

WORKING PAPER · NO. 2022-157

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

*Lars Peter Hansen and Thomas J. Sargent*

NOVEMBER 2022

# Risk, Ambiguity, and Misspecification: Decision Theory, Robust Control, and Statistics\*

Lars Peter Hansen<sup>†</sup> and Thomas J. Sargent<sup>‡</sup>

October 17, 2022

## Abstract

We connect variational preferences with the likelihood functions and prior probabilities over parameters that are building blocks of statistics and econometrics. We use relative entropy and other statistical divergences as cost functions in the variational preferences 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

**JEL Codes**— C10,C14,C18

## 1 Introduction

Practicing econometricians often struggle with uncertainty about their statistical models, but usually with scant guidance from advances in decision theory made after Wald (1947, 1949, 1950), Savage (1954), and Ellsberg (1961).<sup>1</sup> This might be because much recent formal 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

---

\*We thank Tommaso Denti, Marco Loseto, Luciano Pomatto, Giulio Principi, Doron Ravid, and participants of the Econometrics Advising Group at the University of Chicago for criticisms of earlier drafts. We are particularly appreciative of repeated conversations and helpful comments by Simone Cerreia Vioglio, Fabio Angelo Maccheroni, and Massimo Marinacci that helped us understand recent advances in decision theory that address uncertainty with an eye towards model building in the presence of statistical challenges. Diana Petrova, Haomin Qin and Samuel Zhao provided assistance with the preparation of the document. We are thankful for the financial support from the Alfred P. Sloan Foundation (grant G-2018-11113) toward this research.

<sup>†</sup>University of Chicago. Email: lhansen@uchicago.edu

<sup>‡</sup>New York University. Email: thomas.sargent@nyu.edu

<sup>1</sup>Examples of econometricians who explicitly confronted model uncertainty include Onatski and Stock (2002), Brock et al. (2003), Stock and Watson (2006), Brock et al. (2007), Del Negro and Schorfheide (2009), Christensen (2018), Christensen and Connault (2019), Christensen et al. (2020), Andrews and Shapiro (2021), and Bonhomme and Weidner (2021). Chamberlain (2000, 2001) used a post Wald-Savage decision theory of Gilboa and Schmeidler (1989) to confront model uncertainty in his econometric work.

decision maker’s subjective belief about parameters.<sup>2</sup> By distinguishing roles played by likelihood functions and subjective priors over parameters, this paper aims to bring decision theory, post Wald and Savage, into closer contact with statistics and econometrics in ways that can address practical econometric concerns about model misspecifications and selection of prior probabilities.

Although they proceeded differently than we do, Chamberlain (2020), Cerreia-Vioglio et al. (2013), and Denti and Pomatto (2022) studied related issues. Chamberlain (2020) emphasized that likelihoods and priors are both vulnerable to uncertainties. Cerreia-Vioglio et al. (2013) and Denti and Pomatto (2022) focused on uncertainty about predictive distributions that they constructed by integrating likelihoods with respect to priors. Since a likelihood describes probabilities over events that are of direct interest to a decision maker conditioned on parameters, alternative priors over parameters induce ambiguity about probabilities over such events, a focus for both of these papers.<sup>3</sup> But neither of those papers sharply distinguishes prior uncertainty from concerns about possible model misspecifications, which is something that we want to 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 that affect 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 influence the concerns about robustness and ambiguity that we are able to represent with variational preferences.

Some “behavioral” models in economics and finance assume expected utility preferences in which an agent’s subjective probability differs systematically from probabilities assumed to govern the data.<sup>4</sup> This literature also contains discussions of differences in agents’ ‘confidences’ in their views of the world. As we open the door to alternative notions of uncertainty, lack of confidence can take different forms. Preference structures that we describe in this paper allow us to formalize different degrees of “confidence” both about details of specifications of particular statistical models and about subjective probabilities attached to alternative statistical models. Our representations of preferences provide ways to characterize degrees of confidence in terms of perceived statistical plausibilities.<sup>5</sup>

---

<sup>2</sup>The term likelihood can have multiple meanings. We shall use it to represent a probability density of prize-relevant shocks conditioned on parameters. Distinguishing likelihood functions from subjective priors is fundamental to Bayesian formulations of statistical learning. See de Finetti (1937), who recommended exchangeability as a more suitable assumption than iid (independent and identically distributed) to model situations in which a decision maker wants to learn. Putting subjective probabilities over parameters that index likelihood functions for iid sequences of random vectors generates exchangeable sequences of random variables.

<sup>3</sup>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.

<sup>4</sup>We put “behavioral” in quotes to emphasize that most economic models are about agents’ behaviors, including models that impose the rational expectations and common knowledge assumptions that “behavioral” economists want to drop. “Behavioral” economics sometimes means work that is linked more or less informally to psychology.

<sup>5</sup>Although we provide no formal links to psychology here, we think that a promising research plan would explore connections between so-called behavior distortions and the inferential challenges that economic decision makers confront. As is often assumed in behavioral models, degrees of confidence could differ across economic agents.

## 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 prize rules, each of which we represent as a function  $\gamma : W \rightarrow X$ , where  $x \in X$  is a “prize” that he cares about. In our featured examples, the outcome  $\gamma(w)$  determines the decision maker’s exposure to an uncertain random vector that has realization,  $w \in W$ . A set of  $\gamma$ ’s describes the prize rules under consideration by the decision maker. (In section 2, we will provide a more detailed description of a function  $\gamma$  as the outcome from a more fully articulated decision process.) The parameter space  $\Theta$  can be finite or infinite dimensional;  $d\tau(w|\theta)$  is a member of a family of distributions indexed by  $\theta \in \Theta$ . When  $\Theta$  is infinite dimensional, we say that  $d\tau(w|\theta)$  for  $\theta \in \Theta$  is a “nonparametric” family of probability distributions. The “non-informativeness” of a decision maker’s set of possible “prior” probability distributions over  $\Theta$  plays an important role in justifying alternative approaches to “robustness” that we describe in section 4.

We use three key components from decision theory: i) states, ii) acts, and iii) prizes, but we use them differently than many other authors do. We follow Anscombe and Aumann (1963) by defining consequences as lotteries over prizes. An act maps states into consequences. A decision maker’s preferences are defined over acts. In the static setup of this paper, we take a state to be a parameter of a statistical model. That distinguishes our formulation from many other applications of Anscombe and Aumann (1963). For example, decision theorists who connect their work to revealed preference theory typically want states that are “verifiable”. But we are interested in situations, typical in econometrics, where parameters of statistical models are hidden and can be ferreted out eventually, if ever, only by invoking limits associated with the Law of Large Numbers. Because parameter uncertainty is central for us, it is important that the parameter  $\theta$  be included as at least a component of the state.<sup>6</sup>

Gilboa et al. (2010) and Cerreia-Vioglio et al. (2013) introduced parameterized models as a family of primitive probabilities that a decision maker cares about. In effect, Cerreia-Vioglio et al. (2013) considered an expanded state space  $(w, \theta)$  that includes both shocks with realization  $w$  and parameters  $\theta$  and then take a *model* to be a conditional distribution over  $(W, \mathfrak{W})$  given  $\theta$ .<sup>7</sup> Consistent with the framework of Gilboa et al. (2010), Cerreia-Vioglio et al. showed 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 in the family.

In contrast to Cerreia-Vioglio et al. and many other applications of the Anscombe and Aumann (1963) framework, we use lotteries in a more essential way. Anscombe and Aumann (1963) motivate lotteries as “roulette wheels” with known (objective) probabilities, in contrast to “horse races” with unknown (subjective) probabilities. Much previous research used an Anscombe and Aumann (1963) setup as a mathematical vehicle to extend Von Neumann and Morgenstern (1944) preferences defined over lotteries to more general settings that could include subjective uncertainty. In our formulation, the random vector  $W$  induces a probability distribution that according to a particular act implies a particular lottery that can depend on a parameter of a statistical model. This formulation imagines someone who represents a family of models

---

<sup>6</sup>Stephen Stigler showed us a short working paper by Savage (1952) entitled “An Axiomatic Theory of Reasonable Behavior in the Face of Uncertainty,” a prolegomenon to the axiomatic structure presented in Savage (1954). Savage (1952) wrote this: “The set  $S$  represents the conceivable states, or descriptions of the world, or milieu, with which the person is concerned . . .” We think of parameter values or model selection indicators as presenting a “description of the world.”

