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

Lars Peter Hansen† and Thomas J. Sargent‡

October 3, 2023

Abstract

What are "deep uncertainties" and how should their presence influence prudent decisions? To address these questions, we bring ideas from robust control theory into statistical decision theory. Decision theory has its origins in axiomatic formulations by von Neumann and Morgenstern, Wald, and Savage. After Savage, decision theorists constructed axioms that formalize a notion of ambiguity aversion. Meanwhile, control theorists constructed decision rules that are robust to some model misspecifications. We reinterpret axiomatic foundations of decision theories to express ambiguity about a prior over a family of models along with concerns about misspecifications of the corresponding likelihood functions.

Keywords— deep uncertainty, ambiguity, misspecification, variational preferences, statistical divergence, relative entropy

JEL Codes— C10, C14, C18

*Forthcoming in the *Journal of Applied Econometrics*. This paper was prepared for the *Journal of Applied Econometrics Lecture* presented at the Association for Applied Econometrics Meetings held in Oslo in June of 2023. We thank Marco Del Negro, Tommaso Denti, Eric Ghysels, Paul Ho, Marco Loseto, Luciano Pomatto, Giulio Principi, Doron Ravid, Nathan Srebro, Mark Watson, and participants of the Econometrics Advising Group at the University of Chicago for criticisms of earlier drafts. We are especially grateful for comments by Simone Cerreia-Vioglio, Fabio Angelo Maccheroni, and Massimo Marinacci that taught us about recent advances in decision theory that can help applied statisticians. Diana Petrova, Pengyu Chen, Haomin Qin, Judy Yue, 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. The authors have no conflicts of interest to declare that are pertinent to the topic of this paper."

†University of Chicago. Email: lhansen@uchicago.edu

‡New York University. Email: thomas.sargent@nyu.edu

1 Introduction

Climate scientists confront "deep uncertainties."[1] Practicing econometricians also often struggle with uncertainty about their statistical models, but usually with scant guidance from significant advances in decision theory made after Wald (1947, 1949, 1950), Savage (1954), and Ellsberg (1961) because so much recent formal theory of decision making under uncertainty in economics is not cast explicitly in terms of the likelihoods and priors that are the foundations of statistics and econometrics.[2] Likelihoods are probability distributions conditioned on parameters while priors describe a decision maker's subjective belief about parameters.[3] By distinguishing roles played by likelihood functions and subjective priors over parameters, this paper aims to bring contributions to decision theory after Wald and Savage into closer contact with statistics and econometrics in ways that can address practical econometric concerns about model misspecifications and selections 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. His ultimate focus was on uncertainty about predictive distributions that are constructed by integrating likelihoods with respect to priors. Our paper instead formulates a decision theory with distinct uncertainties about priors and likelihoods. Cerreia-Vioglio et al. (2013) (section 4.2) provide a rationalization of the smooth ambiguity preferences proposed by Klibanoff et al. (2005) that includes likelihoods and priors as components. Denti and Pomatto (2022) extend this approach by using an axiomatic revealed preference approach to deduce an implied parameterization of a likelihood function. But neither of those papers sharply distinguishes prior uncertainty from concerns about possible model misspecifications, something that we want to do. We formulate concerns about model misspecifications as uncertainty about likelihoods.

We assemble 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 ideas 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 *parameters* that index alternative statistical models of random variables that affect outcomes that a decision maker cares about. By modifying Gilboa et al. (2010), Cerreia-Vioglio et al. (2013), and Denti and Pomatto (2022), we show how to use variational preferences to represent uncertainty about priors and concerns about statistical model misspecifications.

Some "behavioral" models in economics and finance assume expected utility preferences in which an agent's subjective probability differs systematically from probabilities that govern the data.[4] This literature

[1] Deep uncertainties are defined and discussed by Hallegatte et al. (2012), Maier et al. (2016), Marchau et al. (2019), and Rising et al. (2022).

[2] 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) and Ho (2023) used a post Wald-Savage decision theory of Gilboa and Schmeidler (1989) to confront model uncertainty in his econometric work.

[3] The term likelihood can have multiple meanings. We shall use it to represent a probability density of prize-relevant outcomes, which we refer to as repercussions, 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.

[4] We put "behavioral" in quotes to emphasize that most economic models are about agents' behaviors, including

also contains discussions of differences among agents in their confidence in their view of the world. Lack of confidence can take different forms under different notions of uncertainty. 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 to attach to alternative statistical models. Our representations of preferences provide ways to characterize degrees of confidence in terms of perceived statistical plausibilities.[5]

Objects and Interpretations

Our decision maker knows a parameterized family of probability distributions $\tau(w|\theta)d\upsilon(w)$, where $w \in W$ is a realization of a random vector or "repercussion" that he cares about, $\theta \in \Theta$ is a vector of parameters, and $d\upsilon(w)$ is a measure over W. A realization of w can play two possible roles. It can represent an outcome over which the decision maker has preferences, and it can capture data available to help the decision maker shape decisions. The decision maker has preferences over a set of prize rules, each of which we represent as a function $\gamma : W \times \Theta \to X$, where $x \in X$ is a "prize" that he cares about. In our featured examples, for parameter vector $\theta \in \Theta$, the prize rule $\gamma(w \mid \theta)$ determines the decision maker's exposure to an uncertain random vector that has a realization expressible in terms of $w \in W$ A set of γ's describes prize rules under consideration. In forecasting problems of a type common in time series statistics and econometrics, the prize can depend directly on the error in forecasting a component of w and the forecast rule depends on another component of w. While forecasting problems are interesting in their own right, in many applications, forecasts are intermediate inputs into outcomes of ultimate interest to the decision maker. Examples that appear in Section 2 illustrate a range of applications.[6]

The parameter space Θ can be finite or infinite dimensional; $\tau(w|\theta)$ is a member of a family of densities with respect to a measure $d\upsilon$ indexed by $\theta \in \Theta$. When Θ is infinite dimensional, we say that $\tau(w|\theta)d\upsilon(w)$ for $\theta \in \Theta$ is a "nonparametric" family of probability distributions. A notion of "non-informativeness" of a set of possible "prior" probability distributions over Θ plays an important role in justifying alternative approaches to "robustness" that we describe in section 3. A decision determines a pair (γ, τ). In section 2, we offer a more fully articulated decision process and provide some examples.

We use three components from decision theory, namely, i) states, ii) acts, and iii) prizes, in some new ways. We follow Anscombe and Aumann (1963) by defining consequences as lotteries over prizes. An act maps states into consequences. Preferences are defined over acts. In the static setup of this paper, we take the state to be parameters 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". We are instead interested in a typical econometric situation in which parameters of statistical models remain hidden forever. For us, parameter uncertainty is central, so it is important that parameter vector θ be included as a component of the state.[7]

Gilboa et al. (2010) and Cerreia-Vioglio et al. (2013) introduced parameterized models as a family of

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.

[5] 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.

[6] As a different example, that section illustrates the "parameter estimation problem" as a special case.

[7] 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."

primitive probabilities that a decision maker cares about. Cerreia-Vioglio et al. (2013) in effect consider an expanded state space (w, θ) that includes both repercussions with realization w and parameters θ and then take a *model* to be a conditional distribution over $(W, \mathfrak{W})$ given θ.[8] 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 an act is preferred to another act if it is preferred under all models in the family.

Relative 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) interpret lotteries as "roulette wheels" with known (objective) probabilities, in contrast to "horse races" with unknown (subjective) probabilities. Many authors used an Anscombe and Aumann setup as a vehicle to extend Von Neumann and Morgenstern (1944) preferences defined over lotteries to more general settings that can include subjective uncertainty. In our formulation, the random vector W induces a probability distribution that according to a particular Anscombe and Aumann act implies a particular lottery that can depend on parameters of a statistical model. We represent a family of models as a family of probability distributions indexed by an unknown parameter vector. Parameter vectors can reside in a finite set, a manifold of possible values, or even an infinite dimensional set. With correct statistical models (i.e., likelihoods), each model induces a "roulette wheel" lottery. The possibility of misspecified likelihoods leads us to want a counterpart to an Anscombe-Aumann lottery with unknown probabilities. Our extension of 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 to put over those statistical models. We formulate preferences that express distinct concerns about both types of robustness.

To motivate their axioms, Maccheroni et al. (2006a) and Strzalecki (2011) used Hansen and Sargent's (2001) stochastic formulation of a robust control problem. We use our Anscombe and Aumann formulation to show that the axioms of Maccheroni et al. and Strzalecki actually express prior uncertainty and not the model misspecification concerns that had originally motivated Hansen and Sargent. We go on to 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 selection, including priors meant to represent "nonparametric Bayesian" methods.

