Risk, Ambiguity, and Misspecification: Decision Theory, Robust Control, and Statistics

Lars Peter Hansen and Thomas J. Sargent*

May 5, 2022

Abstract

We use the variational preferences of Maccheroni et al. (2006a) in new ways to make contact with statistics and econometrics. We use relative entropy and other statistical divergences as cost functions in a variational preference representation of someone who is *ambiguous* in the sense of not having a unique prior over a discrete set or manifold of statistical models (i.e., likelihood functions) and who suspects that each statistical model is *misspecified*. We connect variational preferences to theories of robust control and statistical approximation.

Keywords— Variational preferences, statistical divergence, relative entropy, prior, likelihood, ambiguity, misspecification

1 Introduction

Much of the theory of decision making under uncertainty in economics is not cast explicitly in terms of the likelihoods and priors that are foundations of statistics and econometrics. Likelihoods are probability distributions conditioned on parameters while priors describe a decision maker's subjective belief about parameters.[1] By distinguishing roles played by likelihood functions and subjective priors over parameters, this paper aims to bring decision theory into closer contact with statistics, econometrics and addresses concerns about possible model misspecifications and how to select a prior that often arise in practice.

Although they proceeded differently than we do, Chamberlain (2020), Cerreia-Vioglio et al. (2013), and Denti and Pomatto (2022) used decision theories to study related issues. Chamberlain (2020) emphasized that likelihoods and priors are both susceptible to uncertainties. Our paper formulates a decision theory that acknowledges both types of uncertainties. Cerreia-Vioglio et al. (2013) and Denti and Pomatto (2022) focused on uncertainty about predictive distributions that are constructed by integrating likelihoods using priors. Since a likelihood describes probabilities over events that are of direct interest to a decision maker

*We thank Simone Cerreia Vioglio, Marco Loseto, Fabio Angelo Maccheroni, Massimo Marinacci, Giulio Principi, Doron Ravid, and participants of the Econometrics Advising Group at the University of Chicago for criticisms of earlier drafts of this paper.

[1]Distinguishing between them is fundamental to Bayesian formulations of statistical learning. See de Finetti (1937) and Savage (1954). Thus, de Finetti (1937) recommended exchangeability as a more suitable assumption than iid (independent and identically distributed) to model a situation in which a decision maker wants to learn. Importantly, de Finetti showed that putting subjective probabilities over parameters of a manifold of likelihood functions for an iid sequence of random variables leads to an exchangeable sequence of random variables.

conditioned on parameters, alternative priors over parameters induce ambiguity about probabilities over such events, a focus for both of these papers.[2] Neither paper sharply distinguishes prior uncertainty from concerns about possible model misspecifications, which is something that we do. We formulate concerns about model misspecification as uncertainty about likelihoods.

Our approach assembles concepts and practical ways of modeling risks and concerns about model misspecifications from statistics, robust control theory, economics, and decision theory. We align definitions of statistical models, uncertainty, and ambiguity with concepts from decision theories that build on Anscombe and Aumann (1963)'s way of representing subjective and objective uncertainties. We connect our analysis to econometrics and robust control theory by using Anscombe and Aumann *states* as alternative parameterized statistical models of random variables which influence outcomes that a decision maker cares about. We do this differently than Gilboa et al. (2010), Cerreia-Vioglio et al. (2013), and Denti and Pomatto (2022) in ways that affect the concerns about robustness and ambiguity which we are able to represent with variational preferences.

Objects and Interpretations

Our decision maker knows what statisticians call a parameterized family of probability distributions $d\tau(w|\theta)$, where $w \in W$ is a realization of a "shock" that he cares about and $\theta \in \Theta$ is a vector of parameters. The decision maker evaluates alternative decision rules $\rho : W \to X$, where $x \in X$ is a "prize" that he cares about. The parameter space Θ can be finite or infinite dimensional; $d\tau(w|\theta)$ is a member of the family of distributions indexed by $\theta \in \Theta$. When Θ is infinite dimensional, we say that $d\tau(w|\theta)$ for $\theta \in \Theta$ is a family of "nonparametric" probability distributions. The "informativeness" of a decision maker's set of possible "prior" probability distributions over Θ plays an important role in justifying alternative approaches to "robustness" that we describe in section 4.

Let $\mathcal{W}$ be a sigma algebra of events expressed in terms of the shocks. A common way to use an Anscombe and Aumann framework has been to let w be a state and then to explore uncertainty about how to assign a probability to the measurable space $(W, \mathcal{W})$. In such a setting, Anscombe and Aumann acts are lotteries that are conditioned on states. Those lotteries are sources of randomness described by known probabilities. Such lotteries facilitate a mathematically convenient construction that starts with a von Neumann and Morgenstern (2004) expected utility representation with known probabilities and then extends it to include subjective uncertainty. Within such a framework, Gilboa et al. (2010) and Cerreia-Vioglio et al. (2013) introduce parameterized models as a family of primitive probabilities that a decision maker cares about, perhaps for epistemological reasons. In effect, Cerreia-Vioglio et al. (2013) consider an expanded state space (w, θ) that includes both shocks w and parameters θ and then take a model to be a conditional distribution over $(W, \mathcal{W})$ given θ. Consistent with the framework of Gilboa et al. (2010), Cerreia-Vioglio et al. show that a family of models induces a partial ordering according to which one act is preferred to another if it is preferred under all models within the family. They restrict subjectively rational preferences to be consistent with the partial ordering. The role of the partial ordering here is reminiscent of the concept of admissibility from statistics.

In contrast to the Cerreia-Vioglio et al. approach, we use "lotteries" to represent distributions over shocks for each parameter and let Anscombe and Aumann (1963) "states" be parameters. This allows us to

[2] Among other contributions, Cerreia-Vioglio et al. (2013) (section 4.2) provide a rationalization of the smooth ambiguity preferences proposed by Klibanoff et al. (2005) based on likelihood-prior distinctions. Denti and Pomatto (2022) extend this approach by using an axiomatic revealed preference approach to deduce an implied parameterization of a likelihood function.

distinguish robustness to model misspecifications from robustness to choice of prior over parameters, thereby letting us pose decision problems in which both types of robustness concerns are present.

Section 2 sets the stage by reviewing axioms that support Anscombe and Aumann's subjective expected utility representation. Section 3 tells how Maccheroni et al. (2006a) relaxed the Gilboa and Schmeidler (1989) and Anscombe and Aumann axioms to arrive at variational preferences. Section 4 describes a class of variational preferences that use statistical divergences as Maccheroni et al. cost functions. Section 5 describes and applies our formulations of variational preferences, with subsections defining cost functions that distinguish concerns about robustness of likelihoods from concerns about robustness of priors. In subsection 5.1, a decision maker has a unique baseline model that he distrusts and seeks robustness with respect to statistically nearby models. In subsection 5.2, a decision maker knows a set of models but seeks robustness with respect to a set of alternative priors to put over those models. After comparing and contrasting them in subsection 5.3, subsection 5.4 modifies the robust prior analysis to be consistent with the axioms posed by Gilboa and Schmeidler (1989) and subsection 5.5 provides an example of these alternative types of robustness. Next, subsection 6 describes a candidate for a cost function to use for variational preferences of a decision maker who is concerned with *both* types of robustness. Section 7 briefly steps outside the decision theory to discuss how an outside analyst might want to assess "cost" parameters that characterize a decision maker's variational preferences. Section 8 concludes.

2 Preliminaries

Following Gilboa and Schmeidler (1989) and Maccheroni et al. (2006a), we adopt a version of the framework of Anscombe and Aumann (1963) described by Fishburn (1970): $(\Theta, \mathfrak{G})$ is a measurable space of potential *states*; $(X, \mathfrak{X})$ is a measurable space of potential *prizes*; Π is a set of probability measures over states; and Λ is a set of probability measures over prizes.[3] For each $\pi \in \Pi$, $(\Theta, \mathfrak{G}, \pi)$ is a probability space and for each $\lambda \in \Lambda$, $(X, \mathfrak{X}, \lambda)$ is a probability space.

