

Risk, Ambiguity, and Misspecification

Contents

- 10.1. Introduction
- 10.2. Background motivation
- 10.3. Decision theory overview
- 10.4. Subjective expected utility
- 10.5. An extreme response
- 10.6. Divergences
- 10.7. Robust Bayesian preferences and ambiguity aversion
- 10.8. Using uncertainty aversion to represent concerns about model misspecification
- 10.9. Robust Bayes with model misspecification
- 10.10. A dynamic robust decision problem under commitment
- 10.11. A recursive counterpart to a dynamic robust decision problem
- 10.12. Implications for uncertainty quantification

[!\[\]\(666e09182d4cd268646ea700ea60dcdf_img.jpg\) Download PDF here](#)

Authors: Lars Peter Hansen (University of Chicago) and Thomas Sargent (NYU)



Pioneers in Uncertainty and Decision Theory. Frank Knight, Abraham Wald, and Jimmie Savage.

“Uncertainty must be taken in a sense radically distinct from the familiar notion of risk, from which it has never been properly separated.... and there are far-reaching and crucial differences in the bearings of the phenomena depending on which of the two is really present and operating.” [[Knight, 1921](#)]

“The only uncertainty we can rely on is blind chance.” Attributed to Ludvig Beethoven during Napoleonic Wars.

10.1. Introduction

We refer to a likelihood function as a parameterized probability distribution, while a prior probability distribution describes a decision maker’s subjective beliefs about those parameters.^[2] This chapter brings recent contributions to decision theory into contact with statistics and econometrics by distinguishing roles played by likelihood functions and prior probability distributions. We do this to address some practical

econometric concerns about misspecifications of the parameterized probability distributions themselves, as well as doubts about appropriate prior probabilities.

To address these concerns, we combine ideas from control theories — which construct decision rules that are robust to a class of model misspecifications — with axiomatic decision theories created by economic theorists. Such decision theories originated with axiomatic formulations by von Neumann and Morgenstern, Savage, and Wald ([\[Wald, 1947\]](#), [\[Wald, 1949\]](#), [\[Wald, 1950\]](#), [\[Savage, 1954\]](#)). Ellsberg ([\[Ellsberg, 1961\]](#)) pointed out that Savage’s framework seems to include nothing that could be called “uncertainty” as distinct from “risk”. Theorists after Ellsberg constructed coherent systems of axioms that embrace a notion of ambiguity aversion. Remarkably from our point of view, most such axiomatic formulations of decision making under uncertainty in economics are not cast explicitly in terms of likelihood functions and prior distributions over parameters.

We want to reinterpret objects that appear in some of those axiomatic foundations of decision theories in ways useful to an econometrician. We use an axiomatic structure to express ambiguity about a prior over a family of statistical models, on the one hand, along with concerns about misspecifications of those models, on the other hand.

More specifically, 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-Aumann states as *parameters* that index alternative statistical models of random variables affecting outcomes that a decision maker cares about. By modifying aspects of [\[Gilboa et al., 2010\]](#), [\[Cerreia-Vioglio et al., 2013\]](#), and [\[Denti and Pomatto, 2022\]](#), we show how to use variational preferences to represent uncertainty about priors and also concerns about statistical model misspecifications.

Although they proceeded differently than we do here, [\[Chamberlain, 2020\]](#), [\[Cerreia-Vioglio et al., 2013\]](#), and [\[Denti and Pomatto, 2022\]](#) studied related issues. [\[Chamberlain, 2020\]](#) emphasized that likelihoods and priors are both vulnerable to potential misspecifications. He focused on uncertainty about predictive distributions constructed by integrating likelihoods with respect to priors. In contrast to Chamberlain, we formulate a decision theory that distinguishes uncertainties about priors from uncertainties about likelihoods. [\[Cerreia-Vioglio et al., 2013\]](#) (section 4.2) provided a rationalization of the smooth ambiguity preferences proposed by [\[Klibanoff et al., 2005\]](#) that includes likelihoods and priors as components. [\[Denti and Pomatto, 2022\]](#) extended this approach by using an axiomatic revealed preference method to deduce a parameterization of a likelihood function. However, neither [\[Cerreia-Vioglio et al., 2013\]](#) nor [\[Denti and Pomatto, 2022\]](#) sharply distinguished prior uncertainty from concerns about misspecifications of likelihood functions. We want to do that. We formulate concerns about statistical model misspecifications as uncertainty about likelihoods.