Section 2 sets the stage by reviewing axioms that support Anscombe and Aumann's subjective expected utility representation. Subsection 2.4 discusses the extension to max-min utility justified by Gilboa and Schmeidler (1989), while subsection 2.5 tells how Maccheroni et al. (2006a) relaxed the Gilboa and Schmeidler axioms to arrive at variational preferences. Section 3 characterizes a class of variational preferences that use statistical divergences as Maccheroni et al. cost functions. Section 4 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 4.1 decision maker knows a parameterized family of models but seeks robustness with respect to a set of alternative priors to put over those models, while a subsection 4.2 decision maker has a unique prior but a likelihood function that he distrusts and consequently wants robustness with respect to statistically nearby models. After comparing and contrasting these two decision makers in subsection 4.3, subsection 4.4 provides two example of these alternative types of robustness. Section 5 breaks new ground by describing cost functions to use for a variational preferences

[8] 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 approaches are interesting, we suppose that models can have scientific or other sources from outside the specific decision problem. In this, we follow Hansen and Sargent (2022) who refer to such models as "structured models."

representation of a decision maker who is concerned about *both* types of robustness. Section 6 sketches dynamic extensions of our analysis that we want to study in subsequent papers. Section 7 briefly steps outside our decision theory to discuss how someone might want to assess "cost" parameters that characterize a decision maker's variational preferences. Section 8 explores connections to an approach from statistical learning called PAC (probably almost correct) Bayesian analysis. Section 9 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.[9] 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* $f : \Theta \to \Lambda$.

For a given θ, $df(x \mid \theta)$ denotes integration with respect to the probability measure $f(\theta) \in \Lambda$, which is a lottery over possible prizes $x \in X$.[10] For a *given* probability measure $\pi \in \Pi$, $\int_\Theta df(x \mid \theta) d\pi(\theta)$ is a two-stage lottery over prizes, with a lottery over states θ being described by π and another lottery over prizes $x \in X$ being described by $df(x \mid \theta)$, which depends on the outcome θ from the other lottery. We shall introduce uncertainty about probability measure π.

As mentioned in section 1, we interpret objects in the Anscombe and Aumann formulation in ways that help us as statisticians/econometricians. We interpret a state θ as pinning down one among a set Θ of probability models that a decision maker regards as possible. A decision maker make a decision (i.e., "chooses an Anscombe and Aumann act") that generates a probability distribution over outcomes that he/she cares about, i.e., over Anscombe and Aumann prizes $x \in X$.

We use Anscombe and Aumann acts to represent alternative conditional distributions of repercussions and prize rules. An action or decision $\delta \in \Delta$, which is distinct from an Anscombe and Aumann act, can be a vector of real numbers or, more generally, a function that is defined on appropriate spaces. A choice of δ can influence the distribution of repercussions conditioned on the parameter vector. It can also alter the exposure of the prize to repercussions. We represent a decision maker's exposure to repercussions with prize rules γ_δ that are Borel measurable functions that map W into prizes in X. We represent the influence of δ on the distribution of repercussions by a conditional probability measure represented as a density $\tau_\delta(\cdot \mid \theta)$ with respect to a Borel measure υ on $(W, \mathcal{W})$. A $\theta \in \Theta$ implies a probability measure

$$\tau_\delta(w \mid \theta) d\upsilon(w).$$

This formulation is convenient for applied statisticians because for each δ, a parameterized family $\tau_\delta(\cdot \mid \theta)$ can define a manifold of likelihoods indexed by a vector of unknown parameters $\theta \in \Theta$. For a given decision δ, $(\gamma_\delta, \tau_\delta)$ induces a lottery over X conditioned on θ, and hence an Anscombe and Aumann act that can be represented with $df(x \mid \theta)$. As we vary decisions $\delta \in \Delta$, we trace out a collection of such acts. A particular

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

[10] We borrow our basic setup 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, much of probability theory routinely imposes countable additivity. It simplifies our presentation.

decision problem defines both conditional distributions $\tau_\delta dv$ and prize rules γ_δ for alternative decisions δ. Together, they delineate a collection of Anscombe and Aumann acts.

Remark 2.2. *We can expand the collection of acts by randomizing decisions. Given two decisions δ_1 and δ_2, a randomized rule chooses decision δ_1 with probability α and δ_2 with probability $1-\alpha$. Since each decision induces an Anscombe and Aumann act, the randomized decision is a convex combination of the two induced acts.*[11]

We consider several canonical examples.

Example 2.3. *For some situations, it suffices to let Δ be a Borel set of a finite-dimensional Euclidean space and for the conditional distribution τ not to depend on δ. For example, δ could be a particular portfolio of assets whose random return is exposed to a repercussion vector in a particular way. A choice of a portfolio does not affect the joint distribution of returns on individual assets, but it does influence the return on a portfolio of those component assets.*

Example 2.4. *In stochastic optimal control problems like those studied by Bertsekas (1976), a decision maker chooses a "control" that affects the distribution of a repercussion, which in this example takes the form of a next-period state vector. For instance, in linear-quadratic Gaussian optimal control problems, often referred to as "linear regulator" problems, this effect shows up in a mean conditioned on a current state. For example, a repercussion vector w obeys:*

$$w = \mathbb{A} + \mathbb{B}\delta + \mathbb{C}\epsilon,$$

where the probability distribution over ϵ's is a standard, multivariate normal. The vector $\mathbb{A}$ and the matrices $\mathbb{B}$ and $\mathbb{C}$ depend on parameters in Θ.[12] *Suppose that a controller who chooses δ knows parameters only up to an uncertain subjective distribution.*[13] *Think of decision δ as a current period control vector in the sense of Bertsekas (1976). The conditional distribution τ depends on δ: w is distributed as multivariate normal with conditional mean $\mathbb{A} + \mathbb{B}\delta$ and conditional variance $\mathbb{C}\mathbb{C}^\top$. In typical optimal linear regulator control theory problems, prize rules depend on the vector $(w^\top, \delta^\top)$ with a utility function that is the negative of the quadratic form in this vector, for example, $-w^\top \mathbb{R} w - \delta^\top \mathbb{Q}\delta$ where R and Q are positive semidefinite matrices. The linear regular problem is an example of a much larger class of stochastic control problems. For simplicity, we formulate it as a static problem.*[14] *The Example 2.3 portfolio choice problem is a special case of this stochastic control problem in which repercussions are the returns and a control vector of portfolio weights does not influence repercussions.*

Example 2.5. *To build bridges to mathematical statistics, we extend a setup that Ferguson (1967) used to analyze learning from data. We again posit a family of densities $\tau(w \mid \theta)$ for a repercussion vector whose realizations are denoted by w's, and where a parameter vector θ is a "true state of nature." In this example, the decision δ does not affect τ. We can represent the outcome of what Ferguson calls a statistical experiment as a realization of a random vector $y = \zeta(w)$ that contains information about w. This information can be a signal that is correlated with the "prize" of ultimate interest. Let a decision δ be a measurable function that maps observations y from the statistical experiment into a set of what Ferguson calls* actions.[15] *In this way*

[11] 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.

[12] The vector $\mathbb{A}$ absorbs the current state. In a standard optimal linear regular problem, the controller knows $(\mathbb{A}, \mathbb{B}, \mathbb{C})$.

[13] This distribution might depend on past information.

[14] More generally, it would be input into a recursive formulation.

[15] For Ferguson (1967), δ viewed as a function of y is a *decision rule* distinct from our *prize rule*.

Ferguson allows what he calls an "action" – our decision δ – to depend on an observation y. We constrain a prize rule γ_δ to satisfy:[16]

$$\gamma_\delta(w) = \Psi[\delta \circ \zeta(w), w]. \tag{1}$$

Ferguson's actions are not Anscombe and Aumann (1963) acts. *To capture Ferguson's setup, each prize rule γ_δ implies a probability distribution for a prize conditioned on θ that is induced by $\tau(w \mid \theta)d\upsilon(w)$. This probability distribution for a prize conditioned on θ is an Anscombe and Aumann (1963) act.*

By allowing decisions to depend on data that is observed at an intermediate date, the Ferguson formulation allows a richer collection of possible decisions and nests our Remark 2.3 formulation as a special case. It can include problems that seek to construct robustly optimal forecasts from historical data. More generally, we are interested in decision problems for which forecasting is an input but not the ultimate goal.

Although the problem posed in Example 2.5 is static, it can be reinterpreted as a three-stage or three-period decision problem. A decision rule that is chosen at an initial period zero depends on information about the repercussion that will be revealed in a first stage. The decision maker can condition on this information when choosing his exposure to the repercussion with realization w. The repercussion itself is fully realized at the end of stage two. In this formulation, potential likelihood misspecifications affect the decision maker's inferences about the prize distribution in stage one. As posed, this is an *ex-ante* decision problem in which a decision rule, δ, is chosen at period 0. In contrast, we can view Examples 2.3 and 2.4 as *ex post* problems in which the "prior" implicitly conditions on current and past data as does the "decision." As often happens, the timing protocol matters. When a decision maker chooses sequentially, the distinction between priors and posteriors can become obscured when an end of period $j-1$ posterior becomes a period j prior. In a recursive formulation of a dynamic decision problem, concerns about robustness of priors-posteriors can recur in stage-specific components within a multi-stage interpretation of a decision problem. By design, our general formulation invites dynamic extensions.

Example 2.6. *It is common in econometrics and statistics to pose a decision problem as a parameter estimation problem that supposes that prizes are deviations between an estimator $\delta(y)$ and a function $\chi(\theta)$ of the parameter vector. To capture this, we can let*

$$w = \begin{bmatrix} \chi(\theta) \\ y \end{bmatrix}$$

and add a degenerate equation to the τ dynamics that describes how we construct the first component of w. This approach seems to be shorthand for something deeper. Typically, decisions of interest can be expressed in terms of outcomes with probability distributions that depend on the unknown parameter as in the other examples that we mention.