Definition 2.1. *An* **act** *is a* $\mathfrak{G}$ *measurable function* $f : \Theta \to \Lambda$.

For a given θ, $f(\theta) \in \Lambda$ is a lottery over possible prizes $x \in X$.[4] We let $df(x \mid \theta)$ denote integration with respect to probabilities described by that lottery. For a given probability measure $\pi \in \Pi$, $\int_\Theta f(dx \mid \theta)\pi(d\theta)$ is a two-stage lottery over prizes, with one lottery over states θ being described by π and another lottery over prizes being described by $df(x \mid \theta)$ that depends on the outcome θ from the other lottery.

Let $\mathcal{A}$ be the set of all acts. Two collections of acts especially interest us, a set $\mathcal{A}_o$ that lets us represent objective uncertainty, another set $\mathcal{A}_s$ that expresses subjective uncertainty. Formally, let $\mathcal{A}_o \subset \mathcal{A}$ denote the collection of all *constant* acts where a constant act maps all $\theta \in \Theta$ into a unique lottery over prizes $x \in X$. Constant acts express objective uncertainty because they do not depend on the parameter θ. Given this lack of dependence, the probability distribution $\pi \in \Pi$ over states plays no role in shaping an ultimate probability distribution over prizes. The collection $\mathcal{A}_s$ consists of acts, each of which delivers a unique prize for each θ. We let $s(\theta) \in X$ denote an act in $\mathcal{A}_s$.[5] We use a probability distribution $\pi \in \Pi$ over states in conjunction with $\mathcal{A}_s$ to express subjective uncertainty.

[3] For a discussion of the Anscombe-Aumann setup, see Kreps (1988), especially chapters 5 and 7.

[4] The basic setup used here borrows from Marinacci and Cerreia-Vioglio (2021). Formulations of max-min expected utility and variational preferences initially worked within a tradition in decision theory under uncertainty that restricted probabilities to be finitely additive, following de Finetti and Savage. However, countable additivity simplifies the presentation and is routinely imposed in much of probability theory.

[5] Technically an act in $\mathcal{A}_s$ is a degenerate Dirac lottery with a mass point at $s(\theta)$ that is assigned probability one.

Remark 2.2. *Anscombe and Aumann (1963) distinguished "horse race lotteries," represented by acts in $\mathcal{A}_s$, from "roulette lotteries," represented by acts in $\mathcal{A}_o$.*[6] *Savage (1954) did not use "objective" lotteries when he rationalized subjective expected utility.*

We shall often construct a new act from initial acts f and g by using: an $\alpha \in (0,1)$ to form a mixture

$$[\alpha f + (1-\alpha)g](\theta) = \alpha f(\theta) + (1-\alpha)g(\theta) \in \Lambda \quad \forall \theta \in \Theta.$$

As mentioned in section 1, we shall interpret objects in the Anscombe and Aumann formulation in ways that relate to our work as applied econometricians. We interpret a state θ as one among a set Θ of statistical models that a decision maker regards as possible. A decision maker takes an action (i.e., "chooses an Anscombe and Aumann act") that leads to a probability distribution over outcomes that he/she cares about, i.e., over Anscombe and Aumann prizes $x \in X$. A decision maker's *prior* over possible statistical models is a probability measure $\pi \in \Pi$. We shall use special cases of these interpretations to describe a) a Bayesian decision maker with a unique prior over a set Θ of statistical models, b) a decision maker who knows a set Θ of statistical models and who copes with *ambiguity* about those models by considering prospective outcomes under a set of priors Π over those statistical models, and c) a decision maker with concerns that a single known statistical model θ is *misspecified* by using a statistical discrepancy measure to limit the exploration of unknown models surrounding that known model.

Remark 2.3. *Though not in ours, in other applications of Anscombe and Aumann, the state is w, and uncertainty is about a probability distribution to assign to the space W. A lottery conditioned on a state adds additional randomness with known probabilities. Preferences are cast in terms of lotteries conditioned on w as well as uncertainty over W.*

2.1 Preferences

To represent a decision maker's preferences over acts, we use $\sim$ to mean indifference, $\succsim$ to mean a weak preference, and $\succ$ to mean a strict preference. Throughout, we assume that preferences are non-degenerate (there is strict ranking between two acts), complete (we can compare any pair of acts), and transitive ($f \succsim g$ and $g \succsim h$ imply $f \succsim h$). We also impose an Archimedean axiom that provides a form of continuity.[7] A *finite signed measure* on the measurable space $(X, \mathfrak{X})$ is a finite linear combination of probability measures that resides in a linear space $\widehat{\Lambda}$ that contains Λ.

2.2 Objective probability

By analyzing preferences over the constant acts $\mathcal{A}_o$, we temporarily put aside attitudes about ambiguity and model misspecification and focus on objective uncertainty (sometimes called "risk"). There is a unique probability $\lambda \in \Lambda$ associated with every act $f \in \mathcal{A}_o$ and a unique act in $\mathcal{A}_o$ associated with every $\lambda \in \Lambda$. We define a restriction $\succ_\Lambda$ of the preference order $\succ$ to the space of constant acts $f \in \mathcal{A}_o$ by

$$\lambda \succ_\Lambda \kappa \iff f \succ g$$

[6] See Kreps (1988, ch. 5) for more about the distinction.

[7] The Archimedean axiom states: let f, g, h be acts in $\mathcal{A}$ with $f \succ g \succ h$. Then there are $0 < \alpha < 1$ and $0 < \beta < 1$ such that $\alpha f + (1-\alpha)h \succ g \succ \beta f + (1-\beta)h$. See Herstein and Milnor (1953, Axiom 2) for an alternative formulation of a continuity axiom.

where λ is the probability generated by act $f \in \mathcal{A}_o$ and κ is the probability distribution generated by act $g \in \mathcal{A}_o$.

To represent preferences $\succ_\Lambda$, we follow Von Neumann and Morgenstern (1944) who imposed the following restriction on preferences:[8]

Axiom 2.4. *(Independence) If $f, g, h \in \mathcal{A}_o$ and $\alpha \in (0,1)$, then*

$$f \succsim g \Rightarrow \alpha f + (1-\alpha)h \succsim \alpha g + (1-\alpha)h.$$

The Von Neumann and Morgenstern approach delivers an expected utility representation of preferences over constant acts: there exists a utility function $u : X \to \mathbf{R}$ such that for $f, g \in \mathcal{A}_o$

$$f \succsim g \iff U(f) \geqslant U(g) \tag{1}$$

where

$$U(f) = \int_X u(x) d\lambda(x) \tag{2}$$

and $\lambda \in \Lambda$ is the probability distribution generated by constant act f. Representation (2) can be extended to a space $\widehat{\Lambda}$ of finite signed measures to produce a linear functional on this space. The structure of the space of finite signed measures brings interesting properties to representation (2). Thus, although u is in general a nonlinear function of prizes, U is a linear functional of finite signed measures $\lambda \in \widehat{\Lambda}$. Consequently, a representation theorem for linear functionals of finite signed measures justifies (2). According to representation (1), for any real number r_0 and strictly positive real number r_1, utility functions $\mathsf{r}_1 u + \mathsf{r}_0$ and u provide identical preference orderings.

2.3 Subjective probability

To construct subjective expected utility preferences, we extend an expected utility representation of $\succ_\Lambda$ on the set of constant acts to a representation of preferences $\succ$ on the set $\mathcal{A}$ of all acts. To do this we impose restrictions on $\succ$ in the form of two axioms. The first extends the independence axiom to the set of all acts:

Axiom 2.5. *(Independence) If $f, g, h \in \mathcal{A}$ and $\alpha \in (0,1)$, then*

$$f \succsim g \Rightarrow \alpha f + (1-\alpha)h \succsim \alpha g + (1-\alpha)h.$$

The second is:

Axiom 2.6. *(Monotonicity) For any $f, g \in \mathcal{A}$ such that $f(\theta) \succsim_\Lambda g(\theta)$ for each $\theta \in \Theta$, $f \succsim g$.*

We first use a Von Neumann and Morgenstern expected utility representation to represent preferences conditioned on each θ. From this conditional representation, we compute

$$\int_X u(x) df(x \mid \theta) = F(\theta)$$

for any act f where

[8] Completeness, transitivity and the Archimedean axiom carry over directly from $\succ$ to $\succ_\Lambda$, but not necessarily non-degeneracy. Our presentation below presumes non-degeneracy of $\succ_\Lambda$.