<sup>7</sup>Cerreia-Vioglio et al. (2013) deploy a “Dynkin space” and an associated sigma algebra of events. Their conditioning on those events is a counterpart to our conditioning on a model. As an alternative, Denti and Pomatto (2022) used an axiomatic approach to define a parameterized set of models. While both of those approaches are interesting, we suppose that models can have epistemological origins. In this, we follow Hansen and Sargent (2022) who refer to such models as “structured models.”

as a manifold of probability distributions indexed by an unknown parameter vector. Parameter vectors can reside in a finite set or on a manifold of possible values. This way of using the Anscombe and Aumann (1963) framework lets us distinguish robustness to misspecification of each member of a collection of substantively motivated “structured” statistical models from robustness to the choice of a prior distribution over alternative models. We formulate preferences that express and distinguish concerns about both types of robustness.

Maccheroni et al. (2006a) and Strzalecki (2011) used Hansen and Sargent’s (2001) stochastic formulation of a robust control problem as a way to motivate their axioms. We apply our Anscombe and Aumann formulation to describe how those axioms actually express prior uncertainty rather than the model misspecification concerns that originally motivated Hansen and Sargent (2001). But we also show how, by using an appropriate ambiguity index or “cost” function, we can use the variational preferences of Maccheroni et al. (2006a) to express concerns about robustness both to statistical model misspecification and to prior choice, including priors meant to support “nonparametric Bayesian” methods.

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. A subsection 5.1 decision maker has a unique baseline model that he distrusts and seeks robustness with respect to statistically nearby models. A subsection 5.2 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 these two decision makers 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. Section 6 describes a candidate for a cost function to use for a variational preferences representation of a decision maker who is concerned about *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*,  $\Pi$  is a set of probability measures over states, and  $\Lambda$  is a set of probability measures over prizes.<sup>8</sup> 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. Let  $\mathcal{X}$  denote an event in  $\mathfrak{X}$  and  $\mathcal{G}$  denote an event in  $\mathfrak{G}$ .

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

For a given  $\theta$ ,  $f(\theta) \in \Lambda$  is a lottery over possible prizes  $x \in X$ .<sup>9</sup> We let  $df(x | \theta)$  denote integration with respect to probabilities described by that lottery. For a *given* probability measure  $\pi \in \Pi$ ,  $\int_{\Theta} f(dx | \theta)\pi(d\theta)$

<sup>8</sup>For a discussion of the Anscombe-Aumann setup, see Kreps (1988), especially chapters 5 and 7.

<sup>9</sup>The basic setup used here borrows from Marinacci and Cerreia-Vioglio (2021). Following the leads of de Finetti and Savage, 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. However, countable additivity simplifies the presentation and is routinely imposed in much of probability theory.

is a two-stage lottery over prizes, with one lottery over states  $\theta$  being described by  $\pi$  and another lottery over prizes  $x \in X$  being described by  $df(x | \theta)$  that depends on the outcome  $\theta$  from the other lottery. We shall introduce uncertainty about  $\pi$ , the probabilities assigned to the  $\theta$  events.

As mentioned in section 1, we shall interpret objects in the Anscombe and Aumann formulation in ways that relate to our work as statisticians/econometricians. We interpret a state  $\theta$  as pinning down one among a set  $\Theta$  of probability 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  $x \in X$ .

We use “prize rules” to represent alternative acts. A prize rule is a function  $\gamma : W \rightarrow X$ , where  $x \in X$ . A prize rule thus determines how a prize depends on underlying shocks  $W$ . Conditioned on parameter  $\theta$ , the function  $\gamma$  in conjunction with the family of distributions  $d\tau(w | \theta)$  implies a lottery over prizes in  $X$ . Thus, for events  $\mathcal{X} \in \mathfrak{X}$  and a prize rule  $\gamma$ , we induce a lottery  $f(x | \theta)$  by using  $d\tau(w | \theta)$  to assign conditional probabilities to events  $\mathfrak{W}$  of the form

$$\mathcal{W} = \{w : \gamma(w) \in \mathcal{X}\}.$$

In our setting, alternative prize rules  $\gamma$  imply alternative Anscombe and Aumann acts. This is convenient for us as applied statisticians because a parameterized family of distributions  $d\tau(w | \theta)$  can be used to construct a manifold of likelihoods indexed by unknown parameter vector  $\theta \in \Theta$ . We refer to the resulting collection of acts as a set of acts *induced* by prize rules. A particular decision problem will circumscribe a family of prize rules and hence a family of Anscombe and Aumann acts.

**Remark 2.2.** *For some of our purposes, it helps to represent a function  $\gamma$  as*

$$\gamma(w) = \Gamma(d, w) \tag{1}$$

*for a decision  $d \in D$  and another function  $\Gamma$ . Thus,  $\gamma$  depends implicitly on  $d$ , so we sweep out a family of prize rules as we select different possible “decisions”  $d \in D$ . As an example of such a formulation,  $d$  could be a particular investment vehicle whose random return is exposed to shock  $W$  in a particular way. In this stylized setting, we shall interpret  $d\tau(w|\theta)$  as a “statistical model” and the pair  $(\Gamma, \tau)$  as a “substantive model”. We can expand the collection of acts by randomizing the decisions. Given two decisions  $d_1$  and  $d_2$ , a randomized rule makes decision  $d_1$  with probability  $\alpha$  and  $d_2$  with probability  $1 - \alpha$ . Since each decision induces an Anscombe and Aumann act, the randomized decision implies a convex combination of the two induced acts.<sup>10</sup>*

In what follows, a decision maker’s *prior* over possible statistical models indexed by  $\theta$  is a probability measure  $\pi \in \Pi$ .

**Remark 2.3.** *The collection of Anscombe and Aumann acts is typically much larger than ones that can be induced by an available set of prize rules  $\gamma$ . We know that the axioms invoked in this paper apply to preferences over the full collection of Anscombe and Aumann acts. While the randomization of decisions described in remark 2.2 enlarges the set of Anscombe and Aumann acts by including the convex hull of the set of acts induced by prize rules, in general that device will not construct the full set of Anscombe and Aumann acts. We recognize that judging the plausibility or “self-evident quality” of the axioms that we impose generally requires extending the set of the acts to be studied beyond the set of “induced-by-prize rules” or even their convex hull that we focus on in this paper.*

---

<sup>10</sup>In some special cases, the set of acts induced by decisions may itself be convex. In this case, the randomization of decisions merely replicates the collection of induced acts.

**Remark 2.4.** *To build another bridge across literatures, we briefly revisit language from mathematical statistics that Ferguson (1967) used to describe a statistical decision problem. Start by positing a distribution  $d\tau(w | \theta)$  for  $W$ , where  $\theta$  is a “true state of nature” (an object that we instead call a parameter). Represent the outcome of what Ferguson calls a statistical experiment as a realization  $y = \beta(w)$  of a random variable that contains information about  $W$ . Let a decision  $d(y)$  depend on the observed value  $y$  of  $Y$ . We constrain what we call a prize rule  $\gamma$  to satisfy:*

$$\gamma(w) = \Gamma[d(y), w] = \Gamma[d \circ \beta(w), w]$$

*for a pre-specified  $\Gamma$  and a measurable function  $d$  that maps observations  $y$  from the statistical experiment into a set of what Ferguson calls actions.<sup>11</sup> Thus, Ferguson allows for an action to depend on a realization  $y$  of  $Y$ .<sup>12</sup> Ferguson’s actions are distinct from Anscombe and Aumann (1963) acts. For us, each prize rule  $\gamma$  implies a probability distribution for a prize conditioned on  $\theta$  that is induced by  $d\tau(w | \theta)$ . We take this to be an Anscombe and Aumann (1963) act. Our decision problem imposes restrictions on admissible choices of  $\gamma$ . We allow different restrictions than those imposed by Ferguson.<sup>13</sup> By design, our formulation opens up dynamic extensions that we explore in a companion paper.*

Let  $\mathcal{A}$  be the set of all acts. Two collections of acts will interest us, a set  $\mathcal{A}_o$  that lets us represent objective uncertainty and another set  $\mathcal{A}_s$  that Anscombe and Aumann (1963) used to express 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  $\theta$ . Given this lack of dependence, the probability distribution  $\pi \in \Pi$  over states plays no role in shaping an ultimate probability distribution over prizes. A constant act constructed from a prize rule,  $\gamma$ , could emerge as follows. Suppose that some component of  $W$  has a known distribution independent of  $\theta$  and that  $\gamma$  depends only on this component. Such limited dependence implies an act that is independent of  $\theta$ . The collection  $\mathcal{A}_s$  consists of acts, each of which delivers a unique prize for each  $\theta$ . We let  $s(\theta) \in X$  denote an act in  $\mathcal{A}_s$ .<sup>14</sup> We use a probability distribution  $\pi \in \Pi$  over states in conjunction with  $\mathcal{A}_s$  to express subjective uncertainty.