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

Remark 2.7. *The collection of Anscombe and Aumann acts is typically much larger than the set of acts that can be induced by an available pair, $(\gamma_\delta, \tau_\delta)$ for $\delta \in \Delta$ as implied by alternative decisions. 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 previously enlarges the set of Anscombe and Aumann acts by*

[16] 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, θ) under the distribution implied by $\tau(w \mid \theta)d\upsilon(w)$ and ζ.

including the convex hull of the set of acts induced by prize rules, in general that device does 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 would require extending the set of the acts to be studied beyond the set of induced by the potential actions within a "substantive decision model" even if allow randomization of the decisions.

Let $\mathcal{A}$ be the set of all acts. Each act $f \in \mathcal{A}$ implies lotteries $f(\theta)$ for each $\theta \in \Theta$. 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 θ. Absence of dependence means that 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 γ could emerge as follows. Suppose that some component of W has a known distribution independent of θ and that γ depends only on this component. Such limited dependence implies an act that is independent of θ. 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 $\mathcal{A}_s$.[17] We use a probability distribution $\pi \in \Pi$ over states in conjunction with $\mathcal{A}_s$ to express subjective uncertainty.

Remark 2.8. *Anscombe and Aumann (1963) distinguished "horse race lotteries," represented by acts in $\mathcal{A}_s$, from "roulette lotteries," represented by acts in $\mathcal{A}_o$.*[18]

Remark 2.9. *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 benefits and challenges that this flexibility brings.*[19] *There is also flexibility in constructing an act. Cerreia-Vioglio et al. (2012) take advantage of this flexibility to 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.*[20]

We shall often construct a new act from initial acts f and g by using a probability $\alpha \in (0,1)$ 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 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, c) a decision maker with concerns that a single known statistical model θ is *misspecified* by using a statistical discrepancy measure to delineate 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, $\succsim$ a weak preference, and $\succ$ a strict preference. Throughout, we assume that preferences are non-degenerate (there is a strict

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

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

[19] They did not specifically discuss the statistical linkages that we explore here.

[20] More generally, their representation includes an additional curvature adjustment much like the smooth ambiguity model. See Proposition 3 in their appendix.

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.[21] 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

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

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:[22]

Axiom 2.10. *(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 $u : X \to \mathbb{R}$ such that for $f, g \in \mathcal{A}_o$

$$f \succsim g \iff U(f) \geqslant U(g) \tag{2}$$

where

$$U(f) = \int_X u(x) d\lambda(x) \tag{3}$$

and $\lambda \in \Lambda$ is the probability distribution generated by constant act f. Representation (3) 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 (3). 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 (3). According to representation (2), 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. We impose restrictions

[21] The Archimedean axiom states: let f, g, h be acts in $\mathcal{A}$ with $f \succ g \succ h$. Then there are $0 < \alpha_1 < 1$ and $0 < \alpha_2 < 1$ such that $\alpha_1 f + (1-\alpha_1)h \succ g \succ \alpha_2 f + (1-\alpha_2)h$. See Herstein and Milnor (1953, Axiom 2) for an alternative formulation of a continuity axiom.

[22] 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 $\succ_\Lambda$.

on $\succ$ in the form of two axioms. The first extends the independence axiom to the set of all acts:

Axiom 2.11. *(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 2.12. *(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. A set of acts implies an associated collection $\mathcal{B}$ of functions F. From monotonicity axiom 2.12 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 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 positive finite measure.[23] 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 $f \in \mathcal{A}$

$$f \succsim g \iff \int_\Theta \left[\int_X u(x) df(x \mid \theta) \right] d\pi(\theta) \geqslant \int_\Theta \left[\int_X u(x) dg(x \mid \theta) \right] d\pi(\theta), \tag{4}$$

[23] Riesz-type representation theorems provide such representations on the space of continuous functions with compact support on a locally compact Hausdorff space. The decision theory literature typically uses a different space. Our presentation here is informal and suggestive, but it is not intended to be a complete analysis.

where the probability measure π 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 π and then draw a prize $x \in X$ from probability distribution $df(x \mid \tilde{\theta})$. By changing the 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

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

where

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

Equation (6) constructs a single lottery λ over x from the compound lottery generated by $(d\pi(\theta), df(x \mid \theta))$.[24] For a statistician, λ is a "predictive distribution" constructed by integrating over unknown parameter θ. Let f_c be the constant act with lottery λ 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.[25]

2.4 Max-min Expected Utility

To construct a decision maker who has max-min expected utility preferences, Gilboa and Schmeidler (1989) replaced Axiom 2.11 with the following two axioms:

Axiom 2.13. *(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 2.14. *(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 α is known, an assumption that can be interpreted as describing a form of objective uncertainty. Axiom 2.14 asserts a weak preference for mixing with known weights α and $1-\alpha$.

Example 2.15. *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 lottery λ_2 if $\theta = \theta_2$. Let act g be lottery λ_2 if $\theta = \theta_1$ and lottery λ_1 if $\theta = \theta_2$. Suppose that $f \sim g$. Axiom 2.14 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.*

[24] Equation (6) thus expresses the "reduction of compound lotteries" described by Luce and Raiffa (1957, p. 26) and analyzed further by Segal (1990).

[25] Under subjective expected utility, the Remark 2.5 statistical decision problem solves

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

where $d\bar{\pi}(\theta \mid y)$ is the posterior of θ given an observation y. Consequently, a depends implicitly on y, which implies the decision rule $\delta(y)$.

By replacing Axiom 2.11 with Axioms 2.13 and 2.14, Gilboa and Schmeidler obtained preferences described by

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

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 θ, $\int_X u(x) df(x \mid \theta)$ 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 Π_c is a set with a single member.

2.5 Variational preferences

Maccheroni et al. (2006a) relaxed certainty independence Axiom 2.13 of Gilboa and Schmeidler (1989) to obtain preferences with a yet more general representation that they called variational preferences. Maccheroni et al. replaced Axiom 2.13 with the weaker

Axiom 2.16. *(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 2.16 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* α 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 2.16 within the Example 2.3 setting, use the pair (Ψ, τ) to represent the probabilistic outcomes of alternative decisions. Recall that for a given Ψ, prize rule $\gamma(w)$ is described by (1) for some decision δ and some parameterized repercussion distribution $\tau(w \mid \theta) d\upsilon(w)$. Let's compare the uncertainty consequences of decisions δ_1 and δ_2 that give rise to prize rules γ_{δ_1} and γ_{δ_2} via

$$\begin{aligned} \gamma_{\delta_1}(w) &= \Psi[\delta_1 \circ \zeta(w), w] \\ \gamma_{\delta_2}(w) &= \Psi[\delta_2 \circ \zeta(w), w] \end{aligned}$$

for $\delta_1, \delta_2 \in \Delta$. Each of these prize rules specifies how a prize depends on a realization w of the repercussion. Let γ_1 induce act f and γ_2 induce act g.

Now consider two other decisions δ_3 and δ_4 and use them to construct prize rules

$$\begin{aligned} \gamma_{\delta_3} &= \Psi[\delta_3 \circ \zeta(w), w] \\ \gamma_{\delta_4} &= \Psi[\delta_4 \circ \zeta(w), w]. \end{aligned}$$

Suppose that the prize rules γ_{δ_3} and γ_{δ_4} both of which induce distributions of the prize $x \in X$ that do not depend on θ; decisions δ_3 and δ_4 both serve to target risk components of $x \in X$ that are not exposed to parameter uncertainty. In particular, the dependence on the signal $y = \zeta(x)$ could be degenerate for both rules. Denote the constant acts induced by γ_{δ_3} and γ_{δ_4}, respectively, as h and k. For example, consider an investment problem for which some of the available investments (indexed by a subset of the decisions $\delta \in \Delta$)

yield returns that depend only on a component of the repercussion vector that has a known distribution. Two such investments can be used to construct γ_{δ_3} and γ_{δ_4}. Axiom 2.16 requires that if randomizing δ_1 with respect to δ_3 is preferred to randomizing δ_2 with respect to δ_3, the preference order will be preserved if δ_3 is replaced by δ_4 holding fixed the randomization probabilities $(\alpha, 1-\alpha)$.[26]

A specification of Ψ may exclude the possibility described in the previous paragraph. But the axioms refer to hypothetical comparisons. So to explore Axiom 2.16, we extend the substantive model Ψ to $\widetilde{\Psi}$ where the arguments of $\widetilde{\Psi}$ are a realization of an augmented repercussion vector $(w, \tilde{w})$ and the decision δ is in a larger set $\widetilde{\Delta}$ that contains Δ. Suppose that the $\tilde{w}$ component of the augmented repercussion vector has a known distribution that does not depend on θ. Prizes γ that depend only on this second component of representation (1) induce constant acts. To confirm that the $\tilde{\Psi}$ substantive model is an extension of the original Ψ model, we require that

$$\widetilde{\Psi}[\delta \circ \zeta(w), w, \tilde{w}] = \Psi[\delta(w) \circ \zeta, w] \qquad \text{for } \delta \in \Delta, w \in W.$$

We then suppose that

$$\begin{aligned} \gamma_{\delta_3}(\tilde{w}) &= \widetilde{\Psi}[\delta_3 \circ \zeta(w), w, \tilde{w}] \\ \gamma_{\delta_4}(\tilde{w}) &= \widetilde{\Psi}[\delta_4 \circ \zeta(w), w, \tilde{w}] \end{aligned}$$

where $\delta_3, \delta_4 \in \widetilde{\Delta}$ but not necessarily in $\delta_3, \delta_4 \in \Delta$. Decisions δ_3 and δ_4 confine exposure of the resulting prize to $\tilde{w}$ and not to w. As indicated in the previous paragraph, γ_3 and γ_4 induce constant acts. For our Remark 2.3 investment example, construction of the extended substantive model $\widetilde{\Psi}$ might introduce new opportunities that are not exposed to parameter uncertainty. This opens the door to comparisons entertained by Axiom 2.16.