A set of acts implies an associated collection $\mathcal{B}$ of functions F. From monotonicity axiom 2.6 we know that if f and $\tilde{f}$ imply the same F, then $f \sim \tilde{f}$. Consequently, the preference relation $\succ$ induces a unique preference relation $\succ_\Theta$ for which

$$F \succ_\Theta G \iff f \succ g$$

for acts f and g that satisfy

$$\int_X u(x) df(x \mid \theta) = F(\theta)$$
$$\int_X u(x) dg(x \mid \theta) = G(\theta)$$

A mixture of two acts f and g has expected utility:

$$\int_X u(x)[\alpha df(x \mid \theta) + (1-\alpha) dg(x \mid \theta)] = \alpha F(\theta) + (1-\alpha) G(\theta).$$

If the set of acts $\mathcal{A}$ is convex, then so is the set $\mathcal{B}$ of functions of θ. Furthermore, if $F \sim_\Theta G$, the independence axiom guarantees that for any α the associated convex combinations of F and G are also in the same indifference set of acts. From one indifference set we can build other indifference sets by taking an act h and forming convex combinations with members of the initial indifference set. These observations lead us to seek a utility function that is a linear functional $\mathcal{L}$ on $\mathcal{B}$.

Suppose that $F \geqslant G$ on Θ. The monotonicity axiom implies that $\mathcal{L}(F - G) \geqslant 0$, so $\mathcal{L}$ is a positive linear functional. Under general conditions, a positive linear functional can be represented as an integral with respect to a finite measure.[9] Positive multiples of this linear functional imply the same preference ordering. Since the preference ordering is not degenerate, the measure must not be degenerate. This means that we can make it into a probability measure that we denote $\pi(d\theta)$. We thereby arrive at the following representation of preferences over acts $f \in \mathcal{A}$

$$f \gtrsim g \iff \int_\Theta \left[\int_X u(x) df(x \mid \theta) \right] d\pi(\theta) \geqslant \int_\Theta \left[\int_X u(x) dg(x \mid \theta) \right] d\pi(\theta), \tag{3}$$

where the probability measure π describes subjective probabilities.

Representation (3) lets us interpret the expected utility of an act f with a two-stage lottery. First draw a $\tilde{\theta}$ from π and then draw a prize $x \in X$ from probability distribution $df(x \mid \tilde{\theta})$. By changing order of integration, we can write

$$\int_\Theta \left[\int_X u(x) df(x \mid \theta) \right] d\pi(\theta) = \int_X u(x) \left[\int_\Theta df(x|\theta) d\pi(\theta) \right]$$

or equivalently

$$\int_\Theta \left[\int_X u(x) df(x \mid \theta) \right] d\pi(\theta) = \int_X u(x) d\lambda(x), \tag{4}$$

where

$$d\lambda(x) = \int_\Theta df(x \mid \theta) d\pi(\theta). \tag{5}$$

Equation (5) constructs a single lottery λ over x from the compound lottery generated by $(d\pi(\theta), df(x \mid \theta))$.[10]

[9] The Riesz-Markov-Kakutani Representation Theorem provides such a representation on the space of continuous functions with compact support on a locally compact Hausdorff space.

[10] Equation (5) thus expresses the "reduction of compound lotteries" described by Luce and Raiffa (1957, p. 26)

For a statistician, λ is a "predictive density" constructed by integrating over unknown parameter θ. Let f_c be the constant act with lottery λ defined by the left side of (5) for all $\theta \in \Theta$. Equations (4) and (5) assert that a person with expected utility preferences is indifferent between f_c and f.

2.4 Max-min Expected Utility

To construct a decision maker who has max-min expected utility preferences, Gilboa and Schmeidler (1989) replaced Axiom 2.5 with the following two axioms:

Axiom 2.7. *(Certainty Independence) If $f, g \in \mathcal{A}$, $h \in \mathcal{A}_o$, and $\alpha \in (0,1)$, then*

$$f \succsim g \iff \alpha f + (1-\alpha)h \succsim \alpha g + (1-\alpha)h.$$

Axiom 2.8. *(Uncertainty Aversion) If $f, g \in \mathcal{A}$ and $\alpha \in (0,1)$, then*

$$f \sim g \Rightarrow \alpha f + (1-\alpha)g \succsim f.$$

An essential ingredient of this axiom is the mixing weights α and $1-\alpha$ are known. That can be interpreted as a form of objective uncertainty. Axiom 2.8 asserts a weak preference for mixing with known weights α and $1-\alpha$.

Example 2.9. *Suppose that $\Theta = \{\theta_1, \theta_2\}$ and consider lotteries λ_1 and λ_2. Let act f be lottery λ_1 if $\theta = \theta_1$ and be lottery λ_2 if $\theta = \theta_2$. Let act g be lottery λ_2 if $\theta = \theta_1$ and be lottery λ_1 if $\theta = \theta_2$. Suppose that $f \sim g$. Axiom 2.8 allows a preference for mixing the two acts. If, for instance, $\alpha = \frac{1}{2}$, the mixture is a constant act with a lottery $\frac{1}{2}\lambda_1 + \frac{1}{2}\lambda_2$ that is independent of θ. We think of mixing as reducing the exposure to θ uncertainty. In the extreme case, setting $\alpha = \frac{1}{2}$, for example, completely eliminates effects of exposure to θ uncertainty.*

By replacing Axiom 2.5 with Axioms 2.7 and 2.8, Gilboa and Schmeidler obtained preferences described by

$$f \succsim g \iff \min_{\pi \in \Pi_c} \int_\Theta \left[\int_X u(x) df(x \mid \theta) \right] d\pi(\theta) \geqslant \min_{\pi \in \Pi_c} \int_\Theta \left[\int_X u(x) dg(x \mid \theta) \right] d\pi(\theta) \tag{6}$$

for a convex set $\Pi_c \subset \Pi$ of probability measures. An act $f(\theta)$ is still a lottery over prizes $x \in X$ and, as in representation (1), for each θ, $\int_X u(x) df(x \mid \theta)$ is an expected utility over prizes x. Evidently, expected utility preferences (3) are a special case of max-min expected utility preferences (6) in which Π_c is a set with a single member.

3 Variational preferences

Maccheroni et al. (2006a) relaxed certainty independence Axiom 2.7 of Gilboa and Schmeidler (1989) to obtain preferences with a yet more general representation that they called variational preferences. Maccheroni et al. weakened Axiom 2.7 by positing

Axiom 3.1. *(Weak Certainty Independence) If $f, g \in \mathcal{A}_s$, $h, k \in \mathcal{A}_o$, and $\alpha \in (0,1)$, then*

$$\alpha f + (1-\alpha)h \succsim \alpha g + (1-\alpha)h \Rightarrow \alpha f + (1-\alpha)k \succsim \alpha g + (1-\alpha)k$$

and analyzed further by Segal (1990).

Notice that Axiom 3.1 considers only acts that are mixtures of constant acts that can be represented with a single lottery and acts with degenerate lotteries for each α. This axiom states that altering the constant acts does not reverse the decision maker's preferences, but it imposes the same α when making the stated comparison.

Maccheroni et al. showed that preferences that satisfy the weaker Axiom 3.1 instead of Axiom 2.7 are described by

$$f \succsim g \iff \min_{\pi \in \Pi} \int_{\Theta} \left[\int_{X} u(x) df(x \mid \theta) \right] d\pi(\theta) + c(\pi) \geqslant \min_{\pi \in \Pi} \int_{\Theta} \left[\int_{X} u(x) dg(x \mid \theta) \right] d\pi(\theta) + c(\pi) \tag{7}$$

where, as in representation (1), u is uniquely determined up to a linear translation and c is a convex function that satisfies $\inf_{\pi \in \Pi} c(\pi) = 0$. Smaller convex c functions express more aversion to uncertainty. The convex function c in variational preferences representation (7) replaces the restricted set of probabilities Π_c that appears in the max-min expected utility representation (6). In the special case that the convex function c takes on values 0 and $+\infty$ only, Maccheroni et al. show that variational preferences are max-min expected utility preferences.