**Remark 2.5.** *Anscombe and Aumann (1963) distinguished “horse race lotteries,” represented by acts in  $\mathcal{A}_s$ , from “roulette lotteries,” represented by acts in  $\mathcal{A}_o$ .<sup>15</sup>*

**Remark 2.6.** *While Savage (1954) did not include “objective” lotteries when he rationalized subjective expected utility, his framework allows flexibility in defining both a state and an act. Gilboa et al. (2020) exhibit the flexibility of a Savage-style state space with a variety of applications and discuss the benefits and challenges that this flexibility brings.<sup>16</sup> There is also flexibility in constructing an act. Exploiting this*

<sup>11</sup>For Ferguson (1967),  $d$  as a function of  $y$  is a decision rule distinct from our prize rule.

<sup>12</sup>Ferguson’s formulation of the problem introduces a loss function that for us would be the negative of the expectation of a utility function conditioned on  $(Y, \theta)$  under the distribution implied by  $d\tau(w | \theta)$  and  $\beta$ .

<sup>13</sup>Although he posed it as a static problem, Ferguson (1967)’s formulation can be reinterpreted as a multi-stage or multi-period decision problem in which a decision rule chosen at the outset depends on information that will be revealed in a second stage that in turn influences an uncertain outcome to be realized in a subsequent third stage. We want to explore robustness to prior selection. What is pertinent in the second stage is a posterior conditioned on the outcome of a statistical experiment. In a dynamic setting, the distinction between priors and posteriors becomes obscured as today’s posterior becomes tomorrow’s prior. Since a recursive formulation of a dynamic decision problem essentially reduces a multi-period problem to a two-period problem, the prior/posterior robustness sensitivity could occur in a counterpart to the intermediate stage envisioned by Ferguson (1967).

<sup>14</sup>Technically, an act in  $\mathcal{A}_s$  is a degenerate Dirac lottery with a mass point at  $s(\theta)$  that is assigned probability one.

<sup>15</sup>See Kreps (1988, ch. 5) for more about the distinction.

<sup>16</sup>They did not specifically discuss the statistical linkages that we explore here.

flexibility, Cerreia-Vioglio et al. (2012) produce a preference representation for Anscombe and Aumann acts under Savage (1954) axioms augmented with risk independence. This representation coincides with the familiar Savage representation for acts in  $\mathcal{A}_s$  with unique prizes for each state.<sup>17</sup>

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

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

We shall use instances of our Anscombe and Aumann framework to describe a) a Bayesian decision maker with a unique prior over a set  $\Theta$  of statistical models, b) a decision maker who knows a set  $\Theta$  of statistical models and who copes with *ambiguity* about those models by considering prospective outcomes under a set of priors  $\Pi$  over those statistical models, c) a decision maker with concerns that a single known statistical model  $\theta$  is *misspecified* by using a statistical discrepancy measure to discipline the exploration of the unknown models surrounding that known model, and d) a decision maker with ambiguity and concerns about model misspecifications.

## 2.1 Preferences

To represent a decision maker's preferences over acts, we use  $\sim$  to mean indifference,  $\succeq$  a weak preference, and  $>$  a strict preference. Throughout, we assume that preferences are non-degenerate (there is a strict ranking between two acts), complete (we can compare any pair of acts), and transitive ( $f \succeq g$  and  $g \succeq h$  imply  $f \succeq h$ ). We also impose an Archimedean axiom that provides a form of continuity.<sup>18</sup> 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  $\hat{\Lambda}$  that contains  $\Lambda$ .

## 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  $>_\Lambda$  of the preference order  $>$  to the space of constant acts  $f \in \mathcal{A}_o$  by

$$\lambda >_\Lambda \kappa \iff f > g$$

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

To represent preferences  $>_\Lambda$ , we follow Von Neumann and Morgenstern (1944) who imposed the following restriction:<sup>19</sup>

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

$$f \succeq g \implies \alpha f + (1 - \alpha)h \succeq \alpha g + (1 - \alpha)h.$$

<sup>17</sup>More generally, their representation includes an additional curvature adjustment much like the smooth ambiguity model. See Proposition 3 in their appendix

<sup>18</sup>The Archimedean axiom states: let  $f, g, h$  be acts in  $\mathcal{A}$  with  $f > g > h$ . Then there are  $0 < \alpha < 1$  and  $0 < \beta < 1$  such that  $\alpha f + (1 - \alpha)h > g > \beta f + (1 - \beta)h$ . See Herstein and Milnor (1953, Axiom 2) for an alternative formulation of a continuity axiom.

<sup>19</sup>Completeness, transitivity and the Archimedean axiom carry over directly from  $>$  to  $>_\Lambda$ , but not necessarily non-degeneracy. Our presentation below presumes non-degeneracy of  $>_\Lambda$ .



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

$$f \succeq g \iff U(f) \geq U(g) \quad (2)$$

where

$$U(f) = \int_X u(x) d\lambda(x) \quad (3)$$

and  $\lambda \in \Lambda$  is the probability distribution generated by constant act  $f$ . Representation (3) can be extended to a space  $\hat{\Lambda}$  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 (3). Thus, although  $u$  is in general a nonlinear function of prizes,  $U$  is a linear functional of finite signed measures  $\lambda \in \hat{\Lambda}$ . Consequently, a representation theorem for linear functionals of finite signed measures justifies (3). According to representation (2), for any real number  $r_0$  and strictly positive real number  $r_1$ , utility functions  $r_1 u + r_0$  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 2.8.** (*Independence*) If  $f, g, h \in \mathcal{A}$  and  $\alpha \in (0, 1)$ , then

$$f \succeq g \implies \alpha f + (1 - \alpha)h \succeq \alpha g + (1 - \alpha)h.$$

The second is:

**Axiom 2.9.** (*Monotonicity*) For any  $f, g \in \mathcal{A}$  such that  $f(\theta) \succeq_\Lambda g(\theta)$  for each  $\theta \in \Theta$ ,  $f \succeq g$ .

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

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

for any act  $f$ . A set of acts implies an associated collection  $\mathcal{B}$  of functions  $F$ . From monotonicity axiom 2.9 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

$$\begin{aligned} \int_X u(x) df(x | \theta) &= F(\theta) \\ \int_X u(x) dg(x | \theta) &= G(\theta) \end{aligned}$$

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

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

If the set of acts  $\mathcal{A}$  is convex, then so is the set  $\mathcal{B}$  of functions of  $\theta$ . Furthermore, if  $F \sim_{\Theta} G$ , the independence axiom guarantees that for any  $\alpha$  the associated convex combinations of  $F$  and  $G$  are also in the same indifference set of acts. From one indifference set, we 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  $F \geq G$  on  $\Theta$ . The monotonicity axiom implies that  $\mathcal{L}(F - G) \geq 0$ , 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.<sup>20</sup> 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  $f \in \mathcal{A}$

$$f \succeq g \iff \int_{\Theta} \left[ \int_X u(x)df(x | \theta) \right] d\pi(\theta) \geq \int_{\Theta} \left[ \int_X u(x)dg(x | \theta) \right] d\pi(\theta), \quad (4)$$

where the probability measure  $\pi$  describes subjective probabilities.