Maccheroni et al. (2006a) showed that preferences that satisfy the weaker Axiom 2.16 instead of Axiom 2.13 are described by

$$f \succsim g \iff \min_{\pi \in \Pi} \int_\Theta \left[\int_X u(x) df(x \mid \theta) \right] d\pi(\theta) + c(\pi) \geqslant \min_{\pi \in \Pi} \int_\Theta \left[\int_X u(x) dg(x \mid \theta) \right] d\pi(\theta) + c(\pi) \qquad (8)$$

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 functions, c, express more aversion to uncertainty. The convex function c in variational preferences representation (8) replaces the restricted set of probabilities Π_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.

3 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 $(\Pi, \mathfrak{G})$. We construct statistical divergences for these two situations in similar ways.

We first consider repercussion distributions over $(W, \mathfrak{W})$. Consider a family of probabilities represented

[26] We thank Fabio Maccheroni and Massimo Marinacci for suggesting this formulation.

as densities with respect to υ:

$$\mathcal{L} := \left\{ \ell \geqslant 0 : \int \ell(w) d\upsilon(w) = 1 \right\} \tag{9}$$

For a baseline density ℓ_o, a *statistical divergence* is a convex function $D(\ell \mid \ell_o)$ of probability measures $\ell(w)d\upsilon(w)$ that satisfies

- $D(\ell \mid \ell_o) \geqslant 0$
- $D(\ell \mid \ell_o) = 0$ implies $\ell = \ell_o$

Given ℓ_o, write:

$$\ell(w)d\upsilon(w) = m(w)\ell_o(w)d\upsilon(w)$$

for $m = \frac{\ell}{\ell_o}$, where we assume m is not infinite with positive υ measure so that the probability measure ℓ is absolutely continuous with respect to $\ell_o(w)d\upsilon(w)$.[27] The set of such densities is convex as is the set of implied relative densities m. To define a scaled statistical divergence, we set

$$D(\ell \mid \ell_o) = \xi \int_W \phi[m(w)]\ell_o(w)d\upsilon(w),$$

where $\xi > 0$, and ϕ is 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 ϕ functions and the divergences that they lead to are

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

When $\xi = 1$, the divergence, D, is often called a ϕ or f-divergence. When $\phi(m) = m\log(m)$ and $\xi = 1$, we obtain relative entropy

$$D_{KL}(\ell|\ell_o) = \int_W m(w)\log[m(w)]d\upsilon(w).$$

Relative entropy is commonly referred to as Kullback-Leibler divergence.

Remark 3.1. *Other families of divergences can be used in conjunction with preference representations that follow, for instance, from Bregman and Wasserstein divergences. The family ϕ or f divergences featured here has very nice duality properties. as we will see, duality allows us to make formal connections to the extensive literature on smooth ambiguity. Furthermore, these divergences are invariant to one-to-one transformations of the space over which the probability distributions are defined. In addition, some members of this family have useful links to statistical discrimination procedures. The link to likelihood-based statistical discrimination enables statistical constructions that can help us calibrate concerns about robustness.*

4 Basic formulation

We associate a probability measure $\tau(w|\theta)d\upsilon(w)$ 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

[27] For ℓ's for which the implied m is infinite with positive υ measure, we define the divergence to be infinity.

$\gamma \in \Psi$ that map $w \in W$ into an $x \in X$. Think of γ as a prize rule and $\gamma(w)$ as an uncertain scalar prize. For each prize rule γ, let $d\lambda(x \mid \theta)$ be the distribution of the prize x that is induced by distribution $\tau(w|\theta)d\upsilon(w)$ and the prize rule γ. The distribution of the prize thus depends both on the prize rule $\gamma(w)$ and the distribution $\tau(w \mid \theta)d\upsilon(w)$. Within this setting, a decision δ gives rise to a specific pair $(\gamma_\delta, \tau_\delta)$. To avoid cluttering our notation, we will drop the explicit dependence of γ and τ on δ in much of the following discussion.

4.1 Not knowing a prior

Like the robust Bayesian decision maker of Berger (1984), 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.[28] We label such distrust of a single prior "model ambiguity." Here we describe a static version of what Hansen and Sargent (2021, 2022) call structured uncertainty. "Structured" refers to a particular way that we reduce the dimension of a set of alternative models relative to the much larger set considered when we explore likelihood or model misspecification.

A 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:

$$\mathcal{N} \doteq \left\{ n \geqslant 0 : n(\theta) \geqslant 0, \int_\Theta n(\theta) d\pi_o(\theta) = 1 \right\}. \tag{10}$$

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.[29] The decision maker is certain about each of the specific models but is uncertain about a prior to put over them.

4.1.1 Not knowing a prior, I

To express a form of ambiguity aversion, the decision maker uses scaled statistical divergence

$$c(\pi) = \xi \int_\Theta \phi\left[n(\theta)\right] d\pi_o(\theta) \tag{11}$$

and has variational preferences ordered by[30]

$$\min_{n \in \mathcal{N}} \int_\Theta \left(\int_W u[\gamma(w)]\tau(w \mid \theta) d\upsilon(w) \right) n(\theta) d\pi_o(\theta) + \xi \int_\Theta \phi[n(\theta)] d\pi_o(\theta). \tag{12}$$

Remark 4.1. *It is convenient to solve the minimization problem* (12) *by using duality properties of convex functions. Because the objective is separable in θ, we first compute*

$$\phi^*(\mathsf{u} \mid \xi) = \min_{\mathsf{n} \geqslant 0} \mathsf{u}\mathsf{n} + \xi\phi(\mathsf{n}) \tag{13}$$

[28] See Berger (1984) for a robust Bayesian perspective. 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.

[29] See Hansen and Sargent (2022).

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

where $\mathsf{u} = \int u[\gamma(w)]\tau(w|\theta)d\upsilon(w) + \eta$, n *is a nonnegative number, and* η *is a nonnegative real-valued Lagrange multiplier attached to the constraint* $\int_\Theta n(\theta)d\pi_o(\theta) = 1$*;* $\phi^*(\mathsf{u} \mid \xi)$ *is a concave function of* u.[31] *The minimizing value of* n *satisfies*

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

The dual to the minimization problem on the right side of (12) *is*

$$\max_{\eta} \int_\Theta \phi^* \left(u[\gamma(w)]\tau(w \mid \theta)d\upsilon(w) + \eta\right) d\pi_o(\theta) - \eta. \tag{14}$$

Remark 4.2. *(Smooth ambiguity preferences) When statistical divergence is scaled relative entropy, preferences over* $\gamma(w)$ *are ordered by*

$$-\xi \log\left[\int_\Theta \exp\left(-\frac{\int_W u[\gamma(w)]\tau(w \mid \theta)d\upsilon(w)}{\xi}\right) d\pi_o(\theta)\right], \tag{15}$$

a static version of preferences that Hansen and Sargent (2007) used to frame a robust dynamic filtering and control 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 express concerns about model misspecification.[32] *In contradistinction, the robustness concerns being represented in this subsection are about a baseline prior over known models and not about possible misspecifications of those models.*

4.1.2 Not knowing a prior, II

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

$$\Pi = \{\pi : d\pi(\theta) = n(\theta)d\pi_o(\theta), n \in \mathcal{N}, \int_\Theta \phi[n(\theta)]d\pi_o(\theta) \leqslant \kappa\} \tag{16}$$

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

$$\min_{\pi \in \Pi} \int_\Theta \left(\int_W u[\gamma(w)]\tau(w \mid \theta)d\nu(w)\right) d\pi(\theta). \tag{17}$$

Remark 4.3. *It is convenient to solve the minimization problem on the right side of* (17) *by using duality properties of convex functions. The minimized objective for problem* (17) *can again be evaluated using convex duality theory. We now explicitly note the dependence of* ϕ^* *on* ξ *and write the dual problem as:*

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

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

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

[32] 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.