4 Scaled statistical divergences as c functions

Scaled statistical divergences give rise to convex c functions that especially interest us. We use such divergences in two ways, one for distributions over $(W, \mathfrak{W})$, another for distributions over $(\Pi, \mathfrak{G})$. Our constructions of statistical divergences are very similar in both cases.

We first consider shock distributions over $(W, \mathfrak{W})$. For a baseline probability τ_o, a *statistical divergence* is a convex function $D(\tau \mid \tau_o)$ of probability measures τ that satisfies

- $D(\tau \mid \tau_o) \geqslant 0$
- $D(\tau \mid \tau_o) = 0$ implies $\tau = \tau_o$

Now let ϕ be a convex function defined over the nonnegative real numbers for which $\phi(1) = 0$ and impose $\phi''(1) = 1$ as a normalization.[11] Examples of such ϕ functions and the divergences that they lead to are

$\phi(m) = -\log(m)$	Burg entropy
$\phi(m) = -4\left(\sqrt{m} - 1\right)$	Hellinger distance
$\phi(m) = m \log(m)$	relative entropy
$\phi(m) = \frac{1}{2}\left(m^2 - m\right)$	quadratic

Take a baseline distribution τ_o over shocks w and represent alternative distributions that are absolutely continuous with respect to it as

$$d\tau(w) = m(w) d\tau_o(w) \tag{8}$$

for relative densities $m \in \mathcal{M}$, where

$$\mathcal{M} \doteq \left\{ m : m(w) \geqslant 0, \int_{W} m(w) d\tau_o(w) = 1 \right\}. \tag{9}$$

[11] Sometimes this is called a ϕ-divergence.

The set $\mathcal{M}$ is convex. To define a scaled statistical divergence, we set

$$D(\tau \mid \tau_o) = \xi \int_W \phi[m(w)] d\tau_o(w),$$

where $\xi > 0$. When $\phi(m) = m \log(m)$ and $\xi = 1$, we obtain relative entropy

$$D_{KL}(\tau|\tau_o) = \int_W m(w) \log[m(w)] d\tau_o(w).$$

If τ is not absolutely continuous with respect to τ_o, we set $D(\tau \mid \tau_o)$ to infinity. Relative entropy is often called Kullback-Leibler divergence.

5 Basic formulation

We associate a probability measure $d\tau(w|\theta)$ parametrized by $\theta \in \Theta$ with a random vector having possible realizations w in the measurable space $(W, \mathfrak{W})$. Consider alternative real valued, Borel measurable functions $\rho \in \Psi$ that map $w \in W$ into an $x \in X$. Think of ρ as a decision rule and $\rho(w)$ as an uncertain scalar prize. For each decision rule ρ, let $d\lambda(x \mid \theta)$ be the distribution of the prize $x = \rho$ that is induced by distribution $d\tau(w|\theta)$ and the decision rule ρ. The distribution of the prize thus depends both on the decision rule $\rho(w)$ and the distribution $d\tau(w|\theta)$.

5.1 Not knowing alternative models

We consider a decision maker who knows a baseline model $d\tau_o$ of W that he suspects is misspecified in ways that he is unable precisely to describe. He can say that the alternative models that he is most worried about are statistically close to his baseline model. The presence of so many statistically nearby models prevents a Bayesian from deploying a weakly informative prior over them, thereby precluding a robust Bayesian approach.[12]

To formalize concerns that $d\tau_o$ is misspecified, we begin by letting state $\theta = m$ be a likelihood ratio that determines an alternative model of W

$$d\tau(w) = m(w) d\tau_o(w),$$

where $m \in \mathcal{M}$ for $\mathcal{M}$ given by (9) and

$$\Theta = \mathcal{M}.$$

We represent the decision maker's ignorance of specific alternative models by proceeding as if there is a potentially infinite dimensional space $\mathcal{M}$ of such models over which it is mathematically impossible to put a "weakly informative" prior distribution. A decision maker's expected utility under distorted model $m d\tau_o$ is

$$\int u[\rho(w)] m(w) d\tau_o(w). \tag{10}$$

Notice how (10) is cast in terms of a single lottery, not a compound lottery that involves a probability π over a state space Θ. The following important technical considerations cause us to proceed in this way.

Remark 5.1. *Specifying a prior over the infinite dimensinal space $\mathcal{M}$ brings challenges associated with all nonparametric methods, including "nonparametric Bayesian" methods. A Bayesian prior on an infinite*

[12] For example, see Berger (1984) for a robust Bayesian perspective.

dimensional space such as $\mathcal{M}$ must be more "informative" than is required in finite-dimensional estimation problems.[13] *A related "informativeness" requirement carries over to families of priors that are absolutely continuous relative to a baseline prior. The decision maker in this subsection does not want to entertain priors that are "too informative." In subsection 5.2, we describe a decision maker who is concerned about a set of models that is small enough to proceed with a "robust Bayesian" approach with priors over those models that are not "too informative."*

To complete a description of preferences, we require a scaled statistical divergence. We consider alternative probabilities over the space $\Theta = \mathcal{M}$. Under this perspective, a probability model or corresponds to a choice of $m \in \mathcal{M}$. The notation m denotes both a relative density and a state (or parameter value.) Form a scaled divergence measure:

$$c(m) = \xi \int \phi[m(w)] d\tau_o(w) \tag{11}$$

where $\xi > 0$ is a real number.

As an alternative starting point, temporarily consider only a finite set $\Theta_f = \{\theta_i : i = 1, 2, \ldots, \ell\}$, where $m_i = \theta_i$ becomes a "state" that represents a particular alternative model of W via $d\tau(w) = m_i(w) d\tau_o(w)$. We attach a prior probability π_i to each $\theta_i \in \Theta_f$ and form a mixture of relative densities:

$$m = \sum_{i=1}^{\ell} \pi_i m_i. \tag{12}$$

Thus to each prior π, construct a predictive distribution $d\tau(w) = m(w) d\tau_o(w)$. A scaled statistical divergence of this finite set of predictive distributions is

$$c_p(\pi) = \xi \int \phi \left(\sum_{i=1}^{\ell} \pi_i m_i(w) \right) d\tau_o(w) = c(m)$$

where c is given by (11).

Because there is a vast set of potential misspecifications, our decision maker is actually concerned about possible alternatives to the baseline model $d\tau_o$ that are represented by a much larger set of probabilities over $\Theta = \mathcal{M}$ of likelihood ratios. We have started with a finite set of potential probabilities $m_i(w) d\tau_o(w)$ only as a way to set the stage for expanding that set. Thus, imagine that we instead impose a prior π on the much bigger set $\Theta = \mathcal{M}$ and then form the predictive density

$$m = \int_{\Theta} \theta d\pi(\theta). \tag{13}$$

while having made sure to specify π so that the m on the left side is in $\mathcal{M}$. Again, we use the formula:

$$c_p(\pi) = c \left[\int_{\Theta} \theta d\pi(\theta) \right]. \tag{14}$$

Because our decision maker considers all possible priors over $\Theta = \mathcal{M}$, convexity of the set $\mathcal{M}$ lets us think of it either as the original parameter space or as a set of relative densities of predictive distributions formed from priors over that parameter space. When we use c described in (11) and c_p in (14) to construct preferences, we need not distinguish a probability model as a density in $\mathcal{M}$ from a predictive density formed with a prior

[13] Sims (2010) critically surveys an extensive statistical literature on this issue. Foundational papers are Freedman (1963), Sims (1971), and Diaconis and Freedman (1986).

over $\mathcal{M}$.[14]

The implied m is all that matters for $c_p(\pi)$, not how it might have been formed as a convex combination similar to (13) of members of set set $\mathcal{M}$. This means that it is natural to use (11) as a scaled statistical divergence to represent a decision maker's concern about misspecification of $d\tau_o$.