Discrepancies between two probability distributions occur throughout our analysis. They help us to connect our framework to some models in “behavioral” economics and finance. The decision makers inside some of those models have expected utility preferences in which an agent’s subjective probability –

typically a predictive density – differs systematically from the predictive density that the model builder assumed governs the data.^[3] Other “behavioral” models focus on putative differences among agents’ degrees of confidence in their views of the world. Our framework implies that the form taken by a “lack of confidence” ought to depend on the probabilistic concept about which a decision maker is uncertain, in particular, either a likelihood or a prior or a predictive distribution. Preference structures that we describe in this chapter allow us to formalize different amounts of “confidence” about details in the specification of particular statistical models, on one hand, and about subjective probabilities to attach to alternative statistical models, on the other hand. Our representations of preferences provide ways to characterize degrees of confidence in terms of statistical discrepancies between alternative probability distributions.^[4]

10.2. Background motivation

We are sometimes told that we live in a “data rich” environment. Nevertheless, data are often not “rich” along all of the dimensions that we care about for decision making. Furthermore, data don’t “speak for themselves”. We have to make them speak by positing a family of statistical models indexed by parameters that the data can help us infer. For instance, see [\[Hansen, 2019\]](#) for an elaboration. Thus, despite the “data rich” hype, the type of learning we actually do is to infer parameters of a family of statistical models.

Doubts about what existing evidence has taught us about some important dimensions have led some scientists to think about what they call “deep uncertainties.” For example, in a recent paper we read:

“The economic consequences of many of the complex risks associated with climate change cannot, however, currently be quantified. ... these unquantified, poorly understood and often deeply uncertain risks can and should be included in economic evaluations and decision-making processes.” [\[Rising et al., 2022\]](#)

In this chapter, we formulate “deep uncertainties” as a lack of confidence in how we represent probabilities of events and outcomes that are pertinent for designing tax, benefit, and regulatory policies. We do this by confessing ambiguities about probabilities, though necessarily in a restrained way.

In our experience as macroeconomists, model uncertainties are not taken seriously enough, too often being dismissed as being of “second-order”, whatever that means in various contexts. In policy-making settings, there is sometimes a misplaced wisdom that acknowledging uncertainty should tilt decisions toward passivity. But in other times and places, one senses that model uncertainty emboldens pretense. As Hayek warned:

“Even if true scientists should recognize the limits of studying human behavior, as long as the **public has expectations**, there will be people who **pretend** or **believe** that they can do more to meet popular demand than what is really in their power.” [\[Hayek, 1989\]](#)

As economists, part of our job is to delineate tradeoffs. Explicit incorporation of precise notions of uncertainty allows us to explore two tradeoffs pertinent to decision making. The first tradeoff concerns how to weigh alternative models. Difficult tradeoffs emerge when we consider implications from multiple statistical models and alternative parameter configurations. Thus, when making decisions, how much weight should we assign to best “guesses” in the face of our model specification doubts, versus bad outcomes that our doubts might unleash? Focusing exclusively on best guesses can lead us naively to ignore adverse possibilities worth considering. Focusing exclusively on worrisome bad outcomes can lead to extreme policies that perform poorly in more normal outcomes. Such considerations induce us to formalize tradeoffs in terms of explicit expressions of aversions to uncertainty.

The second tradeoff is intertemporal: should we *act now*, or should we *wait* until we have learned more? While waiting is tempting, it can also be so much more costly that it becomes prudent to take at least some actions now even though we anticipate knowing more later.

10.2.1. Aims

With these tradeoffs in mind, in this chapter we allow for uncertainties that include

- **risks**: unknown outcomes with known probabilities.
- **ambiguities**: unknown weights to assign to alternative probability models.
- **misspecifications**: unknown ways in which a model provides flawed probabilities.

We will focus on formulations that are enlightening as well as tractable.