Remark 4.4. *Within a setting like that of Example 2.6, Ho (2023) used another approach to compute robust adjustments to posterior expectations. He used this approach to assess the prior sensitivity of empirical measurements of targets of interest to an investigator. Ho's framework could also be used to define robust preferences defined in terms of posterior expectations. For instance, measurements of interest could be depicted as the maximizer of the negative of an expected loss function of a type common in statistics and econometrics. More formally, Ho used relative entropy divergence to restrict a set of priors. He computed expectations conditioned on a signal and minimized over possible implied posterior distributions given a relative entropy constraint over the priors. The minimizing "prior" from this approach typically depends on the signal,*[33] *unlike the outcome from solving the* ex-ante problem *described in Remark 2.5. Dependence of a minimizing prior on the signal like that in Ho's formulation also emerges in some recursive formulations of dynamic problems, a situation that can lead to statistically inadmissible decisions.*[34]

4.2 Not knowing a likelihood

Instead of being about a prior as the previous subsection, we now suppose that the decision maker's uncertainty is about a likelihood function. We start by supposing that there is a single model that the decision maker fears is misspecified. We then extend the analysis by introducing a parameterized family of probability models that a decision maker thinks might be misspecified.

4.2.1 A misspecified model

Consider first a single model that might be misspecified. We study a decision maker who knows a parameter θ_o. We also fix a decision δ, a determinant of τ that we continue to leave implicit in our notation. The decision maker entertains the possible misspecification of

$$\tau_o(w) := \tau(\cdot \mid \theta_o)$$

in ways that the decision maker cannot precisely 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.)

Notice that $\tau_o \in \mathcal{L}$ where $\mathcal{L}$ is given by (9). To formalize concerns that $\tau_o(w)$ is misspecified, we entertain the following set of repercussion probabilities:

$$m(w)\tau_o(w), \tag{18}$$

where

$$m(w) := \frac{\ell}{\tau_o}$$

$\ell \in \mathcal{L}$ for $\mathcal{L}$ given by (9). We represent the decision maker's ignorance of specific alternative models by assuming that he entertains a potentially infinite dimensional space $\mathcal{L}$ of what we will call "unstructured"

[33] See formula (2.7) in Ho (2023).

[34] See Epstein and Schneider (2003) for a formulation of a dynamic choice problem under ambiguity aversion that deploys multiple priors recursively. Hansen and Sargent (2022) described a possible tension between admissibility and dynamic consistency.

models. A decision maker's expected utility under alternative model $\ell\tau_o(w)dv(w)$ is

$$\int_W u[\gamma(w)]m(w)\tau_o(w)dv(w) = \int_W u[\gamma(w)]\ell(w)dv(w). \tag{19}$$

Notice that (19) evaluates expected utility for a single choice for m.

To complete a description of preferences, we require a scaled statistical divergence. We consider alternative probabilities parameterized by entries in $\mathcal{L}$. Under this perspective, a probability model corresponds to a choice of $m \in \mathcal{M}$. Form a scaled divergence measure:

$$c(m) = \xi \int_W \phi[m(w)]\tau_o(w)dv(w) \tag{20}$$

where $\xi > 0$ is a real number. Variational preferences that use (19) as expected utility over lotteries and (20) as scaled statistical divergence are ordered by

$$\min_{m=\frac{\ell}{\tau_o}, \ell\in\mathcal{L}} \left(\int_W u[\gamma(w)]m(w)\tau_o(w)dv(w) + \xi \int_W \phi[m(w)]\tau_o(w)dv(w) \right). \tag{21}$$

This formulation lets a decision maker evaluate alternative prize rules $\gamma(w)$ 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 $m(w)\tau_o(w)dv(w)$.

Remark 4.5. *As was the case for robust prior analysis, it is again convenient to solve the minimization problem on the right side of* (21) *by using duality properties of convex functions. Because the objective is separable in w, we can first compute*

$$\phi^*(\mathsf{u} \mid \xi) = \min_{\mathsf{m}\geqslant 0} \mathsf{u}\mathsf{m} + \xi\phi(\mathsf{m}) \tag{22}$$

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

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

The dual problem to the minimization problem on the right side of (21) *is*

$$\max_{\eta} \int_W \phi^*(u[\gamma(w)] + \eta)\tau_o(w)dv(w) - \eta. \tag{23}$$

Remark 4.6. *We posed minimum problem* (21) *in terms of a set of probability measures on the measurable space $(W, \mathfrak{W})$ with baseline probability $\tau_o(w)dv(w)$. Since the integrand in the dual problem* (23) *depends on w only through the control law γ, 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 $x = \gamma(w)$ with distribution $d\lambda_o(x)$. Doing that would lead to equivalent outcomes. Representations in sections 2 and 2.5 are all cast in terms of induced distributions over prizes. Because control problems entail searching over alternative γ's, it is more convenient to formulate them in terms of a baseline model $\tau_o(w)dv(w)$, as we originally did in subsection 4.2.*

Remark 4.7. *If we use relative entropy as a statistical divergence, then*

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

and dual problem (23) *becomes*[35]

$$\max_{\eta} -\xi \int_W \exp\left[-\frac{u[\gamma(w)]+\eta}{\xi} - 1\right] \tau_o(w) d\upsilon(w) - \eta = -\xi \log\left(\int_W \exp\left[-\frac{u[\gamma(w)]}{\xi}\right] \tau_o(w) d\upsilon(w)\right). \quad (24)$$

The minimizing m *in problem* (21) *is*

$$m^*(w) = \frac{\exp\left[-\frac{u[\gamma(w)]}{\xi}\right]}{\int_W \exp\left[-\frac{u[\gamma(w)]}{\xi}\right] \tau_o(w) d\upsilon(w)}. \quad (25)$$

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."*

Remark 4.8. *(Risk-sensitive preferences) The right side of equation* (24)*, namely,*

$$-\xi \log\left[\int_W \exp\left(-\frac{u[\gamma(w)]}{\xi}\right) \tau_o(w) d\upsilon(w)\right], \quad (26)$$

defines what are known as "risk-sensitive" preferences over control laws γ*. 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 τ_o *over repercussions. Risk-sensitive preferences are widely used in robust control theory (for example, see Jacobson (1973), Whittle (1990, 1996), and Petersen et al. (2000)).*

Remark 4.9. *Although our notation suppressed it, the* m*'s in the minimization problem can depend on the decision* δ*, as dependence that carries over to implied densities,* ℓ*.*

We could say that (18) gives a parameterization of alternative models expressed in terms of m or ℓ. But since the divergence (21) is not expressed as a divergence in terms of priors over the parameter space, we then could not view preferences (21) as a special case of the robust Bayesian decision theory described in section 4.1.1. With potential misspecifications present, we have deliberately avoided imposing a baseline prior over $\mathcal{L}$.[36] Instead each m induces an alternative Anscombe and Aumann roulette wheel. For reasons that will become clear in the next subsection, we think of m's as ways to introduce ambiguities about lotteries, disarming the "roulette wheel" analogy.

Remark 4.10. *Preferences that use a relative entropy divergence to capture concerns about model misspecification are often referred to as "multiplier preferences." Because of the different ways that we apply the*

[35] 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.

[36] Divergence preferences typically are expressed in terms of probabilities distributions over the collection states, which in the present case would be over the set $\mathcal{L}$'s.

language of decision theory, the preceding construction of multiplier preferences differs from constructions provided by Maccheroni et al. (2006a) and Strzalecki (2011). Specifically, Maccheroni et al. (2006a) define the domain of their cost function to be probabilities over the state space. In our analysis, the state space is Θ*, which means that their application of variational preferences gives rise to the robust Bayesian approach in section 4.1.1.*

4.2.2 A misspecified likelihood function

We now propose a generalization of the previous approach by starting from a parameterized family of probabilities $\tau(w \mid \theta)$ and prior probability measure π. Typically, a family of parameterized family of probability models is specified so that each model is absolutely continuous with respect to an underlying measure, a condition required to apply likelihood-based methods. Consider relative densities $\hat{m}$ that for each θ have been rescaled so that

$$\int_W \hat{m}(w \mid \theta)\tau(w|\theta)d\upsilon(w) = 1. \tag{27}$$

To acknowledge misspecification of a model implied by parameter θ, let $\hat{m}(w \mid \theta)$ represent an "unstructured" relative perturbation with a parameterized family of densities:

$$\hat{\ell}(w \mid \theta) = \hat{m}(w \mid \theta)\tau(w \mid \theta)$$

where $\hat{\ell}(\cdot \mid \theta) \in \mathcal{L}$ for each $\theta \in \Theta$. 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 $\theta \in \Theta$. The pair $(\hat{m}, \theta)$ implies a probability distribution represented as

$$\hat{m}(w \mid \theta)\tau(w \mid \theta)d\upsilon(w)$$

over W conditioned on θ. When $\hat{m}$ is not identically one, we view this as a misspecified likelihood function. Uncertainty about the nature of this misspecification induces corresponding uncertainty in the induced distribution, or the lottery in the language of decision theory.

Preferences that acknowledge this form of model misspecification are ordered by solutions to

$$\min_{\hat{m}\in\widehat{\mathcal{M}}} \left(\int_W u[\gamma(w)]\hat{m}(w \mid \theta)\tau(w \mid \theta)d\upsilon(w) + \xi \int_W \phi[\hat{m}(w \mid \theta)]\tau(w \mid \theta)d\upsilon(w) \right) d\pi_o(\theta), \tag{28}$$

where the decision maker commits to the baseline prior distribution π_o.

