represent 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 $\mathcal{A}^s$ of transitions under the M^S probability is the second-order differential operator

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

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

$$\int \mathcal{A}^s \rho dQ = 0. \tag{12}$$

Equation (12) can be derived heuristically by applying the Law of Iterated expectations.

We can apply characterization (12) of a stationary distribution to obtain the following second-order differential equation whose solution gives relative entropy implied by the drift distortion η:

$$\mathcal{A}^s \rho = \frac{\mathsf{q}^2}{2} - \frac{|\eta|^2}{2}, \tag{13}$$

Relative entropy is $\varepsilon(M^S) = \frac{\mathsf{q}^2}{2}$ and q measures the magnitude of drift distortion η. To compute relative entropy associated with a process defined by generator $\mathcal{A}^s$, we can solve equation (13) simultaneously for q and the function ρ. The right side of (13) must have mean zero under the stationary distribution Q, an outcome that justifies our interpretation of $\frac{\mathsf{q}^2}{2}$ as relative entropy. The function ρ is well defined only up to translation by a constant. This approach to computing relative entropy has direct extensions to Markov jump processes and mixed jump diffusion processes. For diffusion processes, equation (13) is a special case of a Feynman-Kac equation.[12]

To entertain a family of η's, we use relative entropy as a constraint on η or more generally on S. Here we have to be careful. An approach that won't work for us would be to specify only q and then to find all η, or more generally, all S that satisfy $\varepsilon(M^S) \leqslant \frac{\mathsf{q}^2}{2}$. We avoid this approach because it produces a family of probabilities that fails to satisfy

[12]Had our interest been to compute discounted relative entropy, equation (13) 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.

an instant-by-instant constraint $S_t \in \Gamma_t$ for all $t \geqslant 0$ in (10) for some collection of convex sets $\{\Gamma_t : t \geqslant 0\}$. Furthermore, enlarging this set of probabilities to make it rectangular as Epstein and Schneider recommend would yield a set of probabilities that is much too large for max-min preferences, as we shall describe in detail in section 7.2.

Imposing that $|S_t| \leqslant \mathsf{q}$ would be too limiting because, at least in some examples, it would exclude coefficient uncertainty. So to incorporate additional state dependence we use a low frequency refinement to relative entropy. Thus, for $S_t = \eta(X_t)$ consider the log-likelihood-ratio process

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

Following Hansen (2012) and others, decompose L_t as

$$L_t = \frac{\mathsf{q}^2}{2} t + \int_0^t \left[\left(\frac{\partial \rho}{\partial x}(X_\tau) \right)' + \nu(X_\tau) \right] \cdot dW_\tau^S - \rho(X_t) + \rho(X_0),$$

where the second term is a martingale under the probability measure induced by M^S and ρ satisfies (13).[13] Over long horizons, the first term dominates and becomes a time trend with a coefficient equal to relative entropy $\frac{\mathsf{q}^2}{2}$. Assuming that X is stochastically stable and taking expectations under the measure implied by M^S gives

$$\begin{aligned} \lim_{t\to\infty} E\left(M_t^S L_t | X_0 = x\right) - \frac{\mathsf{q}^2}{2} t &= \lim_{t\to\infty} \frac{1}{2} E\left(M_t^S \int_0^t \left[|\eta(X_{t+\tau})|^2 - \frac{\mathsf{q}^2}{2} \right] d\tau \mid X_0 = x \right) \\ &= \rho(x) - \int \rho dQ, \end{aligned}$$

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

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

Thus, $\rho - \int \rho dQ$ provides a long-horizon first-order refinement of relative entropy.

To specify a set of models of potential interest to a decision maker, we construct a

[13]See the discussion in section three of Hansen (2012), which includes the formal result and earlier closely related references.