10.3. Decision theory overview

Decision theories under uncertainty provide axiomatic formulations of “rationality.” Since there are multiple such formulations of decision making under uncertainty, it is perhaps best to replace the term “rational” with “prudent.” While these axiomatic formulations are of intellectual and substantive interest, in this chapter we will focus on the implied representations. We think that this approach is interesting because we sympathise with Savage’s own perspective on his axiomatic formulation:

“Indeed the axioms have served their only function in justifying the existential parts of Theorems 1 and 3; in further exploitation of the theory, ..., the axioms themselves can end, in my experience, and should be forgotten.” [[Savage, 1952](#)]

Later in this chapter, for interested readers, we will delineate specific links to axiomatic formulations of [[Savage, 1954](#)] along with [[Anscombe and Aumann, 1963](#)]. As a bit of intellectual history, [[Anscombe and](#)

[Aumann, 1963](#)] provided a different way to justify Savage’s representation of decision making in the presence of subjective uncertainty. They prominently feature the distinction between a “horse race” and a “roulette wheel”. They rationalize preferences over acts, where an act maps states into lotteries over prizes. The latter is the counterpart to a roulette wheel. Probability assignments over states then become the subjective input and the counterpart to the “horse race.”

[Anscombe and Aumann, 1963](#)] used this formulation to extend the von Neumann-Morgenstern expected utility with known probabilities to decision problems where subjective probabilities also play a central role as in Savage’s approach. While [Anscombe and Aumann, 1963](#)] provides an alternative derivation of subjective expected utility, many subsequent contributions used the Anscombe-Aumann framework to extend the analysis to incorporate forms of ambiguity aversion. Prominent examples include [Gilboa and Schmeidler, 1989](#)] and [Maccheroni et al., 2006](#)]. In what follows, we provide a statistical slant to such analyses.

10.3.1. Approach

In this chapter we will use representations implied by modifications and extensions of Savage-style axiomatic formulations. The extensions allows to incorporate notions of uncertainty beyond risk, motivated by [Knight, 1921](#)] and the subsequent literature that we noted. The overall aim is to make contact with applied challenges in economics and other disciplines. We will start with the basics of statistical decision theory and then proceed to explore extensions that distinguish concerns about potential misspecifications of likelihoods from concerns about the misspecification of priors. This opens the door to better ways of conducting uncertainty quantification for stochastic economic models used for private sector planning and governmental policy assessment. Our task in this chapter is to provide tractable and revealing methods for exploring sensitivity to subjective uncertainties, including potential model misspecification and ambiguity across models. This will allow us systematically to:

- assess the impact of uncertainty on prudent decision or policy outcomes;
- isolate the forms of uncertainty that are most consequential for these outcomes.

To make the methods tractable and revealing, we will utilize tools from probability and statistics to limit the type and amount of uncertainty that is entertained. As inputs, the resulting representations of objectives for decision making will require a specification of aversion to or dislike of uncertainty about probabilities over future events.

10.3.2. Basic setup

Our analysis uses the following components.

1. **A parameterized family of models:** Models are expressed in terms of densities of a random vector with realization x and indexed by a parameter θ with probabilities given by: $\ell(x | \theta)d\tau(x)$ where

$$\int_{\mathcal{X}} \ell(x | \theta) d\tau(x) = 1.$$

In this depiction of probability models, $\theta \in \Theta$ and Θ is a parameter space, \mathcal{X} is the space of possible realizations of x , and $d\tau$ is a measure used conjunction with the model-specific densities. We refer to ℓ as a likelihood and the probability implied by each θ as a *structured* probability model.

2. **A partitioning of $\mathbf{x} = (\mathbf{w}, \mathbf{z})$:** The vector z , consists of observations or signals that are observed when decisions are made, and the vector w is an outcome or shock that is realized after decisions are made. For instance, in an intertemporal context, w may reflect future shocks.
3. **A prize rule:** Denoted by γ , a prize rule maps \mathcal{X} into prizes, which are the argument of a utility function, U . We allow the utility function to depend on the unknown parameter θ , as is common in statistical decision problems. Arguably, such a formulation is a shorthand for a more primitive specification in which in the parameter has ramifications for a future prize that is left unmodeled.
4. **A risk function:** The risk function is the expected utility for a prize rule as a function conditioned on the parameter vector θ :^[5]

$$\bar{U}(\gamma | \theta) = \int_{\mathcal{X}} U[\gamma(x), \theta] \ell(x | \theta) d\tau(x).$$

This construct was initiated in a foundational paper by [\[Wald, 1947\]](#) and embraced in part expected utility under known probabilities as rationalized previously by [\[Ramsey, 1931\]](#) and [\[Von Neumann and Morgenstern, 1944\]](#).