Remark 5.2. *By applying a Gilboa and Schmeidler (1989) representation of ambiguity aversion to a decision maker who has multiple predictive distributions, Cerreia-Vioglio et al. (2013) forge a link between ambiguity aversion as studied in decision theory and the robust approach to statistics. They also cast corresponding links in terms of variational preferences.*

Variational preferences that use (10) as expected utility over lotteries and (11) as scaled statistical divergence are ordered by

$$\min_{m\in\mathcal{M}} \left(\int u[\rho(w)]m(w)d\tau_o(w) + \xi \int \phi[m(w)]d\tau_o(w) \right). \tag{15}$$

This formulation lets a decision maker evaluate alternative decision rules $\rho(w)$ while guarding against a concern that his baseline model τ_o is misspecified without having in mind specific alternative models τ. Key ingredients are the single baseline probability τ_o and a statistical divergence over probability distributions $m(w)d\tau_o(w)$.

Remark 5.3. *It is convenient to solve the minimization problem on the right side of* (15) *by using duality properties of convex functions. Because the objective is separable in w, we can first compute*

$$\phi^*(\mathsf{u} \mid \xi) = \min_{\mathsf{m}\geqslant 0} \mathsf{u}\mathsf{m} + \xi\phi(\mathsf{m}) \tag{16}$$

where $\mathsf{u} = u[\rho(w)] + \eta$, m is a nonnegative number, and η is a nonnegative real-valued Lagrange multiplier that we attach to the constraint $\int m(w)d\tau_o(w) = 1$; $\phi^*(\mathsf{u} \mid \xi)$ is a concave function of u.[15] *The minimizing value of m satisfies:*

$$\mathsf{m}^* = \phi'^{-1}\left(-\frac{\mathsf{u}}{\xi}\right).$$

The dual problem to the minimization problem on the right side of (15) *is*

$$\max_{\eta} \int_W \phi^*(u[\rho(w)] + \eta \mid \xi)d\tau_o(w) - \eta. \tag{17}$$

Remark 5.4. *We posed minimum problem* (15) *in terms of a set of probability measures on the measurable space $(W, \mathfrak{W})$ with baseline probability $d\tau_o(w)$. Since the integrand in the dual problem* (17) *depends on w only through the control law ρ, we could instead have used the same convex function ϕ to pose a minimization in terms of a set of probability distributions $d\lambda(x)$ with the baseline being the probability distribution over prizes induced $x = \rho(w)$ with distribution $d\lambda_o(x)$. Doing that would lead to equivalent outcomes. Representations in sections 2 and 3 are all cast in terms of induced distributions over prizes. Because control problems entail*

[14] While this distinction between a model and predictive density is not important for defining static preferences, a model builder with repeated observations will draw a distinction between the two objects. When confronted with a single model that generates the data, Bayesian learning is degenerate. In contrast, when there is prior over a family of models, each of which could generate data, there will be scope for the model builder to update the weights over the alternative model in accord with Bayesian leaning. In dynamic settings posed in Hansen and Sargent (2019, 2021), possible misspecifications vary over time in general ways that render Bayesian learning impossible.

[15] The function $-\phi^*(-\mathsf{u} \mid \xi)$ is the Legendre transform of $\xi\phi(\mathsf{m})$.

searching over alternative ρ's, it is more convenient to formulate them in terms of a baseline model $d\tau_o(w)$, as we originally did in subsection 5.1.

Remark 5.5. *If we use relative entropy as a statistical divergence, then*

$$\phi^*(\mathsf{u} \mid \xi) = -\xi \exp\left(-\frac{\mathsf{u}+\eta}{\xi} - 1 \mid \xi\right)$$

and dual problem (17) *becomes*[16]

$$\max_{\eta} -\xi \int \exp\left[-\frac{u[\rho(w)]+\eta}{\xi} - 1\right] d\tau_o(w) - \eta = -\xi \log\left(\int \exp\left[-\frac{u[\rho(w)]}{\xi}\right] d\tau_o(w)\right). \tag{18}$$

The minimizing m in problem (15) *is*

$$m^*(w) = \frac{\exp\left[-\frac{u[\rho(w)]}{\xi}\right]}{\int \exp\left[-\frac{u[\rho(w)]}{\xi}\right] d\tau_o(w)}. \tag{19}$$

The worst-case likelihood ratio m^* exponentially tilts a lottery toward low-utility outcomes. Bucklew (2004) calls this adverses tilting a statistical version of Murphy's law:

> *"The probability of anything happening is in inverse proportion to its desirability."*

Preferences associated with a relative entropy divergence are often referred to as "multiplier preferences." The preceding construction of multiplier preferences is distinct from constructions provided by Maccheroni et al. (2006a) and Strzalecki (2011). Nevertheless, the Maccheroni et al. axiomatic formulation of variational preferences includes our construction as a special case.

Remark 5.6. *(risk-sensitive preferences) The right side of equation* (18)*, namely,*

$$-\xi \log\left[\int_W \exp\left(-\frac{u[\rho(w)]}{\xi}\right) d\tau_o(w)\right], \tag{20}$$

defines what are known as "risk-sensitive" preferences over control laws ρ. Since logarithm is a monotone function, these are evidently equivalent to Von Neumann and Morgenstern expected utility preferences with utility function:

$$-\exp\left[-\frac{u(\cdot)}{\xi}\right]$$

in conjunction with the baseline distribution τ_o over shocks. Risk-sensitive preferences are widely used in robust control theory.

5.2 Not knowing a prior, I

Unlike subsection 5.1, we now adopt a setting in which a decision maker has a parametrized family of models and a baseline prior distribution over those models. Like the decision maker of Gilboa et al. (2010) and Cerreia-Vioglio et al. (2013), our decision maker has multiple prior distributions because he does not trust the baseline prior. Following Gilboa et al., Cerreia-Vioglio et al. and others, we dub such distrust of a single

[16] See Dupuis and Ellis (1997, sec. 1.4) for a closely related connection between relative entropy and a variational formula that occurs in large deviation theory.

prior "model ambiguity." (We use "fear of misspecificaton" to refer to other concerns analyzed in subsection 5.1.) Here, we describe a static version of what Hansen and Sargent (2019, 2021) call structured uncertainty. "Structured" refers to the particular way that we reduce the dimension of a set of alternative models relative to the much larger set entertained by a subsection 5.1 decision maker. The distribution of the prize again depends both on a decision rule $\rho(w)$ and on a shock vector distribution $d\tau(w|\theta)$. Let Θ be a parameter space, and let π_o be a baseline prior probability measure over models θ. The baseline π_o anchors a set of priors π over which a decision maker wishes to be robust. We describe the set of priors by:

$$\pi(d\theta) = n(\theta)\pi_o(d\theta),$$

where n is in the set $\mathcal{N}$ defined by:

$$\mathcal{N} \doteq \left\{ n \geqslant 0 : n(\theta) \geqslant 0, \int_\Theta n(\theta) d\tau_o(\theta) = 1 \right\}. \tag{21}$$

This specification includes a form of "structured" uncertainty in which all models have the same parametric "structure" but each is associated with a different vector of parameter values.[17] The decision maker is certain about each of the specific models $m = \theta$ in the set but is uncertain about a prior to put over them. The decision maker uses scaled statistical divergence

$$c(\pi) = \xi \int_\Theta \phi\left[n(\theta)\right] d\pi_o(\theta). \tag{22}$$

and has variational preferences ordered by[18]

$$\min_{n \in \mathcal{N}} \int_\Theta \left(\int_W u[\rho(w)] d\tau(w \mid \theta) \right) n(\theta) d\pi_o(\theta) + \xi \int_\Theta \phi[n(\theta)] d\pi_o(\theta). \tag{23}$$

Remark 5.7. *From an appropriate counterpart to dual formulation* (17)*, we can represent variational preferences ordered by* (23) *as:*

$$\max_\eta \int_\Theta \phi^* \left(\int_W u[\rho(w)] d\tau(w \mid \theta) + \eta \mid \xi \right) d\pi_o(\theta) - \eta.$$

Remark 5.8. *(Smooth ambiguity preferences) When statistical divergence is scaled relative entropy, preferences over $\rho(w)$ are ordered by*

$$-\xi \log \left[\int \exp\left(-\frac{\int_W u[\rho(w)] d\tau(w \mid \theta)}{\xi} \right) d\pi_o(\theta) \right]. \tag{24}$$

a static version of preferences that Hansen and Sargent (2007) used to frame a robust dynamic filtering problem. It turns out that these preferences are also a special case of the smooth ambiguity preferences that Klibanoff et al. (2005) justified with a set of axioms different from the ones we have used here. Furthermore, Maccheroni et al. (2006a) and Strzalecki (2011) use this construction to justify "multiplier preferences" in contrast to the approach taken here.[19] *Notice that the robustness being discussed in this subsection is with*

[17] See Hansen and Sargent (2022).