Representation (4) lets us interpret the expected utility of an act  $f$  with a two-stage lottery. First, draw a  $\tilde{\theta}$  from  $\pi$  and then draw a prize  $x \in X$  from probability distribution  $df(x | \tilde{\theta})$ . By changing the order of integration, we can write

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

or equivalently

$$\int_{\Theta} \left[ \int_X u(x)df(x | \theta) \right] d\pi(\theta) = \int_X u(x)d\lambda(x), \quad (5)$$

where

$$d\lambda(x) = \int_{\Theta} df(x | \theta)d\pi(\theta). \quad (6)$$

Equation (6) constructs a single lottery  $\lambda$  over  $x$  from the compound lottery generated by  $(d\pi(\theta), df(x | \theta))$ .<sup>21</sup> For a statistician,  $\lambda$  is a “predictive distribution” constructed by integrating over unknown parameter  $\theta$ . Let  $f_c$  be the constant act with lottery  $\lambda$  defined by the left side of (6) for all  $\theta \in \Theta$ . Equations (5) and (6) assert that a person with expected utility preferences is indifferent between  $f_c$  and  $f$ .<sup>22</sup>

<sup>20</sup>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.

<sup>21</sup>Equation (6) thus expresses the “reduction of compound lotteries” described by Luce and Raiffa (1957, p. 26) and analyzed further by Segal (1990).

<sup>22</sup>The statistical decision problem specified by Ferguson (1967) can be solved by computing

$$\max_a \int_{\Theta} \Gamma(a, w)\ell(w | \theta)d\bar{\pi}(\theta | w)$$

where  $d\bar{\pi}(\theta | w)$  is the posterior of  $\theta$  given  $Y = y$ . Notice that  $a$  will depend implicitly on  $y$  which implies the decision rule  $d(y)$ .

## 2.4 Max-min Expected Utility

To construct a decision maker who has max-min expected utility preferences, Gilboa and Schmeidler (1989) replaced Axiom 2.8 with the following two axioms:

**Axiom 2.10.** (*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 2.11.** (*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 that mixing weight  $\alpha$  is known, an assumption that can be interpreted as describing a form of objective uncertainty. Axiom 2.11 asserts a weak preference for mixing with known weights  $\alpha$  and  $1 - \alpha$ .

**Example 2.12.** Suppose that  $\Theta = \{\theta_1, \theta_2\}$  and consider lotteries  $\lambda_1$  and  $\lambda_2$ . Let act  $f$  be lottery  $\lambda_1$  if  $\theta = \theta_1$  and lottery  $\lambda_2$  if  $\theta = \theta_2$ . Let act  $g$  be lottery  $\lambda_2$  if  $\theta = \theta_1$  and lottery  $\lambda_1$  if  $\theta = \theta_2$ . Suppose that  $f \sim g$ . Axiom 2.11 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  $\theta$ . We think of mixing as reducing the exposure to  $\theta$  uncertainty. In the extreme case, setting  $\alpha = \frac{1}{2}$ , for example, completely eliminates effects of exposure to  $\theta$  uncertainty.

By replacing Axiom 2.8 with Axioms 2.10 and 2.11, Gilboa and Schmeidler obtained preferences described by

$$f \succsim g \iff \min_{\pi \in \Pi_c} \int_{\Theta} \left[ \int_X u(x) df(x | \theta) \right] d\pi(\theta) \geq \min_{\pi \in \Pi_c} \int_{\Theta} \left[ \int_X u(x) dg(x | \theta) \right] d\pi(\theta) \quad (7)$$

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 (2), for each  $\theta$ ,  $\int_X u(x) df(x | \theta)$  is an expected utility over prizes  $x$ . Evidently, expected utility preferences (4) are a special case of max-min expected utility preferences (7) in which  $\Pi_c$  is a set with a single member.

## 3 Variational preferences

Maccheroni et al. (2006a) relaxed certainty independence Axiom 2.10 of Gilboa and Schmeidler (1989) to obtain preferences with a yet more general representation that they called variational preferences. Maccheroni et al. replaced Axiom 2.10 with the weaker

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

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

Axiom 3.1 considers only acts that are mixtures of constant acts that can be represented with a single lottery. The axiom states that altering the constant act from  $h$  to  $k$  does not reverse the decision maker's preferences. The *same*  $\alpha$  appears in all three acts being compared. This axiom imparts to preferences a smooth tradeoff between separate contributions that come from an expected utility, on the one hand, and from statistical uncertainty, on the other hand. Mixing with pure lotteries continues to support linearity in evaluations of risks conditioned on states.

To place Axiom 3.1 within the Remark 2.2 setting, use a substantive model  $(\Gamma, \tau)$  to represent the probabilistic outcomes of alternative decisions. Recall that for a given  $\Gamma$ , prize rule  $\gamma(w)$  is described by (1) for some decision  $d$  and some parameterized shock distribution  $d\tau(w \mid \theta)$ . Let's compare the uncertainty consequences of decisions  $d_1$  and  $d_2$  that give rise to prize rules  $\gamma_1$  and  $\gamma_2$  via

$$\begin{aligned}\gamma_1(w) &= \Gamma(d_1, w) \\ \gamma_2(w) &= \Gamma(d_2, w)\end{aligned}$$

for  $d_1, d_2 \in D$ . Each of these prize rules specifies how a prize depends on a realization  $w$  of the shock. Let  $\gamma_1$  induce act  $f$  and  $\gamma_2$  induce act  $g$ .

Now consider two other decisions  $d_3$  and  $d_4$  and use them to construct prize rules

$$\begin{aligned}\gamma_3 &= \Gamma(d_3, w) \\ \gamma_4 &= \Gamma(d_4, w).\end{aligned}$$

Suppose that there exist prize rules  $\gamma_3$  and  $\gamma_4$  both of which induce distributions of the prize  $x \in X$  that do not depend on  $\theta$ ; decisions  $d_3$  and  $d_4$  both serve to target risk components of  $x \in X$  that are not exposed to parameter uncertainty. Denote the constant acts induced by  $\gamma_3$  and  $\gamma_4$ , respectively, as  $h$  and  $k$ . For instance, consider an investment problem for which some of the available investments (indexed by a subset of the decisions  $d \in D$ ) yield returns that depend only on a component of the shock vector that has a known distribution. Two such investments can be used to construct  $\gamma_3$  and  $\gamma_4$ . Axiom 3.1 requires that if randomizing  $d_1$  with respect to  $d_3$  is preferred to randomizing  $d_2$  with respect to  $d_3$ , the preference order will be preserved if  $d_3$  is replaced by  $d_4$  holding fixed the randomization probabilities  $(\alpha, 1 - \alpha)$ .<sup>23</sup>

The substantive model  $\Gamma$  may not include the possibility described in the previous paragraph. But the axioms refer to hypothetical comparisons. To explore Axiom 3.1, we now extend the substantive model  $\Gamma$  to  $\tilde{\Gamma}$  where the arguments of  $\tilde{\Gamma}$  are a realization of an augmented shock vector  $(w, \tilde{w})$  and the decision  $d$  is in a larger set  $\tilde{D}$  that contains  $D$ . Suppose that the  $\tilde{w}$  component of the augmented shock vector has a known distribution that does not depend on  $\theta$ . "Prizes" that depend only on this second component induce constant acts. To confirm that the  $\tilde{\Gamma}$  substantive model is an extension of the original  $\Gamma$  model, we require that

$$\tilde{\Gamma}(d, w, \tilde{w}) = \Gamma(d, w) \quad \text{for } d \in D, w \in D.$$

We then suppose that

$$\begin{aligned}\gamma_3(\tilde{w}) &= \tilde{\Gamma}(d_3, w, \tilde{w}) \\ \gamma_4(\tilde{w}) &= \tilde{\Gamma}(d_4, w, \tilde{w})\end{aligned}$$

where  $d_3, d_4 \in \tilde{D}$  but not necessarily in  $D$ . Decisions  $d_3$  and  $d_4$  confine exposure of the resulting prize to  $\tilde{w}$  and not to  $w$ . As indicated in the previous paragraph,  $\gamma_3$  and  $\gamma_4$  induce constant acts. For our investment example, construction of the extended substantive model  $\tilde{\Gamma}$  could conceivably introduce new opportunities that are not exposed to parameter uncertainty. This opens the door to comparisons entertained by Axiom 3.1.