5. **A decision rule:** Denoted by δ , A decision rule is a function of the observations or signals. Decision rules are restricted to be in a set Δ . Since a decision rule can depend on z , *ex post* learning is entertained. A decision rule gives a restriction on a prize rule via a mapping:

$$\gamma_{\delta}(x) = \Psi[\delta(z), x].$$

6. **A prior distribution, π , over the parameter space Θ :** We will sometimes have reason to focus on a particular prior, π , that we will call a baseline prior. The unique specification of prior from a Bayesian perspective is a subjective input sometimes referred to as “personalized probability.”

Note that preferences over prize rules imply a ranking over decision rules, the δ 's, in Δ . While we feature the impact of a decision rule on a prize rule, we may extend the analysis to allow δ to influence ℓ as happens when we entertain experimentation.

10.3.3. Acts

We now digress a bit and provide a mapping from our formulas into languages used in axiomatic decision theory. A central construct in that literature is an **act**. We suggest two constructions of an act, we call the first one a Savage act and the second one an Anscombe-Aumann act. While [[Savage, 1954](#)] is an earlier formulation of an axiomatic defense for expected utility constructed with a subjective probabilities, the subsequent [[Anscombe and Aumann, 1963](#)] formulation has been featured in a variety of extensions of extensions of expected utility.

We connect to [[Savage, 1954](#)] by defining a **state** to be a parameter value $\theta \in \Theta$, and a **consequence** to be a probability measure over \mathcal{C} . A **Savage act** maps each θ into a probability distribution over \mathcal{C} . For our analysis the acts of interest use the probability measure that is induced by prize rule $\gamma(x)$ in conjunction with $\ell(x | \theta)d\tau(x)$ for each state θ . We assume a linear utility function over consequences as given by:

$$U(c) = \int_{\mathcal{C}} u(c)dp(c)$$

for a probability measure p over \mathcal{C} .

To provide connection to [[Anscombe and Aumann, 1963](#)], we let \mathcal{C} be the space of consequences. An **Anscombe-Aumann act** maps each $\theta \in \Theta$ into a probability measure over \mathcal{C} . Again we may use a prize rule, $\gamma(x)$ in conjunction with the parameterized family $\ell(x | \theta)d\tau(x)$ to induce a probability measure over \mathcal{C} . The utility function remains the same.

[[Anscombe and Aumann, 1963](#)] used a roulette wheel analogy for the outcome of an act, and a horse race analogy for a distribution over Θ . In our setup, we may think of Wald's risk function as reflecting a roulette wheel type contribution to expected utility.

10.3.4. A simple application

As an illustration, we consider a model selection problem. Suppose $\Theta = \{\theta_1, \theta_2\}$, and that the decision maker can use the entire underlying random vector. Thus, we make no distinction between γ and δ and no distinctions between x and z .

The decision rule, $\delta = \gamma$ is a mapping from \mathcal{Z} to $[0, 1]$ where $\delta(z) = 1$ means that the decision maker selects model θ_1 and $\delta(z) = 0$ means that the decision maker selects model θ_2 . We allow for intermediate values between zero and one, which can be interpreted as randomization. These intermediate choices will end up not being of particular interest for this example.

The utility function for assessing risk is:

$$U(\delta, \theta_1) = v_1 \delta$$

$$U(\delta, \theta_2) = v_2(1 - \delta)$$

where v_1 , and v_2 are positive utility parameters.