[18] See Theorem 4 of Cerreia-Vioglio et al. (2013) for their counterpart to this representation.

[19] Strzalecki (2011) showed that when Savage's Sure Thing Principle augments axioms imposed by Maccheroni et al. (2006a), the cost functions capable of representing variational preferences are proportional to scalar multiples of entropy divergence relative to a unique baseline prior. The Sure Thing Principle also plays a significant role in

respect to a baseline prior over known models and not with respect to possible misspecifications of those models.

Remark 5.9. *When we formulate the set of priors, as we did to obtain criterion* (24)*, we cannot interpret them as expected utility preferences, unlike the situation is described in remark 5.6.*

5.3 Robustness

It is useful to compare and contrast two approaches to "robustness" that we have taken. The decision maker in section 5.1 explores potential model misspecifications by searching over the entire space $\mathcal{M}$, subject to a penalty on statistical divergence from a baseline model. The decision maker in section 5.2 searches over a small space by starting with a baseline prior over the space $\mathcal{M}$ and considering consequences of misspecifying it. As an application of the section 5.2 approach, we let the state or "parameter vector" be $\ell(w \mid \theta)$ and:

$$d\tau(w \mid \theta) = \ell(w \mid \theta) d\tau_o(w),$$

where $\ell(\cdot \mid \theta) \in \mathcal{M}$ for each $\theta \in \Theta$ and $d\tau_o(w)$ is a baseline distribution. The parameter space $\mathcal{M}$ is potentially infinite dimensional. We want to specify a prior π_o over the parameter space $\mathcal{M}$ that is consistent with a Bayesian approach to "nonparametric" estimation and inference. We represent misspecification of $d\pi_o$ in terms of alternative priors $nd\pi_o$ associated with n's in $\mathcal{N}$. We constrain the set of such priors by penalizing a statistical divergence of $n\pi_o$ relative to π_o. To make this approach work technically, it is necessary to start with a prior distribution π_o that is "informative". Requiring that they be absolutely continuous with respect to $d\pi_o$ limits the range of alternative distributions $d\tau(w \mid \theta)$ and renders each of them "informative" as well. Consequently, the associated collection of distributions entertained in section 5.2 is more limited than those in the section 5.1 formulation. See remark 5.1 for more about this issue. The distinct ways in which the section 5.1 and 5.2 formulations use statistical discrepancies lead to substantial differences in the resulting variational preferences, namely, representation (15) for the section 5.1 setting of not knowing the distribution of $d\tau(w)$ and (23) for the section 5.2 formulation.

5.4 Not knowing a prior, II

We modify preferences by using a statistical divergence to constrain a set of prior probabilities. The resulting preferences satisfy the axioms of Gilboa and Schmeidler (1989). Consider:

$$\Pi = \{\pi : d\pi(\theta) = n(\theta) d\pi_o(\theta), n \in \mathcal{N}, \int \phi[n(\theta)] d\pi_o(\theta) \leqslant \kappa\} \tag{25}$$

where $\kappa > 0$ pins down the size of the set of priors. Preferences over $\rho(w)$ are ordered by

$$\min_{\pi \in \Pi} \int_\Theta \left(\int_W u[\rho(w)] d\tau(w \mid \theta) \right) d\pi(\theta). \tag{26}$$

Remark 5.10. *The minimized objective for problem* (26) *can again be evaluated using convex duality theory via*

$$\max_{\eta, \xi \geqslant 0} \int_\Theta \phi^* \left[\int_W u[\rho(w)] d\tau(w \mid \theta) + \eta \mid \xi \right] d\pi_o(\theta) - \eta - \xi\kappa.$$

Maximization over $\xi \geqslant 0$ enforces a constraint on the set of admissible priors.

Denti and Pomatto (2022)'s axiomatic construction of a parameterized likelihood to be used in Klibanoff et al. (2005) preferences.

5.5 An Example

It is instructive to apply the distinct approaches of subsections 5.1 and 5.2 to a simple example. To apply the subsection 5.1 approach, we take the following constituents:

- Baseline model $d\tau_o(w) \sim \mathcal{N}(\mu_o, \sigma_o^2)$
- Prize $c(w) = \rho(w)$
- Utility function $u[c(w)] = \log[c(w)]$, where $c(w)$ is consumption
- Decision rule $\rho(w) = \exp(\rho_0 + \rho_1 w)$

When we use relative entropy as statistical divergence, variational preferences for a subsection 5.1 decision maker are ordered by

$$\rho_0 + \rho_1\mu_0 - \frac{1}{2\xi}(\sigma_0\rho_1)^2$$

Larger values of the positive scalar ξ call for smaller adjustments $-\frac{1}{2\xi}(\sigma_0\rho_1)^2$ of expected utility $\rho_0 + \rho_1\mu_0$ for concerns about misspecification of $d\tau_o$.

To study a subsection 5.2 decision maker, we add the following constituents to the example:

- Alternative structured models $\sim \mathcal{N}(\mu_i, \sigma_i^2), i = 1, \ldots, \ell$, where potential parameter values (states) are $\theta_i = (\mu_i, \sigma_i)$ and parameter space $\Theta = \{\theta_i : i = 1, 2, \ldots, k\}$
- Baseline prior over structured models is a uniform distribution $\pi_o(\theta_i) = \frac{1}{k}, i = 1, \ldots, \ell$

To obtain an alternative prior π_i for $i = 1, \ldots, \ell$, we set $n_i = k\pi_i$ so that the product of n_i times the baseline prior is:

$$\frac{n_i}{k} = \pi_i.$$

Expected utility conditioned on parameter vector θ_i is

$$\int u\left[\exp(\rho_0 + \rho_1 w)\right] d\tau(w \mid \theta) = \rho_0 + \rho_1\mu_i$$

and a statistical divergence is

$$\frac{1}{k}\sum_{i=1}^{k} \phi\left(k\pi_i\right).$$

A subsection 5.2 decision maker with variational preferences orders decision rules $\rho(w) = \exp(\rho_0 + \rho_1 w)$ according to

$$\min_{\pi_i \geqslant 0, \sum_{i=1}^k \pi_i} \rho_0 + \rho_1 \sum_{i=1}^{k} \pi_i\mu_i + \frac{\xi}{k}\sum_{i=1}^{k} \phi\left(k\pi_i\right).$$

For a relative entropy divergence, decision rules are ordered by

$$-\xi \log \sum_{i=1}^{k} \left(\frac{1}{k}\right) \exp\left[-\frac{1}{\xi}\left(\rho_0 + \rho_1\mu_i\right)\right] = \rho_0 - \xi \log \sum_{i=1}^{k} \left(\frac{1}{k}\right) \exp\left(-\frac{\rho_1\mu_i}{\xi}\right)$$

and the associated minimizing π_i is

$$\frac{\exp\left(-\frac{\rho_1\mu_i}{\xi}\right)}{\sum_{i=1}^{k} \exp\left(-\frac{\rho_1\mu_i}{\xi}\right)}$$

6 Hybrid models

We now use components described above as inputs into a representation of preferences that includes uncertainty about a prior to put over structured models as well as concerns about possible misspecifications of those structured models. We use probability perturbations in the form of relative densities in $\mathcal{M}$ to capture uncertainty about models and probability perturbations in the form of relative densities $\mathcal{N}$ to capture uncertainty about a prior over models. To represent a family of structured models for W, it is helpful to write a parameterized family of relative densities as

$$\ell(w|\theta) \geqslant 0$$

where

$$\ell(w \mid \theta) \in \mathcal{M} \qquad \forall \theta \in \Theta.$$

We represent a family of structured models as

$$d\tau(w \mid \theta) = \ell(w \mid \theta) d\tau_o(w)$$

where $\tau_o(w)$ is now used to represent the family of structured models. Absolute continuity of a parameterized family of models is widely used in likelihood theory. The probability measure $d\tau_o$ does not itself have to be a structured model.[20]

Let $\pi_o(\theta)$ is a baseline prior over θ. To conduct a prior robustness analysis, consider alternative priors

$$d\pi(\theta) = n(\theta) d\pi_o(\theta)$$

for $n \in \mathcal{N}$.