Maccheroni et al. showed that preferences that satisfy the weaker Axiom 3.1 instead of Axiom 2.10 are

---

<sup>23</sup>We thank Fabio Maccheroni and Massimo Marinacci for suggesting this formulation.

described by

$$f \succeq g \iff \min_{\pi \in \Pi} \int_{\Theta} \left[ \int_X u(x) df(x | \theta) \right] d\pi(\theta) + c(\pi) \geq \min_{\pi \in \Pi} \int_{\Theta} \left[ \int_X u(x) dg(x | \theta) \right] d\pi(\theta) + c(\pi) \quad (8)$$

where, as in representation (2),  $u$  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 (8) replaces the restricted set of probabilities  $\Pi_c$  that appears in the max-min expected utility representation (7). 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 ways of constructing statistical divergences for these two situations are very similar.

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

- $D(\tau | \tau_o) \geq 0$
- $D(\tau | \tau_o) = 0$  implies  $\tau = \tau_o$

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

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

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

$$d\tau(w) = m(w)d\tau_o(w) \quad (9)$$

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

$$\mathcal{M} \doteq \left\{ m : m(w) \geq 0, \int_W m(w)d\tau_o(w) = 1 \right\}. \quad (10)$$

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

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

where  $\xi > 0$ . When  $\xi = 1$ , the divergence is often called a  $\phi$  or  $f$ -divergence. 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  $\tau$  is not absolutely continuous with respect to  $\tau_o$ , we set  $D(\tau | \tau_o)$  to infinity. Relative entropy is commonly referred to as Kullback-Leibler divergence.

**Remark 4.1.** *Other families of divergences can be use in conjunction with the preference representations that follow, for instance, Bregman and Wasserstein divergences. The family  $\phi$  or  $f$  divergences featured here has very nice duality properties. Moreover, the divergence is invariant to one-to-one transformations of the space over which the probability distributions are defined. Finally, some members of this family have very close ties to statistical discrimination. Duality allow us to make formal connections to the extensive literature on smooth ambiguity. The link to likelihood-based statistical discrimination enables statistical constructions that can help us calibrate concerns about robustness.*

## 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  $\gamma \in \Psi$  that map  $w \in W$  into an  $x \in X$ . Think of  $\gamma$  as a prize rule and  $\gamma(w)$  as an uncertain scalar prize. For each prize rule  $\gamma$ , let  $d\lambda(x | \theta)$  be the distribution of the prize  $x = \gamma$  that is induced by distribution  $d\tau(w|\theta)$  and the prize rule  $\gamma$ . The distribution of the prize thus depends both on the prize rule  $\gamma(w)$  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. But he can say that the alternative models that he is most worried about are statistically close to his baseline model. The presence of too many statistically nearby models would prevent a Bayesian from deploying a proper prior over them. Later we will compare our approach here to a robust Bayesian approach that requires a family of priors that are mutually absolutely continuous.<sup>24</sup>

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

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

where  $m \in \mathcal{M}$  for  $\mathcal{M}$  given by (10) 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. A decision maker's expected utility under model  $md\tau_o$  is

$$\int_W u[\gamma(w)]m(w)d\tau_o(w). \tag{11}$$

Notice that (11) evaluates expected utility for a single choice for  $\theta = m$ . The following important technical considerations induce us to proceed in this way.

---

<sup>24</sup>For example, see Berger (1984) for a robust Bayesian perspective.

To complete a description of preferences, we require a scaled statistical divergence. We consider alternative probabilities parameterized by entries in  $\mathcal{M}$ . Under this perspective, a probability model corresponds to a choice of  $m \in \mathcal{M}$ . The object  $m$  is now both a relative density and a state (or parameter value.) Form a scaled divergence measure:

$$c(m) = \xi \int_W \phi[m(w)] d\tau_o(w) \quad (12)$$

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

We explore potential misspecification by entertaining alternative models in the set  $\mathcal{M}$ . Consider first a starting point in which  $\{\theta_i : i = 1, 2, \dots, I\}$ , where  $m_i = \theta_i$  is a “state” that represents a particular alternative model of  $W$  via  $d\tau(w) = m_i(w)d\tau_o(w)$ . Here  $I$  is either a positive integer or infinite. Form

$$\Theta = \left\{ m : m = \sum_{i=1}^I \varpi_i m_i, \text{ where } \varpi_i \geq 0 \text{ and } \sum_{i=1}^I \varpi_i = 1 \right\}$$

Since  $\Theta$  is convex, any subjective probability distribution applied to  $\Theta$  can be represented as:

$$\sum_{i=1}^I \varpi_i m_i$$

for some vector of  $\varpi_i$ 's.<sup>25</sup> We use the convex cost function

$$\tilde{c}(\varpi_1, \varpi_2, \dots, \varpi_I) = c \left( \sum_{i=1}^I \varpi_i m_i(w) \right) = \xi \int_W \phi \left[ \sum_{i=1}^I \varpi_i m_i(w) \right] d\tau_o(w). \quad (13)$$

**Example 5.1.** *Suppose that the probability measure  $\tau_o$  is discrete, with  $I$  points of support and with support point  $i$  having probability  $\varpi_i^o > 0$ . Let*

$$m_i(w) = \begin{cases} \frac{1}{\varpi_i^o} & : w = \text{support point } i \\ 0 & : w \neq \text{support point } i, \end{cases}$$

*so that  $m_i$  assigns probability one to support point  $i$ . Then a probability measure associated with  $\sum_{i=1}^I \varpi_i m_i$  assigns probability  $\varpi_i$  to support point  $i$  and  $\Theta$  consists of all probability models that concentrate probability on all  $I$  of the support points for  $\tau_o$ . Here  $\mathcal{M} = \Theta$  consists of all possible probabilities over the support set of  $\tau_o$ . Cost (13) becomes small when  $m$  is close to one on this same set.*

Consider extending this example to study a decision maker who wants to explore possible misspecifications of his baseline model  $\tau_o$ . The decision maker considers a vast set of possible alternatives to the baseline model  $d\tau_o$  that are in the set  $\Theta = \mathcal{M}$  of likelihood ratios. We use cost  $c$  from (12) to specify costs for deviating from baseline model  $\tau_o$ . When we use (12) to construct preferences, we need not distinguish a probability model as a (relative) density in  $\mathcal{M}$  from a “predictive density” formed from a prior over  $\mathcal{M}$ .<sup>26</sup>

<sup>25</sup>Although the distinction between a model expressed as a parameterized family  $\sum_{i=1}^I \varpi_i m_i$  and a subjective mixture of models formed with a prior probability  $\pi_i = \varpi_i$  is inconsequential when defining static preferences, a model builder with repeated observations will distinguish the two objects. For instance, Bayesian updating rules differ. In dynamic settings posed in Hansen and Sargent (2019, 2021), possible misspecifications are allowed to vary over time in general ways that render Bayesian learning impossible.

<sup>26</sup>While distinguishing between a model and a predictive density is not essential to define static preferences, a model builder with repeated observations will want to distinguish between them. When confronted with a single model that generates the data, Bayesian learning is degenerate. In contrast, when there is a prior over a family of

The implied  $m$  is what matters and not how  $m$  might have been formed as a convex combination of some primitive  $m$ 's in  $\mathcal{M}$ .

**Remark 5.2.** *Maccheroni et al. (2006a) define the domain of their cost function to be probabilities  $\pi$  over the state space, in this case  $\mathcal{M}$ . To map into their framework, consider any probability measure  $\pi$  over  $\mathcal{M}$  and compute*

$$m_\pi = \int_{\mathcal{M}} m d\pi(m).$$

Then define the cost

$$\hat{c}(\pi) = c(m_\pi) = \xi \int_W \phi[m_\pi(w)] d\tau_o(w).$$

Notice that  $\hat{c}(\pi) = 0$  for

$$1 = \int_{\mathcal{M}} m d\pi(m),$$

which is trivially true when  $\pi$  assigns probability one to  $m = 1$  but will also be true for other choices of  $\pi$ .