"boundary" by imposing both q and ρ where $s = \eta(x)$ is restricted by

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

Different choices of $s = \eta(x)$ that satisfy this equality have the same relative entropy q and the same first-order refinement ρ. Notice that we need to specify ρ only up to a translation. Many alternative probability models have the same relative entropy and the same first-order refinement. We include convex combinations of η's on the boundary by imposing the restriction

$$\Gamma_t = \left\{ s : \mathcal{A}^s(X_t)\rho \leqslant \frac{\mathsf{q}^2}{2} - \frac{|s|^2}{2} \right\}, \tag{14}$$

where we use the properties that $\mathcal{A}^s$ is linear in s and $\frac{|s|^2}{2}$ is concave in s. The inequality on the right-hand side of (14) allows us to include other models as well including ones with time-varying parameters. One way to implement this approach is to posit alternative drift configurations associated with specifications of $\eta(x)$ and then to solve (13) for ρ and q.

To explore further implications of residing in the set Γ_t defined in (14), let $\widetilde{S}_t = \tilde{\eta}(X_t)$ where $\widetilde{S}_t = s$ in Γ_t. Integrating the negative of both sides of the inequality used in the definition of Γ_t in (14) implies

$$\frac{1}{2} E\left[M_t^{\widetilde{S}} \int_0^t |\tilde{\eta}(X_\tau)|^2 d\tau \mid X_0 = x \right] - \frac{\mathsf{q}^2}{2} t \leqslant -E\left[M_t^{\widetilde{S}} \rho(X_t) \mid X_0 = x \right] + \rho(x).$$

Bound 5.1. *Let $\widetilde{S}_t = \tilde{\eta}(X_t)$ and let X be stochastically stable under the probability measure implied by the martingale $M^{\widetilde{S}}$ with limiting stationary distribution $\widetilde{Q}$. If $\widetilde{S}_t = s$ is in Γ_t, then*

$$\lim_{t\to\infty} \left(E\left[M_t^{\widetilde{S}} |\tilde{\eta}(X_t)|^2 \mid X_0 = x \right] - \frac{\mathsf{q}^2}{2} t \right) \leqslant \rho(x) - \int \rho d\widetilde{Q}. \tag{15}$$

Dividing by t, entropy bound (15) implies

$$\frac{1}{t} E\left[M_t^{\widetilde{S}} |\tilde{\eta}(X_t)|^2 \mid X_0 = x \right] - \frac{\mathsf{q}^2}{2} \leqslant 0.$$

In summary, we have i) shown how to use relative entropy along with additional *a priori* information to construct a family of structured models; and ii) conveyed precisely the sense in which we can use a restriction on relative entropy to expand the set of models that concern the decision maker. 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.

5.3 Example

In this subsection, we offer an example of a set $\mathcal{M}^o$ for structured models that can be constructed by the approach either of subsection 5.1 or of subsection 5.2. We start with a baseline parametric model for a representative investor's consumption process, then form a family of parametric structured probability models. We deduce the pertinent version of the second-order differential equation (13) to be solved for computing ρ. The baseline model is

$$\begin{aligned} dY_t &= .01\left(\widehat{\alpha}_y + \widehat{\beta}_y Z_t\right)dt + .01\sigma_y \cdot dW_t \\ dZ_t &= \left(\widehat{\alpha}_z - \widehat{\beta}_z Z_t\right)dt + \sigma_z \cdot dW_t. \end{aligned} \tag{16}$$

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

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

We focus on the following collection of structured parametric models:

$$\begin{aligned} dY_t &= .01\left(\alpha_y + \beta_y Z_t\right)dt + .01\sigma_y \cdot dW_t^S \\ dZ_t &= \left(\alpha_z - \beta_z Z_t\right)dt + \sigma_z \cdot dW_t^S, \end{aligned} \tag{17}$$

where W^S is a Brownian motion and (6) continues to describe the relationship between the processes W and W^S. Collection (17) nests the baseline model (16). Here $(\alpha_y, \beta, \alpha_z, \kappa)$ are parameters that distinguish the structured models (17) from the baseline model, and (σ_y, σ_z) are parameters common to models (16) and (17).