Consider relative densities $\hat{m}$ that for each θ have been rescaled so that

$$\int \hat{m}(w \mid \theta) \ell(w|\theta) d\tau_o(w) = 1.$$

To acknowledge misspecification of a model implied by parameter θ, let $\hat{m}(w|\theta)$ to represent an "unstructured" perturbation of that model. With this in mind, let $\widehat{\mathcal{M}}$ be the space of admissible relative densities $\hat{m}(w|\theta)$ associated with model θ for each $\theta \in \Theta$. We then consider a composite parameter $(\hat{m}, \theta)$ for $\hat{m} \in \widehat{\mathcal{M}}$ and $\theta \in \Theta$. The composite parameter $(\hat{m}, \theta)$ implies a distribution $\hat{m}(w \mid \theta)\ell(w \mid \theta)d\tau_o(w)$ over W conditioned on θ.

To measure a statistical discrepancy brought by applying $\hat{m}$ to the density ℓ of w conditioned on θ and by applying n to the baseline prior over θ, we first acknowledge possible misspecification of each of the θ models by computing:

$$\mathbb{T}_1[\rho](\theta) = \min_{\hat{m} \in \widehat{\mathcal{M}}} \int_W \left(u[\rho(w)] \hat{m}(w \mid \theta) + \xi_1 \phi_1 \left[\hat{m}(w \mid \theta) \right] \right) \ell(w \mid \theta) d\tau_o(w)$$

The $\mathbb{T}_1$ operator maps control laws, ρ, into functions of θ. We use this for both hybrid approaches.

[20] The counterpart to $d\tau_o(w)$ in likelihood theory is a measure, but not necessarily a probability measure. However, a parameterized family can typically also be represented with a baseline probability measure.

6.1 First hybrid model

We can rank alternative decision rules ρ by including an adjustment for possible misspecification of the baseline prior π_o:

$$\mathbb{T}_2 \circ \mathbb{T}_1[\rho] = \min_{n \in \mathcal{N}} \int_\Theta \left(\mathbb{T}_1[\rho](\theta) n(\theta) + \xi_2 \phi_2[n(\theta)] \right) d\pi_o(\theta).$$

This hybrid formulation allows possibly distinct convex functions ϕ_1 and ϕ_2 with properties like ones that we imposed on ϕ in section 4.

This two-step construction leads to an implied one-step variational representation with a composite divergence. For $\hat{m} \in \widehat{\mathcal{M}}$ and $n \in \mathcal{N}$, form a composite scaled statistical discrepancy

$$d(\hat{m}, n) = \xi_1 \int_\Theta \left(\int_W \phi_1 \left[\hat{m}(w \mid \theta) \right] d\ell(w \mid \theta) \right) n(\theta) d\pi_o(\theta) + \xi_2 \int_\Theta \phi_2 \left[n(\theta) \right] d\pi_o(\theta) \tag{27}$$

for $\xi_1 > 0, \xi_2 > 0$. Then variational preferences are ordered by:

$$\min_{\hat{m} \in \widehat{\mathcal{M}}, n \in \mathcal{N}} \int_\Theta \left(\int_W u[\rho(w)] \hat{m}(w \mid \theta) \ell(w \mid \theta) d\tau_o(w) \right) n(\theta) d\pi_o(\theta) + d(\hat{m}, n)$$

In Appendix A we establish that divergence (27) is convex over the family of probability measures that concerns the decision maker.

6.2 Second hybrid model

As an alternative to the section 6.1 approach, we could instead constrain the set of priors to satisfy:

$$\int_\Theta \phi_2[n(\theta)] d\pi_o(\theta) \leqslant \kappa \tag{28}$$

so that a decision maker's preferences over decision rules ρ would be ordered by:

$$\min_{n \in \mathcal{N},} \int_\Theta \mathbb{T}_1 \left[\rho \right] (\theta) n(\theta) d\pi_o(\theta), \tag{29}$$

where minimization is subject to (28).[21]

In the spirit of Cerreia-Vioglio et al. (2021), preferences ordered by (29) subject to constraint (28) can be thought of as using a divergence between a potentially misspecified probability distribution and a set of predictive predictive distributions that have been constructed from priors over a parameterized family of probability densities within the constrained set Θ. Notice how the first term in discrepancy measure (27) uses a prior $nd\pi_o$ to construct a weighted averaged over $\theta \in \Theta$ of the the following conditioned-on-θ misspecification measure

$$\xi_1 \left(\int_W \phi_1 \left[\hat{m}(w \mid \theta) \right] d\ell(w \mid \theta) \right).$$

The objective in problem (29) is to make the divergence between a given distribution and each of the parameterized probability models small on average by minimizing over how to weight divergence measures

[21] Cerreia-Vioglio et al. (2021) provide an axiomatic justification of set-based divergences as a way to capture model misspecification within a Gilboa et al. (2010) setup with multiple models. Their divergence measures feature sets of probability models or sets of predictive distributions, but they do not include divergences using a family of priors as we do here.

indexed by θ subject to the constraint that $\pi \in \Pi$.[22] Equivalently, in place of (27), this approach uses cost function

$$d(\hat{m}, n) = \xi_1 \min_{n \in \mathcal{N}} \int \left(\int_W \phi_1 \left[\hat{m}(w \mid \theta) \right] d\ell(w \mid \theta) \right) n(\theta) d\pi_o(\theta).$$

Remark 6.1. *It is possible to simplify computations by using dual versions of the hybrid approaches delineated in subsections 6.1 and 6.2. Such formulations closely parallel ones described in our discussions of robust prior analysis and potential model misspecification in remarks 5.3, 5.4, and 5.5.*

7 An approach to uncertainty quantification

Subsection 6 posed a minimum problem that comes from variational preferences with a two-parameter cost function that we constructed from two statistical divergences. The minimum problem produces a robust decision rule along with a worst-case probability distribution. Strictly speaking, the decision theory tells us that particular values of cost function parameters (ξ_1, ξ_2) reflect a decision maker's concerns about uncertainty, broadly conceived. Questioning those values is none of our business as outside observers of, say, someone who makes public policy or private decisions. Nevertheless, in the spirit of Good (1952), it can be enlightening to study worst-case distributions as functions of (ξ_1, ξ_2). The concluding paragraph of Chamberlain (2020) recommends such sensitivity analyses of both a likelihood and a prior. Sensitivity of worst-case distributions to cost function parameters provides evidence about the forms of subjective uncertainty and potential model misspecification that *should* be of most concern. That can provide better understandings of the consequences of uncertainty aversion to decision makers and outside analysts.

Motivated partly by a robust Bayesian approach, we have used decision theory to suggest a new approach to uncertainty quantification. By varying two aversion parameters (ξ_1, ξ_2), we can trace out two-dimensional representations of decision rules and worst-case probabilities. The representation of worst-case probabilities includes both worst-case priors and a worst-case alteration to each member of a parametric family of models. A decision maker can explore alternative choices and the expected utilities by varying the aversion parameters (ξ_1, ξ_2) that trace out the two-dimensional set of worst-case probabilities. In this way, we reduce potentially high-dimensional subjective uncertainties to a two-dimensional collection of alternative probability specifications that should most concern a decision maker along with robust decision rules for responding to those uncertainties.