Remark 4.11. *Please remember that we have left dependencies of* τ *and* γ *on* δ *implicit. Consequently, constraint* (27) *holds for each* $\delta \in \Delta$*, where* τ *depends implicitly on* δ*.*

Remark 4.12. *Another approach would be to use the baseline prior to construct:*

$$\int_\Theta \tau(w \mid \theta)d\pi_o(\theta)$$

and treat this as the baseline model of section 4.2.1; this would correspond to a predictive distribution provided that learning is not formally incorporated into the analysis or that the "prior" $d\pi_o$ *has already conditioned on what has been learned from available data.*[37]

[37] See Chamberlain (2020) for a discussion of robustness relative to a predictive distribution.

4.3 Robustness reconsidered

It is useful to compare two approaches to robustness that we have taken. The subsection 4.1.1 decision maker starts with a baseline prior over parameter vectors and considers consequences of misspecifying that prior. This decision maker takes as given the parameterized family of densities $\tau(w \mid \theta)$ for $\theta \in \Theta$. In contrast, the section 4.2 decision maker searches over the entire space $\widehat{\mathcal{M}}$, subject to a penalty on a statistical divergence from a baseline parameterized family of models. This decision maker considers only the baseline prior distribution.

Our setup allows the parameter space to be infinite dimensional. Consider a prior π_o that is consistent with a Bayesian approach to "nonparametric" estimation and inference. Since $\tau(\cdot \mid \theta)$ can be viewed as a mapping from Θ into $\mathcal{L}$, a prior distribution π_o over Θ implies a corresponding distribution over $\mathcal{L}$. This procedure necessarily assigns prior probability zero to a substantial portion of the space $\mathcal{L}$. Specifying a prior over the infinite dimensional space $\mathcal{L}$ 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.[38] 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 likelihood misspecifications. Indeed, the section 4.2 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 4.1 and 4.2 formulations use statistical discrepancies lead to substantial differences in the associated variational preferences, namely, representation (12) or (17) for the section 4.1 way of prior ambiguity and representation (28) way of ambiguity about the parameterized family of densites, $\tau(\cdot \mid \theta)$.

4.4 Two examples

It is instructive to apply the distinct approaches of sections 4.1 and 4.2 to simple examples. The first example gives a simple illustration of preference inputs into robust control problems, and the second one explores a forecasting problem.

4.4.1 Robust preferences

Assume the following constituents:

- Baseline model is $\tau_o(w) \sim Normal(\mu_o, \sigma_o^2)$
- Alternative structured models $\tau(w \mid \theta_i) \sim Normal(\mu_i, \sigma_i^2), i = 1, \ldots, k$, where potential parameter values (states) are $\theta_i = (\mu_i, \sigma_i)$ and parameter space $\Theta = \{\theta_i : i = 1, 2, \ldots, k\}$.. The baseline model can be one of these k models.
- Baseline prior over structured models is a uniform distribution $\pi_o(\theta_i) = \frac{1}{k}, i = 1, \ldots, k$.
- Prize is the induced distribution of $c(w) = \gamma(w)$.
- Utility function is $u[c(w)] = \log[c(w)]$, where $c(w)$ is consumption

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

- Prize rule is $\gamma(w) = \exp(\gamma_0 + \gamma_1 w)$

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

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

The expected utility conditioned on parameter vector θ_i is

$$\int_W u\left[\exp(\gamma_0 + \gamma_1 w)\right] \tau(w|\theta) d\upsilon(w) = \gamma_0 + \gamma_1 \mu_i,$$

and a statistical divergence applied to alternative priors is

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

A section 4.1.1 decision maker with variational preferences orders prize rules $\gamma(w) = \exp(\gamma_0 + \gamma_1 w)$ according to

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

For a relative entropy divergence, prize rules are ordered by

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

and the associated minimizing π_i is

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

A section 4.1.2 decision maker, in effect, chooses the multiplier ξ to hit a relative entropy constraint on the prior.

A criterion that expresses robustness to prior misspecification with a relative entropy divergence ranks prizes as either

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

or,

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

When we use relative entropy as a statistical divergence, variational preferences for a section 4.2 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 ξ call for smaller adjustments $-\frac{1}{2\xi}(\sigma_o \gamma_1)^2$ of expected utility $\gamma_0 + \gamma_1 \mu_o$ for concerns about misspecification of $\tau_o(w) d\upsilon(w)$.

4.4.2 Robust forecasting

Consistent with Example 2.5, partition

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix},$$

where w_1 is scalar outcome of a variable to be forecast and w_2 constitutes data underlying a forecast. Assume that:

- The baseline model is $\tau_o(w)$.
- Alternative structured models are $\tau(w \mid \theta)$ for a parameter space $\Theta = \{\theta_i : i = 1, 2, \ldots, k\}$. The baseline model can be one of the k models.
- The baseline prior over structured models is a uniform distribution $\pi_o(\theta_i) = \frac{1}{k}, i = 1, \ldots, k$.
- The prize is the induced distribution of the forecast error $w_1 - \delta(w_2)$, where δ is the forecast rule.
- The utility function is $-[w_1 - \delta(w_2)]^2$.
- The prize rule is $\gamma_\delta(w) = w_1 - \delta(w_2)$.

To study a subsection 4.1.1 decision maker, we proceed as follows. As in the subsection 4.4.1 example, to obtain an alternative prior π_i for $i = 1, \ldots, k$, we set $n_i = k\pi_i$ so that n_i times the baseline prior is

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

and a statistical divergence is

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

Given a forecast rule δ, for each model form the second moment of the forecast error:

$$\sigma_\delta^2(i) = \int_W [w_1 - \delta(w_2)]^2 d\tau(w \mid \theta_i),$$

and then solve

$$\min_{\pi_i \geqslant 0, \sum_{i=1}^k \pi_i = 1} -\sum_{i=1}^{k} \pi_i \sigma_\delta^2(i) + \frac{\xi}{k} \sum_{i=1}^{k} \phi\left(k\pi_i\right).$$

For a relative entropy divergence, prize rules are ordered by

$$-\xi \log \sum_{i=1}^{k} \left(\frac{1}{k}\right) \exp\left[\frac{1}{\xi}\sigma_\delta(i)\right],$$

and the associated minimizing π_i is

$$\frac{\exp\left[\frac{1}{\xi}\sigma_\delta(i)\right]}{\sum_{i=1}^{k} \exp\left[\frac{1}{\xi}\sigma_\delta(i)\right]},$$

which places more emphasis on larger second moments of forecast errors than would the equal weighting that is implied by setting ξ to be arbitrarily large. A section 4.1.2 decision maker can be interpreted as choosing the multiplier ξ to hit a relative entropy constraint on the prior.

When we use relative entropy as statistical divergence, variational preferences for a section 4.2 decision maker with model misspecification concerns are ordered by

$$-\xi \log \int_W \exp\left(\frac{1}{\xi}[w_1 - \delta(w_2)]^2\right) d\tau_o(w). \tag{29}$$

Criterion (29) reshapes a baseline density to place more weight on larger squared forecast-errors. When the baseline probability measure $d\tau_o$ is a normal distribution and δ is the mean conditioned on the forecasting variables under the reshaped distribution, the change in probability measure preserves normality and leaves the conditional mean unaltered; however, it increases the conditional variance.[39]

5 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 $\widehat{\mathcal{M}}$ 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.

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

$$\int_W \hat{m}(w \mid \theta)\tau(w|\theta)d\upsilon(w) = 1.$$

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 $\hat{m}(w \mid \theta)\ell(w \mid \theta)\tau(w|\theta)d\upsilon(w)$ over W conditioned on θ.

To measure a statistical discrepancy that comes from 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:

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

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

[39] An extension of this observation shows that in this special case, the conditional mean under the baseline normal distribution is robust to misspecification concerns represented with relative entropy.

5.1 First hybrid model

We can rank alternative prize rules γ by including the following adjustment for possible misspecification of the baseline prior π_o:

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

Here ϕ_1 and ϕ_2 are possibly distinct convex functions with properties like the ones that we imposed on ϕ in section 3.

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

$$\widehat{D}(\hat{m}, n \mid \tau, \pi_o) = \xi_1 \int_\Theta \left(\int_W \phi_1 \left[\hat{m}(w \mid \theta)\right] \tau(w \mid \theta) d\upsilon(\theta) \right) n(\theta) d\pi_o(\theta) + \xi_2 \int_\Theta \phi_2 \left[n(\theta)\right] d\pi_o(\theta) \tag{30}$$

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

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

In Appendix A we establish that divergence (30) is convex over the family of probability measures that concerns the decision maker.

Remark 5.1. *As noted earlier, Cerreia-Vioglio et al. (2013) posit a state space that includes parameters but also can include what we call repercussions. Thus, think of the state as the pair (w, θ). In this setting, one could apply a statistical divergence to a joint distribution over possible realizations of (w, θ). Since the joint distribution can be factored into the product of a distribution over W conditioned on θ and a marginal distribution over Θ, such an approach can capture robustness in the specification of both τ and π_o, albeit in a very specific way. For instance, for the relative entropy divergence, this results in the joint divergence measure:*

$$\widehat{D}(\hat{m}, n \mid \tau, \pi_o) = \xi_1 \int_\Theta \left[\int_W \hat{m}(w \mid \theta) \log \hat{m}(w \mid \theta) \tau(w \mid \theta) d\upsilon(w) \right] n(\theta) d\pi_o(\theta) + \xi_2 \int_\Theta n(\theta) \log n(\theta) d\pi_o(\theta)$$

for $\xi_1 = \xi_2$.