Variational preferences that use (11) as expected utility over lotteries and (12) as scaled statistical divergence are ordered by

$$\min_{m \in \mathcal{M}} \left( \int_W u[\gamma(w)] m(w) d\tau_o(w) + \xi \int_W \phi[m(w)] d\tau_o(w) \right). \quad (14)$$

This formulation lets a decision maker evaluate alternative prize rules  $\gamma(w)$  while guarding against a concern that his baseline model  $\tau_o$  is misspecified without having in mind specific alternative models  $\tau$ . Key ingredients are the single baseline probability  $\tau_o$  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 (14) by using duality properties of convex functions. Because the objective is separable in  $w$ , we can first compute*

$$\phi^*(\mathbf{u} \mid \xi) = \min_{\mathbf{m} \geq 0} \mathbf{u} \mathbf{m} + \xi \phi(\mathbf{m}) \quad (15)$$

where  $\mathbf{u} = u[\gamma(w)] + \eta$ ,  $\mathbf{m}$  is a nonnegative number, and  $\eta$  is a nonnegative real-valued Lagrange multiplier that we attach to the constraint  $\int m(w) d\tau_o(w) = 1$ ;  $\phi^*(\mathbf{u} \mid \xi)$  is a concave function of  $\mathbf{u}$ .<sup>27</sup> The minimizing value of  $\mathbf{m}$  satisfies

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

The dual problem to the minimization problem on the right side of (14) is

$$\max_{\eta} \int_W \phi^*(u[\gamma(w)] + \eta \mid \xi) d\tau_o(w) - \eta. \quad (16)$$

**Remark 5.4.** *We posed minimum problem (14) 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 (16) depends on  $w$  only through the control law  $\gamma$ , we could instead have used the same convex function  $\phi$  to pose a minimization in*

---

models, each of which could generate the data, there is scope to use Bayes' Law to update weights over alternative models. In dynamic settings studied by Hansen and Sargent (2019, 2021), possible misspecifications vary over time in a vast number of ways that render Bayesian learning impossible.

<sup>27</sup>The function  $-\phi^*(-\mathbf{u} \mid \xi)$  is the Legendre transform of  $\xi\phi(\mathbf{m})$ .



terms of a set of probability distributions  $d\lambda(x)$  with the baseline being the probability distribution over prizes induced  $x = \gamma(w)$  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 searching over alternative  $\gamma$ '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*

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

and dual problem (16) becomes<sup>28</sup>

$$\max_{\eta} -\xi \int \exp\left[-\frac{u[\gamma(w)] + \eta}{\xi} - 1\right] d\tau_o(w) - \eta = -\xi \log\left(\int \exp\left[-\frac{u[\gamma(w)]}{\xi}\right] d\tau_o(w)\right). \quad (17)$$

The minimizing  $m$  in problem (14) is

$$m^*(w) = \frac{\exp\left[-\frac{u[\gamma(w)]}{\xi}\right]}{\int \exp\left[-\frac{u[\gamma(w)]}{\xi}\right] d\tau_o(w)}. \quad (18)$$

The worst-case likelihood ratio  $m^*$  exponentially tilts a lottery toward low-utility outcomes. Bucklew (2004) calls this adverse 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) because of the different way we apply the language of decision theory. Nevertheless, the Maccheroni et al. axiomatic formulation of variational preferences includes our formulation as a special case.

**Remark 5.6.** *(risk-sensitive preferences) The right side of equation (17), namely,*

$$-\xi \log\left[\int_W \exp\left(-\frac{u[\gamma(w)]}{\xi}\right) d\tau_o(w)\right], \quad (19)$$

defines what are known as “risk-sensitive” preferences over control laws  $\gamma$ . Since a 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  $\tau_o$  over shocks. Risk-sensitive preferences are widely used in robust control theory (for example, see Jacobson (1973), Whittle (1990, 1996), and Petersen et al. (2000)).

---

<sup>28</sup>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.

## 5.2 Not knowing a prior, I

Unlike subsection 5.1, we now adopt a setting in which a decision maker has a parameterized 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.<sup>29</sup> Following Gilboa et al., Cerreia-Vioglio et al. and others, we label such distrust of a single prior “model ambiguity.” (We use “fear of misspecificatons” 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 considered by a subsection 5.1 decision maker. The distribution of the prize again depends both on a prize rule  $\gamma(w)$  and on a shock vector distribution  $d\tau(w|\theta)$ . Let  $\Theta$  be a parameter space, and let  $\pi_o$  be a baseline prior probability measure over models  $\theta$ . The baseline  $\pi_o$  anchors a set of priors  $\pi$  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:

$$\mathcal{N} \doteq \left\{ n \geq 0 : n(\theta) \geq 0, \int_{\Theta} n(\theta) d\pi_o(\theta) = 1 \right\}. \quad (20)$$

This specification includes a form of “structured” uncertainty in which all models have the same parametric “structure” but in which each is associated with a different vector of parameter values.<sup>30</sup> 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. To capture a form of ambiguity aversion, the decision maker uses scaled statistical divergence

$$c(\pi) = \xi \int_{\Theta} \phi[n(\theta)] d\pi_o(\theta) \quad (21)$$

and has variational preferences ordered by<sup>31</sup>

$$\min_{n \in \mathcal{N}} \int_{\Theta} \left( \int_W u[\gamma(w)] d\tau(w | \theta) \right) n(\theta) d\pi_o(\theta) + \xi \int_{\Theta} \phi[n(\theta)] d\pi_o(\theta). \quad (22)$$

**Remark 5.7.** *From an appropriate counterpart to dual formulation (16), we can represent variational preferences ordered by (22) as*

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

**Remark 5.8.** *(Smooth ambiguity preferences) When statistical divergence is scaled relative entropy, prefer-*

<sup>29</sup>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.

<sup>30</sup>See Hansen and Sargent (2022).

<sup>31</sup>See Theorem 4 of Cerreia-Vioglio et al. (2013) for their counterpart to this representation.

ences over  $\gamma(w)$  are ordered by

$$-\xi \log \left[ \int \exp \left( -\frac{\int_W u[\gamma(w)] d\tau(w | \theta)}{\xi} \right) d\pi_o(\theta) \right], \quad (23)$$

a static version of preferences that Hansen and Sargent (2007) used to frame a robust dynamic filtering problem. 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 formulation to justify “multiplier preferences” rather than the approach taken here.<sup>32</sup> We emphasize that the robustness being discussed in this subsection is with respect to a baseline prior over known models and not with respect to possible misspecifications of those models.

**Remark 5.9.** If we formulate the set of priors as we have in order to obtain criterion (23), we cannot interpret them as expected utility preferences, unlike the situation described in remark 5.6.

### 5.3 Robustness

It is useful to compare two approaches to robustness that we have taken. The section 5.1 decision maker explores potential model misspecifications by searching over the entire space  $\mathcal{M}$ , subject to a penalty on statistical divergence from a baseline model. The section 5.2 decision maker starts with a baseline prior over parameter vectors and considers consequences of misspecifying that prior. In this subsection, we impose additional structure that allows us to sharpen the comparisons and opens the door to hybrid approaches that we will describe later.

As an application of our section 5.2 approach, we represent the parameterized family of models with  $\ell(w | \theta)$  and

$$d\tau(w | \theta) = \ell(w | \theta) d\tau_o(w).$$

We restrict  $\ell(\cdot | \theta) \in \mathcal{M}$  for each  $\theta \in \Theta$ , where  $d\tau_o(w)$  is a baseline distribution. Even though we do not require that  $\ell(\cdot | \theta) = 1$  identically for some  $\theta \in \Theta$ , each of the parameterized distributions is absolutely continuous with respect to  $d\tau_o(w)$ , as required to apply likelihood-based methods.

This setup allows the parameter space to be infinite dimensional. Consider a prior  $\pi_o$  that is consistent with a Bayesian approach to “nonparametric” estimation and inference, in particular, one that induces a prior over  $\mathcal{M}$ . For each parameter  $\theta \in \Theta$ , a specification of  $\ell(\cdot | \theta)$  determines an element of  $\mathcal{M}$ . Given this mapping from  $\Theta$  into  $\mathcal{M}$ , a prior distribution  $\pi_o$  over  $\Theta$  implies a corresponding distribution over  $\mathcal{M}$ . This procedure necessarily assigns prior probability zero to a substantial portion of the space  $\mathcal{M}$ . Specifying a prior over the infinite dimensional space  $\mathcal{M}$  brings challenges associated with all nonparametric methods, including “nonparametric Bayesian” methods that must assign probability one to what is called a “meager set.” A meager set is defined topologically as a countable union of nowhere dense sets and is arguably small within an infinite-dimensional space.<sup>33</sup> This conclusion carries over to situations with families of priors that are absolutely continuous with respect to a baseline prior, as we have here. To us, prior robustness of this form is interesting, although it is distinct from robustness to potential model misspecifications. Indeed,

<sup>32</sup>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 Denti and Pomatto (2022)’s axiomatic construction of a parameterized likelihood to be used in Klibanoff et al. (2005) preferences.