We represent members of a parametric class defined by (17) in terms of our section 3.1 structure with drift distortions S of the form

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

then use (1), (6), and (17) to deduce the following restrictions on η_1:

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

where

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

Given an η that satisfies these restrictions, we compute a ρ that for this example will be quadratic and will depend only on z ($\rho(x) = \widetilde{\rho}(z)$). Relative entropy $\frac{\mathsf{q}^2}{2}$ emerges from the solution to differential equation (13) appropriately specialized:

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

Under parametric alternatives (17), the solution for ρ is quadratic in $z - \bar{z}$. Write:

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

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

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

We compute an example in which $\rho_1 = 0$ and ρ_2 satisfies:

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

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

While Figure 1 displays contours of time-invariant parameters with the same relative entropies, the parameters are allowed to vary over time provided that they remain within appropriate regions. Indeed, we use (10) as a convenient way to build a set of structured models that include ones with time varying parameters that lack probabalistic 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 the implied 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 a dynamic consistency axiom of Epstein and Schneider (2003) that among other things 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 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.

5.4 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 notion of ambiguity that is rich enough to allow for uncertainty in conditional volatilities of Brownian increments.[14] Alternative probability specifications there fail to be absolutely continuous (over finite time intervals), and standard likelihood ratio analysis does not apply. This approach would push us outside the Chen and Epstein (2002) formulation but would still let us employ a rectangular embedding. Provided that we impose date specific constraints on the conditional variances, we could extend the construction of the set of structured models to accommodate volatility uncertainty.[15]

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 $\theta > 0$, we define a scaled discrepancy of martingale M^U from a set of

[14] 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.

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

martingales $\mathcal{M}^o$ as

$$\begin{aligned}\Xi(M^U|\mathfrak{F}_0) &= \xi \inf_{M^S \in \mathcal{M}^o} \Delta\left(M^U; M^S | \mathcal{F}_0\right) \\ &= \frac{\xi\delta}{2}\int_0^\infty \exp(-\delta t) E\left[M_t^U \gamma_t(U_t) \middle| \mathfrak{F}_0\right] dt. \end{aligned} \tag{18}$$

where

$$\gamma_t(U_t) = \inf_{S_t \in \Gamma_t} |U_t - S_t|^2, \tag{19}$$

and $S(U_t)$ is the infimum of the right side. 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 much we penalize the expected utility minimizer for choosing unstructured models that are statistically close to 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 $\Xi(M^U|\mathfrak{F}_0)$ defined in (18).[16]

7 Recursive Representation of Preferences

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

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

7.1 Continuation values

A 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. For Markovian plans, a Hamilton-Jacobi-Bellman (HJB) equation restricts the evolution of continuation values. In particular, for a plan $\{C_t\}$, a continuation value process $\{V_t\}_{t=0}^{\infty}$ is defined by

$$V_t = \min_{\{U_\tau : t \leqslant \tau < \infty\}} E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^U}{M_t^U}\right)\left[\psi(C_{t+\tau}) + \left(\frac{\xi\delta}{2}\right)\gamma_{t+\tau}(U_{t+\tau})\right] d\tau \mid \mathfrak{F}_t\right) \tag{20}$$

where ψ is an instantaneous utility function. We can use (20) 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 first deriving and discussing this inequality and the associated robustness bound, we shall use (20) to present 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 (20) and let $S^o = S(U^o)$ be the minimizing S implied by equation (19). 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.

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

$$\begin{aligned}&\frac{\delta}{2}E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^U}{M_t^U}\right)|S_{t+\tau} - U_{t+\tau}|^2 d\tau \mid \mathfrak{F}_t\right)\\ &\leqslant \frac{\delta}{2}E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^{U^o}}{M_t^{U^o}}\right)|S_{t+\tau}^o - U_{t+\tau}^o|^2 d\tau \mid \mathfrak{F}_t\right)\end{aligned} \tag{21}$$

then

$$\begin{aligned}&E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^U}{M_t^U}\right)\psi(C_{t+\tau}) d\tau \mid \mathfrak{F}_t\right)\\ &\geqslant E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^{U^o}}{M_t^{U^o}}\right)\psi(C_{t+\tau}) d\tau \mid \mathfrak{F}_t\right).\end{aligned} \tag{22}$$

for all $t \geqslant 0$.