In earlier work, we have demonstrated important limits to such an approach in dynamic settings.[40] *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 ϕ divergence.*

5.2 Second hybrid model

As an alternative to the section 5.1 approach, we could instead constrain the set of priors to satisfy:

$$\int_\Theta \phi_2[n(\theta)] d\pi_o(\theta) \leqslant \kappa \tag{31}$$

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

$$\min_{n \in \mathcal{N}} \int_\Theta \mathbb{T}_1 \left[\gamma\right](\theta) n(\theta) d\pi_o(\theta), \tag{32}$$

[40] See Hansen and Sargent (2007, 2011) and Hansen and Miao (2018).

where minimization is subject to (31).

As in Cerreia-Vioglio et al. (2022), preferences ordered by (32) subject to constraint (31) 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 Θ.[41] Notice how the first term in discrepancy measure (30) uses a prior $nd\pi_o$ to construct a weighted averaged over $\theta \in \Theta$ of the following conditioned-on-θ misspecification measure

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

The objective in problem (32) 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 θ subject to the constraint that $\pi \in \Pi$.[42] Equivalently, in place of (30), this approach uses cost function

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

Remark 5.2. *It is possible to simplify computations by using dual versions of the hybrid approaches delineated in subsections 5.1 and 5.2. Such formulations closely parallel those described in our discussions of robust prior analysis and potential model misspecification in remarks 4.5, 4.6, and 4.7.*

6 Dynamic extension

Although a complete treatment of dynamics deserves its own paper, here we describe briefly how to extend the familiar recursive utility specification of Kreps and Porteus (1978) and Epstein and Zin (1989) to accommodate our two robustness concerns to an intertemporal environment. We accomplish this by using conditional counterparts to the preceding analysis to explore consequences of mis-specifying Markov transition dynamics and prior distributions over unknown parameters. The resulting preferences have a recursive structure. There is an inherent tension between dynamic consistency and statistical consistency in these preferences that we discuss elsewhere (Hansen and Sargent (2022)).

6.1 A deterministic warm up

We represent preferences using recursions that apply to continuation values. Abstracting from uncertainty, a commonly used intertemporal preference specification is captured by the value recursion:

$$V_t = \left[(1-\beta) (C_t)^{1-\rho} + \beta (V_{t+1})^{1-\rho} \right]^{\frac{1}{1-\rho}}$$

for $0 < \beta < 1$ and $\rho > 0$. V_t is the date t continuation value and C_t is date t consumption. The parameter β governs discounting and the parameter ρ is the reciprocal of the intertemporal elasticity of substitution. Applying the recursion over an infinite horizon leads to the following expression for the continuation value:

$$V_t = \left[(1-\beta) \sum_{j=0}^{\infty} \beta^j (C_{t+j})^{1-\rho} \right]^{\frac{1}{1-\rho}}$$

[41] 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.

[42] 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.

Since the logarithmic transformation is increasing, we can use the following recursion in the logarithm $\widehat{V}_t$ of the continuation value to represent preferences:

$$\widehat{V}_t = \frac{1}{1-\rho}\log\left[(1-\beta)\exp\left[(1-\rho)\widehat{C}_t\right] + \beta\exp\left[(1-\rho)\widehat{V}_{t+1}\right]\right]$$

where $\widehat{C}_t$ is the logarithm of consumption.

6.2 Introducing uncertainty

Let $\mathfrak{A}_t$ denote a sigma algebra capturing information available to the decision maker at date t. Think of the repercussion W_{t+1} as generating new information relative to $\mathfrak{A}_t$ that is pertinent for constructing $\mathfrak{A}_{t+1}$. Think of the continuation value, $\widehat{V}_{t+1}$ as the counterpart to a prize that can depend on a repercussion vector W_{t+1}. A continuation value $\widehat{V}_{t+1}$ is constrained to be measurable with respect to $\mathfrak{A}_{t+1}$. We explore model misspecification by using nonnegative random variables M_{t+1} that are $\mathfrak{A}_{t+1}$ measurable and satisfy $\mathbb{E}\left(M_{t+1} \mid \mathfrak{A}_t, \theta\right) = 1$. We explore prior/posterior misspecification using nonnegative random variables N_t that are measurable with respect $\mathfrak{A}_t$ augmented by knowledge of θ and satisfy $\mathbb{E}\left(N_t \mid \mathfrak{A}_t\right) = 1$.

To accommodate robustness concerns in decision making, define preferences with three recursions for updating the continuation value

$$\begin{aligned}
\widehat{V}_t &= \frac{1}{1-\rho}\log\left[(1-\beta)\exp\left[(1-\rho)\widehat{C}_t\right] + \beta\exp\left[(1-\rho)\overline{R}_t\right]\right] \\
\widehat{R}_t &= \min_{M_{t+1}\geqslant 0, \mathbb{E}(M_{t+1}|\mathfrak{A}_t,\theta)=1} \mathbb{E}\left[M_{t+1}\widehat{V}_{t+1} + \xi_1\phi_m\left(M_{t+1}\right) \mid \mathfrak{A}_t, \theta\right] \\
\overline{R}_t &= \min_{N_t\geqslant 0, \mathbb{E}(N_t|\mathfrak{A}_t)=1} \mathbb{E}\left[N_t\widehat{R}_t + \xi_2\phi_n\left(N_t\right) \mid \mathfrak{A}_t\right] \qquad (33)
\end{aligned}$$

where $\widehat{R}_t$ adjusts next-period's continuation value for potential model misspecification captured by conditioning the unknown parameter θ, and $\overline{R}_t$ adjusts for "prior robustness." Date t "priors" actually condition on $\mathfrak{A}_t$. The three recursions affect values in different ways:

- the first adjusts for discounting and intertemporal substitution;
- the second adjusts for model misspecification;
- the third adjusts for prior misspecification.

The second and third recursions provide a dynamic counterpart to the approach in section 5.1. Replacing the third recursion in (33) with a constrained counterpart gives a dynamic counterpart to the approach in section 5.2.[43]

6.3 Shadow valuation

Following Hansen and Richard (1987) and others, we can use stochastic discount factors to value an assets having an uncertain one-period ahead payoff. We deduce shadow values by computing a one-period intertemporal marginal rate of substitution. Of particular interest to us are contributions that our model-misspecification operator $\widehat{R}_t$ and our prior-robustness operator $\overline{R}_t$ make to this shadow value.

[43] See Hansen and Sargent (2021, 2022) for elaboration and application of this alternative approach.

A contribution to the shadow value that comes from the first recursion in (33) looks at marginal contributions in adjacent time periods. Date t marginal contributions of C_t and $\overline{R}_t$ to the current period continuation value are:

$$\begin{aligned} MC_t &= (1-\beta)\exp\left[(\rho-1)\widehat{V}_t\right](C_t)^{-\rho} \\ M\overline{R}_t &= \beta\exp\left[(\rho-1)\widehat{V}_t\right]\exp\left[(1-\rho)\overline{R}_t\right]. \end{aligned}$$

Since our aim is to infer the one-period intertemporal marginal rate of substitution, we look across adjacent time periods using consumption at each date as a numeraire:

$$\frac{MC_{t+1}M\overline{R}_t}{MC_t} = \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\exp\left[(\rho-1)\left(\widehat{V}_{t+1}-\overline{R}_t\right)\right].$$

This would give the deterministic intertemporal marginal rate of substitution if we were to substitute $\widehat{V}_{t+1}$ for $\overline{R}_t$ in this expression.

For the uncertainty adjustments, we deduce the marginal contributions by applying the Envelope Theorem to the minimization problems in the second and third recursions in (33):

- $M\widehat{V}_{t+1} = M^*_{t+1}$
- $M\widehat{R}_t = N^*_t$

Thus, the minimizing changes in probabilities contribute directly to the shadow valuation. The resulting increment to a stochastic discount factor process is:

$$\frac{S_{t+1}}{S_t} = \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\exp\left[(\rho-1)\left(\widehat{V}_{t+1}-\overline{R}_t\right)\right]M^*_{t+1}N^*_t$$

where

- M^*_{t+1} adjusts for possible model misspecification
- N^*_t adjusts for possible prior misspecification

7 An approach to uncertainty quantification

Subsection 5 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 (ξ_1, ξ_2) express 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 (ξ_1, ξ_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 (ξ_1, ξ_2) provides evidence about the forms of subjective uncertainty and potential model misspecification that *should* be of most concern. That can provide 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 (ξ_1, ξ_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 (ξ_1, ξ_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 Relation to statistical learning

We briefly compare our approach to related analyses coming from statistical learning theory and, in particular, PAC (probably approximately correct) Bayesian analysis. See Guedj (2019) for a recent survey of PAC Bayesian methods and see McAllester (1999) and Cantani (2007), among others, for fundamental contributions. While their formulations of a decision problem differ from ours, there are intriguing connections.