<sup>33</sup>Sims (2010) critically surveys an extensive statistical literature on this issue. Foundational papers are Freedman (1963), Sims (1971), and Diaconis and Freedman (1986).

the section 5.1 decision maker who is concerned about model misspecification does not restrict himself to priors that are absolutely continuous with respect to a baseline prior because doing so would exclude many probability distributions he is concerned about.

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 (14) for the section 5.1 way of not knowing the distribution  $d\tau(w)$  and (22) for the section 5.2 way of not knowing a prior.

## 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:

$$\Pi = \{\pi : d\pi(\theta) = n(\theta)d\pi_o(\theta), n \in \mathcal{N}, \int_{\Theta} \phi[n(\theta)]d\pi_o(\theta) \leq \kappa\} \quad (24)$$

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

$$\min_{\pi \in \Pi} \int_{\Theta} \left( \int_W u[\gamma(w)]d\tau(w | \theta) \right) d\pi(\theta). \quad (25)$$

**Remark 5.10.** *The minimized objective for problem (25) can again be evaluated using convex duality theory via*

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

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

## 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  $c(w) = \gamma(w)$
- Utility function  $u[c(w)] = \log[c(w)]$ , where  $c(w)$  is consumption
- Prize rule  $\gamma(w) = \exp(\gamma_0 + \gamma_1 w)$

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

$$\gamma_0 + \gamma_1\mu_0 - \frac{1}{2\xi}(\sigma_0\gamma_1)^2$$

Larger values of the positive scalar  $\xi$  call for smaller adjustments  $-\frac{1}{2\xi}(\sigma_0\gamma_1)^2$  of expected utility  $\gamma_0 + \gamma_1\mu_0$  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  $\sim \mathcal{N}(\mu_i, \sigma_i^2), i = 1, \dots, \ell$ , where potential parameter values (states) are  $\theta_i = (\mu_i, \sigma_i)$  and parameter space  $\Theta = \{\theta_i : i = 1, 2, \dots, k\}$
- Baseline prior over structured models is a uniform distribution  $\pi_o(\theta_i) = \frac{1}{k}, i = 1, \dots, \ell$

To obtain an alternative prior  $\pi_i$  for  $i = 1, \dots, \ell$ , 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  $\theta_i$  is

$$\int u[\exp(\gamma_0 + \gamma_1 w)] d\tau(w | \theta) = \gamma_0 + \gamma_1 \mu_i$$

and a statistical divergence is

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

A subsection 5.2 decision maker with variational preferences orders prize rules  $\gamma(w) = \exp(\gamma_0 + \gamma_1 w)$  according to

$$\min_{\pi_i \geq 0, \sum_{i=1}^k \pi_i} \gamma_0 + \gamma_1 \sum_{i=1}^k \pi_i \mu_i + \frac{\xi}{k} \sum_{i=1}^k \phi(k\pi_i).$$

For a relative entropy divergence, prize rules are ordered by

$$-\xi \log \sum_{i=1}^k \left(\frac{1}{k}\right) \exp\left[-\frac{1}{\xi}(\gamma_0 + \gamma_1 \mu_i)\right] = \gamma_0 - \xi \log \sum_{i=1}^k \left(\frac{1}{k}\right) \exp\left(-\frac{\gamma_1 \mu_i}{\xi}\right)$$

and the associated minimizing  $\pi_i$  is

$$\frac{\exp\left(-\frac{\gamma_1 \mu_i}{\xi}\right)}{\sum_{i=1}^k \exp\left(-\frac{\gamma_1 \mu_i}{\xi}\right)}$$

## 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 alternative relative densities in  $\mathcal{M}$  to capture uncertainty about models and probability perturbations in the form of alternative 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 we did in section 5.3 where we form

$$\ell(w | \theta) \in \mathcal{M} \quad \forall \theta \in \Theta.$$

We represent a family of structured models as

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

where  $\tau_o(w)$  is now used to represent the family of structured models. The probability measure  $d\tau_o$  does not itself have to be a structured model.<sup>34</sup>

---

<sup>34</sup>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.

Let  $\pi_o(\theta)$  is a baseline prior over  $\theta$ . 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  $\theta$  have been rescaled so that

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

To acknowledge misspecification of a model implied by parameter  $\theta$ , 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  $\theta$  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  $\hat{m}(w | \theta)\ell(w | \theta)d\tau_o(w)$  over  $W$  conditioned on  $\theta$ .

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

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

The  $\mathbb{T}_1$  operator maps prize rules  $\gamma$  into functions of  $\theta$ . We use this for both hybrid approaches.

## 6.1 First hybrid model

We can rank alternative prize rules  $\gamma$  by including the following adjustment for possible misspecification of the baseline prior  $\pi_o$ :

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

Here  $\phi_1$  and  $\phi_2$  are possibly distinct convex functions with properties like the ones that we imposed on  $\phi$  in section 4.

Such a two-step adjustment for possible misspecification leads to an implied one-step variational representation with a composite divergence that we can define in the following way. For  $\hat{m} \in \widehat{\mathcal{M}}$  and  $n \in \mathcal{N}$ , form a composite scaled statistical discrepancy

$$d(\hat{m}, n) = \xi_1 \int_{\Theta} \left( \int_W \phi_1[\hat{m}(w | \theta)] d\ell(w | \theta) \right) n(\theta)d\pi_o(\theta) + \xi_2 \int_{\Theta} \phi_2[n(\theta)] d\pi_o(\theta) \quad (26)$$

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

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

In Appendix A we establish that divergence (26) is convex over the family of probability measures that concerns the decision maker.

**Remark 6.1.** *As noted earlier, Cerreia-Vioglio et al. (2013) posit a state space that includes parameters but also can include what we call shocks. Thus, think of the state as the pair  $(w, \theta)$ . In this setting, one could apply a statistical divergence to a joint distribution over possible realizations of  $(w, \theta)$ . Since the joint*

distribution can be factored into the product of a distribution over  $W$  conditioned on  $\theta$  and a marginal distribution over  $\Theta$ , such an approach can capture robustness in the specification of both  $\ell$  and  $\pi_o$ , albeit in a very specific way. For instance, for the relative entropy divergence, this results in the joint divergence measure:

$$d(\hat{m}, n) = \xi_1 \int_{\Theta} \left[ \int_W \hat{m}(w | \theta) \log \hat{m}(w | \theta) d\ell(w | \theta) \right] n(\theta) d\pi_o(\theta) + \xi_2 \int_{\Theta} n(\theta) \log n(\theta) d\pi_o(\theta)$$

for  $\xi_1 = \xi_2$ .

In earlier work, we have demonstrated important limits to such an approach in dynamic settings.<sup>35</sup> As we have shown here, we find both robustness to model misspecification and robustness to prior specification to be interesting in their own rights and see little reason to group them into a single  $\phi$  divergence.

## 6.2 Second hybrid model

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

$$\int_{\Theta} \phi_2[n(\theta)] d\pi_o(\theta) \leq \kappa \tag{27}$$

so that a decision maker's preferences over prize rules  $\gamma$  would be ordered by:

$$\min_{n \in \mathcal{N}} \int_{\Theta} \mathbb{T}_1[\gamma](\theta) n(\theta) d\pi_o(\theta), \tag{28}$$

where minimization is subject to (27).

As in Cerreia-Vioglio et al. (2022), preferences ordered by (28) subject to constraint (27) can be thought of as using a divergence between a potentially misspecified probability distribution and a set of predictive distributions that have been constructed from priors over a parameterized family of probability densities within the constrained set  $\Theta$ .<sup>36</sup> Notice how the first term in discrepancy measure (26) uses a prior  $n d\pi_o$  to construct a weighted averaged over  $\theta \in \Theta$  of the following conditioned-on- $\theta$  misspecification measure

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

The objective in problem (28) 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 indexed by  $\theta$  subject to the constraint that  $\pi \in \Pi$ .<sup>37</sup> Equivalently, in place of (26), this approach uses cost function

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

**Remark 6.2.** *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 those described in our discussions of robust prior analysis and potential model misspecification in remarks 5.3, 5.4, and 5.5.*

<sup>35</sup>See Hansen and Sargent (2007), Hansen and Sargent (2011), and Hansen and Miao (2018).