Inequality (22) is a direct implication of minimization problem (20). It gives probability specifications that have date t discounted expected utilities that are at least as large as the one parameterized by U^o. All of the structured models 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 (20) implies that

$$\begin{aligned} V_t = \min_{\{U_\tau : t \leqslant \tau < t+\epsilon\}} \Bigg\{ & E\left[\int_0^\epsilon \exp(-\delta\tau)\left(\frac{M^U_{t+\tau}}{M^U_t}\right)\left[\psi(C_{t+\tau}) + \left(\frac{\xi\delta}{2}\right)\gamma_{t+\tau}(U_{t+\tau})\right] d\tau \mid \mathfrak{F}_t\right] \\ & + \exp(-\delta\epsilon) E\left[\left(\frac{M^U_{t+\epsilon}}{M^U_t}\right) V_{t+\epsilon} \mid \mathfrak{F}_t\right]\Bigg\} \end{aligned} \tag{23}$$

for $\epsilon > 0$. Heuristically, we can "differentiate" the right side of (23) 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 (23) is then

$$\begin{aligned} 0 &= \min_{U_t} \left[\psi(C_t) - \frac{\xi\delta}{2}\gamma_t(U_t) - \delta V_t + U_t \cdot \varsigma_t + \nu_t\right] \\ &= \min_{S_t \in \Gamma_t} \min_{U_t} \left[\psi(C_t) + \frac{\xi\delta}{2}|U_t - S_t|^2 - \delta V_t + U_t \cdot \varsigma_t + \nu_t\right] \\ &= \min_{S_t \in \Gamma_t} \left[\psi(C_t) - \frac{1}{2\xi\delta}\varsigma_t \cdot \varsigma_t - \delta V_t + S_t \cdot \varsigma_t + \nu_t\right] \end{aligned} \tag{24}$$

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 (24) comes from an Ito adjustment to the local covariance between $\frac{dM^U_t}{M^U_t}$ 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, (24) 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* (24)*, we can still discuss admissibility relative* *to a set of structured models using the representation on the third line of* (24)*. 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 C plans for a given S probability. By posing a Markov decision problem, we can apply a Minimax theorem along with a so-called Bellman-Isaacs condition for a dynamic two-person game to study admissibility. 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 (21) *and conclude that a counterpart to inequality* (22) *holds in which we maximize both the right side and the left side by choice of a plan subject to the imposed constraints on the decision problem. Thus, the entropy of U^* relative to S^* tells us over what probabilities we can bound discounted expected utility.*

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 example of variational preferences.*[17] *Their baseline model is a unique structured model, distrust of which motivates a decision maker to compute a worst-case 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 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*

[17]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 (24) 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.

associated with martingales in $\mathcal{M}^o$.

7.2 With relative entropy and rectangularity, anything goes

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

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

$$\overline{\mathcal{M}} = \left\{ M^U \in \mathcal{M} : \Xi(M^U | \mathfrak{F}_0) < \epsilon \right\} \tag{25}$$

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 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.[18] 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:

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

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

$$= \left[\frac{1 - \exp(-\delta h)}{2\delta}\right] \exp(-\delta\tau) E \left(|\overline{U}_\tau|^2 \mid \mathfrak{F}_0 \right).$$

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. We summarize this argument in the following proposition:

Proposition 7.4. *Any rectangular set of probabilities that contains the probabilities induced by martingales in* (25) *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 the smallest rectangular set that embeds the set described by (25). That set is far too big to pose a meaningful max-min decision problem.

8 Conclusion

Our continuous-time formulation (24) 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.[19]

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

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, demographics, and so on. Such uncertainties confront both private decision makers and public policy makers.

A Computing relative entropy

We show how to compute relative entropies for parametric models of the form (17). Recall that relative entropy $\frac{\mathsf{q}^2}{2}$ emerges as part of the solution to the second-order differential equation (13) appropriately specialized to become:

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

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

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

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

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

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,[20]

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

[20]We could also 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 P. and Thomas J. Sargent. 2001. Robust Control and Model Uncertainty. *American Economic Review* 91 (2):60–66.

———. 2007. Recursive Robust Estimation and Control Without Commitment. *Journal of Economic Theory* 136 (1):1 – 27.

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

Hansen, Lars 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. 2018. Macroeconomic Uncertainty Prices when Beliefs are Tenuous. University of Chicago.

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.