To understand some of the connections, partition a random vector, Y, with realization y as

$$Y' := \begin{bmatrix} Y_1' & Y_2' & Y_2' & ... & Y_K' \end{bmatrix},$$

and regard it as "training data" for a machine learning method. For an objective function, construct an "empirical risk" criterion:

$$\widehat{\Phi}(Y, \theta) := \frac{1}{K} \sum_{k=1}^{K} \Phi(Y_k, \theta).$$

The object of interest θ can be an element of a collection of functions. Define the population counterpart to $\widehat{\Phi}(Y, \theta)$ as $\overline{\Phi}(\theta)$, the Law of Large Numbers limit of $\widehat{\Phi}(W, \theta)$ as $K \to \infty$. Suppose that an idealized target decision solves:

$$\theta^* = \arg\min_{\theta \in \Theta} \overline{\Phi}(\theta).$$

This estimator appears in an extensive literature on M estimation; the idealized optimized decision θ^* defines a parameter or decision of interest. In this setting, we use decisions and parameters interchangeably, in contrast to our formulation.

A common approach to M estimation is to solve the finite sample analog problem

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \widehat{\Phi}(Y, \theta).$$

This approach struggles when the space is Θ is "large", the typical case with machine learning methods that fit flexible functional forms with many parameters. More generally, statistical learning often seeks meaningful worst-case bounds of finite sample approximates to a solution of the population problem.

Motivated by concerns for applications when the space Θ in standard M estimation is expansive, the PAC Bayesian approach proceeds differently. The approach seeks a probability distribution π over the space Θ given the data W, rather than a single value, θ. By analogy to Bayesian methods, such a distribution is referred to as a 'generalized posterior.' The approach imposes a baseline prior distribution π_o over the space Θ, and considers generalized posteriors in a family:

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

for $\int n d\pi_o = 1$.

Instead of solving the finite sample M estimation problem, consider a family of problems:

$$\min_{n, \int n d\pi_o = 1} \int_\Theta \widehat{\Phi}(Y, \theta) d\pi(\theta) + \xi \int_\Theta \log n(\theta) n(\theta) d\pi_o(\theta) \tag{34}$$

indexed by ξ. Applying the same mathematics we have used in previous sections, minimization brings exponential tilting:

$$n^*(\theta) = \frac{\exp\left[-\frac{1}{\xi}\widehat{\Phi}(Y, \theta)\right]}{\int_\Theta \exp\left[-\frac{1}{\xi}\widehat{\Phi}(Y, \tilde{\theta})\right] d\pi_o(\tilde{\theta})}$$

The Bayesian PAC uses this minimizer to construct an approximation to a minimizer of the underlying (infeasible) population problem.

Problem (34) provides a way to incorporate probabilistic restrictions into the M estimation problem. There are some interesting special cases. When $\widehat{\Phi}(W, \cdot)$ is the negative of the log likelihood function and $\xi = 1$, we are led to a standard calculation of a Bayesian posterior. Zhang (2006), Grünwald (2011) and others propose and defend a 'safe Bayesian framework' by exploring other values of $\xi > 1$ based on robustness considerations. The Bayesian PAC approach studies alternative specifications of $\widehat{\Phi}(W, \cdot)$ based on more general loss functions. As in M-estimation more generally, this construction may embed some robustness concerns. When ξ tends to infinity the generalized posterior collapses to the prior distribution. When ξ tends to zero the prior becomes inconsequential, and the generalized posterior collapses to the solution the finite-sample M estimation solution. The penalty parameter ξ governs a tradeoff between the importance of the objective $\widehat{\Phi}(Y, \cdot)$ and the baseline prior π_o. The PAC-Bayesian literature discusses extensively the role of the parameter in approximation.

While our approach shares much mathematical structure with PAC-Bayesian methods, it differs in ways that are significant for applications. The M estimation formulation ties its decision problem directly to an underlying unknown "parameter" and contains no counterpart to the maximization steps that we use to represent uncertainty aversions. Furthermore, the PAC-Bayesian Problem (34) conditions on W and focuses exclusively on uncertainties about unknown states or parameters. Also, PAC-Bayesian methods use Problem (34) as a device to approximate a solution to an infeasible population problem, which is not a component of our analysis.

To elaborate more on the differences between PAC-Bayesian methods and our approach, we study decision problems in which parameters are not the objects of ultimate interest but instead are just intermediate "means to ends" of constructing decision rules for making choices of economics quantities that are robust to misspecifications. Rather than replacing a log-likelihood function with an M estimation objective and possibly down-weighting its importance, we introduce potential likelihood misspecifications explicitly; we also formally acknowledge possible misspecification of priors. By appropriately adjusting the divergence "cost" structure, our approach allows us to explore tradeoffs between concerns about misspecifications of likelihoods, on the one hand, and priors, on the other hand.

9 Concluding remarks

Except for our brief section 6 excursion, we have confined ourselves to a "static" setting that allowed us to apply and extend a framework created by Maccheroni et al. (2006a) to distinguish ambiguity about a prior and from concerns for misspecifications of likelihood functions. In doing this we reinterpret objects that appear

in the Anscombe and Aumann (1963) formulation to represent both types of doubts, distinct specification concerns that are familiar to applied statisticians. We intend the present paper as a prolegomenon to a sequel in which we shall extend and reinterpret the dynamic variational preferences of Maccheroni et al. (2006b). That dynamic formulation will connect 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 the issues studied here will arise that framework, additional ones such as dynamic consistency and appropriate choices of state variables for recursive formulations of preferences also appear.[44]

[44]To apply quantum methods dynamic asset pricing models, Ghysels and Mogan (2023) deploy extensions of the formulation developed here. An interesting notion of a "state" again comes into play.

A Convexity of composite divergence

To verify convexity of (30), consider two joint probability measures on $W \times \Theta$:

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

A convex combination of these two probability measures 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

$$\phi_2[n_\alpha(\theta)] \leqslant (1-\alpha)\phi_2[n_0(\theta)] + \alpha\phi_2[n_1(\theta)]. \tag{35}$$

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,

$$\phi_1[\hat{m}_\alpha(w \mid \theta)]n_\alpha(\theta) \leqslant (1-\alpha)n_0(\theta)\phi_1[\hat{m}_0(w \mid \theta)] + \alpha n_1(\theta)\phi_1[\hat{m}_1(w \mid \theta)]. \tag{36}$$

Multiply (36) by ξ_1 and (35) by ξ_2, add the resulting two terms, and integrate with respect to $\tau(w \mid \theta)dv(w)d\pi_o(\theta)$ to verify that divergence (30) 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.

Bertsekas, Dimitri P. 1976. *Dynamic programming and stochastic control.* New York: Academic Press.

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.

Cantani, O. 2007. PAC-Baysian Supervised Classification: The Thermodynamics of Statistical Learning. In *IMS Lecture Notes Monograph Series*, vol. 56, 1–163.

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 Macchernoni, 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 Martin Schneider. 2003. Recursive multiple-priors. *Journal of Economic Theory* 113 (1):1–31.

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 prévision: 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.

Ghysels, Eric and Jack Mogan. 2023. On Potential Exponential Computatoinal Speed-Ups in Solving Dynamic Asset Pricing Models. University of North Carolina.

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).

Grünwald, Peter. 2011. Safe learning: Bridging the gap between Bayes, MDL and statistical learning theory via empirical convexity. In *Proceedings of the 24th Annual Proceedings of the 24th Annual Conference on Learning Theory*, vol. 19, 397–420. Cambridge, MA: MIT Press.

Guedj, Benjamin. 2019. A Primer on PAC-Bayesian Learning. In *Proceedings of the 2nd congress of the Société Mathématique de France*, 391–414. Paris: French Mathematical Society.

Hallegatte, Stéphane, Ankur Shah, Casey Brown, Robert Lempert, and Stuart Gill. 2012. Investment decision making under deep uncertainty–application to climate change. Tech. Rep. 6193, World Bank Policy Research Working Paper.

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 Scott F. Richard. 1987. The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica* 50:587–614.

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.

———. 2011. Robustness and Ambiguity in Continuous Time. *Journal of Economic Theory* 146:1195–1223.

———. 2021. Macroeconomic Uncertainty Prices when Beliefs are Tenuous. *Journal of Econometrics* 223 (1):222–250.

———. 2022. Structured Ambiguity and Model Misspecification. *Journal of Economic Theory* 199:105165.

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

Ho, Paul. 2023. Global robust Bayesian analysis in large models. *Journal of Econometrics* 235:608–642.

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.

Maier, Holger R, Joseph HA Guillaume, Hedwig van Delden, Graeme A Riddell, Marjolijn Haasnoot, and Jan H Kwakkel. 2016. An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? *Environmental modelling & software* 81:154–164.

Marchau, Vincent AWJ, Warren E Walker, Pieter JTM Bloemen, and Steven W Popper. 2019. *Decision making under deep uncertainty: from theory to practice*. Springer Nature.

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

McAllester, David A. 1999. PAC-Bayesian model averaging. In *Proceedings of the 12th Annual ACM Conference on Computational Learning Theory*, 164–170.

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.

Rising, James, Marco Tedesco, Franziska Piontek, and David A Stainforth. 2022. The missing risks of climate change. *Nature* 610 (7933):643–651.

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.

Zhang, Tong. 2006. From epsilon-entropy to KL-entropy: Analysis of minimum information complexity density estimation. *Annals of Statistics* 34:2180–2210.