8 Concluding remarks

Because we have confined ourselves to a "static" setting, we have worked within the framework created by Maccheroni et al. (2006a). We intend this as a prolegomenon to a paper that will analyze relate issues in dynamic contexts in which our starting point will instead be the dynamic variational preferences of Maccheroni et al. (2006b) together with a link to a dynamic measure of statistical divergence based on relative entropy and the recursive preferences of Kreps and Porteus (1978) and Epstein and Zin (1989). While many issues studied here recur in that framework, additional issues such as dynamic consistency and appropriate state variables for recursive formulations of preferences arise.

[22] By emphasizing a family of structured models, this set-divergence concept differs from an alternative that could be constructed in terms of an implied family of predictive distributions.

A Convexity of composite divergence (27)

To verify convexity of (27), consider two joint probability measures on $W \times \Theta$:

$$\hat{m}_0(w \mid \theta)\ell(w \mid \theta)d\tau_o(w)n_0(\theta)d\pi_o(\theta)$$
$$\hat{m}_1(w \mid \theta)\ell(w \mid \theta)d\tau_o(w)n_1(\theta)d\pi_o(\theta).$$

A convex combination of these two probability measure is itself a probability measure. Use weights $1-\alpha$ and α to construct a convex combination and then factor it in the following way. First, compute the marginal probability distribution for θ expressed as $n_\alpha(\theta)d\pi_o(\theta)$:

$$n_\alpha(\theta) = (1-\alpha)n_0(\theta) + \alpha n_1(\theta).$$

By the convexity of ϕ_2, it follows that

$$\phi_2[n_\alpha(\theta)] \leqslant (1-\alpha)\phi_2[n_0(\theta)] + \alpha\phi_2[n_1(\theta)]. \tag{30}$$

Next note that

$$\begin{aligned}\hat{m}_\alpha(w \mid \theta) = &\left[\frac{(1-\alpha)n_0(\theta)}{(1-\alpha)n_0(\theta) + \alpha n_1(\theta)}\right]\hat{m}_0(w \mid \theta) \\ &+ \left[\frac{\alpha n_1(\theta)}{(1-\alpha)n_0(\theta) + \alpha n_1(\theta)}\right]\hat{m}_1(w \mid \theta).\end{aligned}$$

By the convexity of ϕ_1

$$\begin{aligned}\phi_1[\hat{m}_\alpha(w \mid \theta)] \leqslant &\left[\frac{(1-\alpha)n_0(\theta)}{(1-\alpha)n_0(\theta) + \alpha n_1(\theta)}\right]\phi_1[\hat{m}_0(w \mid \theta)] \\ &+ \left[\frac{\alpha n_1(\theta)}{(1-\alpha)n_0(\theta) + \alpha n_1(\theta)}\right]\phi_1[\hat{m}_1(w \mid \theta)].\end{aligned}$$

Thus,

$$\phi_1[\hat{m}_\alpha(w \mid \theta)]n_\alpha(\theta) \leqslant (1-\alpha)n_0(\theta)\phi_1[\hat{m}_0(w \mid \theta)] + \alpha n_1(\theta)\phi_1[\hat{m}_1(w \mid \theta)]. \tag{31}$$

Multiply (31) by ξ_1 and (30) by ξ_2, add the resulting two terms, and integrate with respect to $\ell(w \mid \theta)d\tau_o(w)d\pi_o(\theta)$ to verify that divergence (27) is indeed convex in probability measures that concern the decision maker.

References

Anscombe, F J and R J Aumann. 1963. A Definition of Subjective Probability. *The Annals of Mathematical Statistics* 34 (1):199–205.

Berger, James O. 1984. The Robust Bayesian Viewpoint. In *Robustness of Bayesian Analysis*, vol. 4 of *Studies in Bayesian Econometrics*, edited by Joseph B Kadane, 63–124. North-Holland, Amsterdam.

Bucklew, James A. 2004. *An Introduction to Rare Event Simulation*. New York: Springer Verlag.

Cerreia-Vioglio, Simone, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio. 2013. Ambiguity and Robust Statistics. *Journal of Economic Theory* 148 (3):974–1049.

Cerreia-Vioglio, Simone, Lars Peter Hansen, Fabio Macchernoni, and Massimo Marinacci. 2021. Making Decisions under Model Misspecification. Available at SSRN.

Chamberlain, Gary. 2020. Robust Decision Theory and Econometrics. *Annual Review of Economics* 12:239–271.

Denti, Tommaso and Luciano Pomatto. 2022. Model and predictive uncertainty: A foundation for smooth ambiguity preferences. *Econometrica* 90 (2):551–584.

Diaconis, Persi and David A. Freedman. 1986. On the Consistency of Bayes Estimates. *Annals of Statistics* 14 (1):1–26.

Dupuis, Paul and Richard S Ellis. 1997. *A Weak Convergence Approach to the Theory of Large Deviations*. New York: John Wiley & Sons.

Epstein, Larry G. and Stanley E. Zin. 1989. Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica* 57 (4):937–969.

de Finetti, Bruno. 1937. La prevision: ses lois logiques, ses sources subjectives. *Annales de l'Institute Henri Poincaré* 7:1–68.

Fishburn, Peter C. 1970. *Utility Theory for Decision Making*. New York: Wiley.

Freedman, David A. 1963. "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case. *Annals of Mathematical Statistics* 34 (4):1386–1403.

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

Gilboa, Itzhak, Fabio Maccheroni, Massimo Marinacci, and David Schmeidler. 2010. Objective and Subjective Rationality in a Multiple Prior Model. *Econometrica* 78 (2).

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

Hansen, Lars Peter and Thomas J. Sargent. 2007. Recursive Robust Estimation and Control without Commitment. *Journal of Economic Theory* 136 (1):1–27.

———. 2019. Acknowledging and Pricing Macroeconomic Uncertainties. VOXEU.

Hansen, Lars Peter and Thomas J Sargent. 2021. Macroeconomic Uncertainty Prices when Beliefs are Tenuous. *Journal of Econometrics* 223 (1):222–250.

Hansen, Lars Peter and Thomas J. Sargent. 2022. Structured Ambiguity and Model Misspecification. *Journal of Economic Theory* 199:105–165.

Herstein, I. N. and John Milnor. 1953. An Axiomatic Approach to Measurable Utility. *Econometrica* 21 (2):291–297.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. A Smooth Model of Decision Making Under Uncertainty. *Econometrica* 73 (6):1849–1892.

Kreps, David M. 1988. *Notes on the Theory of Choice*. Boulder, Colorado: Westview Press.

Kreps, David M. and Evan L. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice. *Econometrica* 46 (1):185–200.

Luce, R. Duncan and Howard Raiffa. 1957. *Games and Decisions: Introduction and Critical Survey*. New York: John Wiley & Sons.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. 2006a. Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica* 74 (6):1147–1498.

———. 2006b. Dynamic Variational Preferences. *Journal of Economic Theory* 128 (1):4–44.

Marinacci, Massimo and Simone Cerreia-Vioglio. 2021. Countable Additive Variational Preferences. Bocconi University.

von Neumann, John and Oskar Morgenstern. 2004. *Theory of Games and Economic Behavior: Sixtieth Anniversary Edition*. Princeton and Oxford: Princeton University Press.

Savage, L J. 1954. *The Foundations of Statistics*. New York: John Wiley and Sons.

Segal, Uzi. 1990. Two-Stage Lotteries without the Reduction Axiom. *Econometrica* 58 (2):349–377.

Sims, Christopher A. 1971. Distributed Lag Estimation When the Parameter-Space is Explicitly Infinite-Dimensional. *Annals of Mathematical Statistics* 42 (5):1622–1636.

———. 2010. Understanding non-bayesians. Unpublished chapter, Department of Economics, Princeton University.

Strzalecki, Tomasz. 2011. Axiomatic Foundations of Multiplier Preferences. *Econometrica* 79 (1):47–73.

Von Neumann, J and O Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.