<sup>36</sup>Cerreia-Vioglio et al. (2022) 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.

<sup>37</sup>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.

## 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. Along with a robust prize rule, the minimum problem produces a worst-case probability distribution that rationalizes that prize rule. Strictly speaking, the decision theory tells us that particular values of cost function parameters  $(\xi_1, \xi_2)$  reflect a decision maker’s concerns about uncertainty, broadly conceived. In the spirit of Good (1952), it can be enlightening to study how worst-case distributions depend on  $(\xi_1, \xi_2)$ . The concluding paragraph of Chamberlain (2020) recommends exploring sensitivities with respect to a likelihood and with respect to a prior. Sensitivity of worst-case distributions to  $(\xi_1, \xi_2)$  provides evidence about the forms of subjective uncertainty and potential model misspecification that *should* be of most concern. That can provide both decision makers and outside analysts better understandings of the consequences of uncertainty aversion.

Motivated partly by a robust Bayesian approach, we have used decision theory to suggest a new approach to uncertainty quantification. By varying the aversion parameters  $(\xi_1, \xi_2)$ , we can trace out two-dimensional representations of prize rules and worst-case probabilities. A 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 associated expected utilities by studying how  $(\xi_1, \xi_2)$  trace out a 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 accompanying robust prize rules for responding to those uncertainties.

## 8 Concluding remarks

We have confined ourselves to a “static” setting and so have worked within the framework created by Maccheroni et al. (2006a). We intend this as a prolegomenon to another paper that will analyze related 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 will recur in that framework, additional issues such as dynamic consistency and appropriate state variables for recursive formulations of preferences will also arise.



## A Convexity of composite divergence

To verify convexity of (26), consider two joint probability measures on  $W \times \Theta$ :

$$\begin{aligned}\hat{m}_0(w | \theta)\ell(w | \theta)d\tau_o(w)n_0(\theta)d\pi_o(\theta) \\ \hat{m}_1(w | \theta)\ell(w | \theta)d\tau_o(w)n_1(\theta)d\pi_o(\theta).\end{aligned}$$

A convex combination of these two probability measures is itself a probability measure. Use weights  $1 - \alpha$  and  $\alpha$  to construct a convex combination and then factor it in the following way. First, compute the marginal probability distribution for  $\theta$  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  $\phi_2$ , it follows that

$$\phi_2[n_\alpha(\theta)] \leq (1 - \alpha)\phi_2[n_0(\theta)] + \alpha\phi_2[n_1(\theta)]. \quad (29)$$

Next note that

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

By the convexity of  $\phi_1$

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

Thus,

$$\phi_1[\hat{m}_\alpha(w | \theta)]n_\alpha(\theta) \leq (1 - \alpha)n_0(\theta)\phi_1[\hat{m}_0(w | \theta)] + \alpha n_1(\theta)\phi_1[\hat{m}_1(w | \theta)]. \quad (30)$$

Multiply (30) by  $\xi_1$  and (29) by  $\xi_2$ , add the resulting two terms, and integrate with respect to  $\ell(w | \theta)d\tau_o(w)d\pi_o(\theta)$  to verify that divergence (26) is indeed convex in probability measures that concern the decision maker.

## References

- Andrews, Isaiah and Jesse M Shapiro. 2021. A model of scientific communication. *Econometrica* 89 (5):2117–2142.
- 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.
- Bonhomme, Stéphane and Martin Weidner. 2021. Minimizing sensitivity to model misspecification. *arXiv preprint arXiv:1807.02161* .
- Brock, William A, Steven N Durlauf, and Kenneth D West. 2003. West. 2003. “Policy Evaluation in Uncertain Economic Environments.”. *Brookings Papers on Economic Activity* 1 (34):235–322.
- . 2007. Model uncertainty and policy evaluation: Some theory and empirics. *Journal of Econometrics* 136 (2):629–664.
- Bucklew, James A. 2004. *An Introduction to Rare Event Simulation*. New York: Springer Verlag.
- Cerreia-Vioglio, Simone, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio. 2012. Probabilistic sophistication, second order stochastic dominance and uncertainty aversion. *Journal of Mathematical Economics* 48:271–283.
- . 2013. Ambiguity and Robust Statistics. *Journal of Economic Theory* 148 (3):974–1049.
- Cerreia-Vioglio, Simone, Lars Peter Hansen, Fabio Maccheroni, and Massimo Marinacci. 2022. Making Decisions under Model Misspecification. Available at SSRN.
- Chamberlain, Gary. 2000. Econometric applications of maxmin expected utility. *Journal of Applied Econometrics* 15 (6):625–644.
- . 2001. Minimax Estimation and Forecasting in a Stationary Autoregression Model. *American Economic Review* 91 (2):55–59.
- . 2020. Robust Decision Theory and Econometrics. *Annual Review of Economics* 12:239–271.
- Christensen, Timothy and Benjamin Connault. 2019. Counterfactual sensitivity and robustness. *arXiv preprint arXiv:1904.00989* .
- Christensen, Timothy, Hyungsik Roger Moon, and Frank Schorfheide. 2020. Robust Forecasting. *arXiv preprint arXiv:2011.03153* .
- Christensen, Timothy M. 2018. Dynamic Models with Robust Decision Makers: Identification and Estimation. *arXiv preprint arXiv:1812.11246* .
- Del Negro, Marco and Frank Schorfheide. 2009. Monetary policy analysis with potentially misspecified models. *American Economic Review* 99 (4):1415–50.

- 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.
- Ellsberg, Daniel. 1961. Risk, Ambiguity, and the Savage axioms. *The Quarterly Journal of Economics* 643–669.
- 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.
- Ferguson, Thomas S. 1967. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press.
- 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).
- Gilboa, Itzhak, Stefania Minardi, Larry Samuelson, and David Schmeidler. 2020. States and contingencies: How to understand savage without anyone being hanged. *Revue Economique* 71:365–385.
- Good, Irving J. 1952. Rational Decisions. *Journal of the Royal Statistical Society. Series B (Methodological)* 14 (1).
- Hansen, Lars Peter and Jianjun Miao. 2018. Aversion to Ambiguity and Model Misspecification in Dynamic Stochastic Environments. *Proceedings of the National Academy of Sciences* 115 (37):9163–9168.
- Hansen, Lars Peter and Thomas J. Sargent. 2001. Robust Control and Model Uncertainty. *The American Economic Review* 91 (2):60–66.
- . 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.
- Hansen, L.P. and T.J. Sargent. 2011. Robustness and Ambiguity in Continuous Time. *Journal of Economic Theory* 146:1195–1223.
- Herstein, I. N. and John Milnor. 1953. An Axiomatic Approach to Measurable Utility. *Econometrica* 21 (2):291–297.
- Jacobson, David H. 1973. Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Differential Games. *IEEE Transactions for Automatic Control* AC-18:1124–1131.
- 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.
- Onatski, Alexei and James H Stock. 2002. Robust monetary policy under model uncertainty in a small model of the US economy. *Macroeconomic Dynamics* 6 (1):85–110.
- Petersen, Ian R., Matthew R. James, and Paul Dupuis. 2000. Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints. *IEEE Transactions on Automatic Control* 45:398–412.
- Savage, L. J. 1952. An Axiomatic Theory of Reasonable Behavior in the face of Uncertainty. Statistical research center, University of Chicago.
- . 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.

- Stock, James H and Mark W Watson. 2006. Forecasting with many predictors. *Handbook of economic forecasting* 1:515–554.
- 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.
- Wald, Abraham. 1947. An essentially complete class of admissible decision functions. *The Annals of Mathematical Statistics* 549–555.
- . 1949. Statistical decision functions. *The Annals of Mathematical Statistics* 165–205.
- . 1950. *Statistical Decision Functions*. New York: John Wiley and Sons, Inc.
- Whittle, Peter. 1990. *Risk-Sensitive Optimal Control*. New York: John Wiley & Sons.
- . 1996. *Optimal Control: Basics and Beyond*. John Wiley & Sons, Inc.