A class of decision rules, called threshold rules, will be of particular interest. Partition \mathcal{Z} into two sets:

$$\mathcal{Z} = \mathcal{Z}_1 \cup \mathcal{Z}_2,$$

where the intersection of \mathcal{Z}_1 and \mathcal{Z}_2 is empty. A threshold rule has the form:

$$\delta(z) = \begin{cases} 1 & \text{if } z \in \mathcal{Z}_1 \\ 0 & \text{if } z \in \mathcal{Z}_2. \end{cases}$$

For a threshold rule, the risk function (the expected utility conditioned on θ is:

$$\bar{U}(\gamma | \theta_1) = v_1 \int_{\mathcal{Z}_1} \ell(z | \theta_1) d\tau(z)$$

$$\bar{U}(\gamma | \theta_2) = v_2 \int_{\mathcal{Z}_2} \ell(z | \theta_2) d\tau(z).$$

Suppose that the utility weights v_1 and v_2 are both one. Under this threshold decision rule, $1 - \bar{U}(\gamma | \theta_1)$ is the probability of making a mistake when model θ_1 generates the data, and $1 - \bar{U}(\gamma | \theta_2)$ is the probability of making a mistake if model θ_2 generates the data. In statistics, the first of these is called a Type I error, and the second a Type II error, assuming we consider θ_1 to be the “null model” and θ_2 to be the “alternative model.” The utility weights determine the relative importance, to the decision maker, of making a correct identification of the model.

10.4. Subjective expected utility

Under subjective expected utility, we order preferences over γ , and hence δ , by

$$\int_{\Theta} \left[\int_{\mathcal{X}} U[\gamma(x), \theta] \ell(x|\theta) d\tau(x) \right] d\pi(\theta) = \int_{\Theta} \bar{U}(\gamma | \theta) d\pi(\theta)$$

for a specific π . This representation is supported by Savage and Anscombe-Aumann axioms, but imposes full confidence with **no potential misspecification** of the priors or the likelihood.

We use these preferences for a decision problem where prize rules are restricted to be in the set $\Gamma(\Delta)$:

Problem 11.1

$$\max_{\gamma \in \Gamma(\Delta)} \int_{\Theta} \left(\int_{\mathcal{X}} U[\gamma(x), \theta] \ell(x|\theta) d\tau(x) \right) d\pi(\theta). \quad (10.1)$$

Recall that partitioning of $x = (w, z)$ where the decision rule can only depend on z and the prize rule on the entire x vector. Factor $\ell(\cdot | \theta)$ and τ as:

$$\begin{aligned} d\tau(x) &= d\tau_2(w) d\tau_1(z) \\ \ell(x | \theta) &= \ell_2(w | z, \theta) \ell_1(z | \theta) \end{aligned} \quad (10.2)$$

In particular, the density ℓ_2 captures the probabilities of the shock after observing data, z . These factorizations allow us to write the objective as:

$$\int_{\Theta} \left[\int_{\mathcal{Z}} \left(\int_{\mathcal{W}} U[\gamma(x), \theta] \ell_2(w | z, \theta) d\tau_2(w) \right) \ell_1(z | \theta) d\tau_1(z) \right] d\pi(\theta).$$

Notice that even if the utility function U does not depend on θ , this dependence may emerge after we integrate over \mathcal{W} in the innermost integration because of the dependence of $\ell_2(\cdot | \theta)$ on the unknown parameter.

10.4.1. Making a decision

To solve problem [\(10.1\)](#), it is convenient to exchange the orders of integration in the objective:

$$\int_{\mathcal{Z}} \left[\int_{\Theta} \left(\int_{\mathcal{W}} U[\gamma(x), \theta] \ell_2(w | z, \theta) d\tau_2(w) \right) \ell_1(z | \theta) d\pi(\theta) \right] d\tau_1(z)$$

The γ 's of interest are those that are built from decisions rules, the δ 's. As δ only depends on z and the objective is additively separable in z , we may equivalently solve a **conditional problem** using the objective:

$$\tilde{U}[\delta(z)] \stackrel{\text{def}}{=} \int_{\Theta} \left(\int_{\mathcal{W}} U(\Psi[\delta(z), w, z], \theta) \ell_2(w | z, \theta) d\tau_2(w) \right) \ell_1(z | \theta) d\pi(\theta)$$

for each value of z provided that the restrictions imposed on δ by the construction of the set of decision rules Δ are separable in z . That is, provided that we can write:

$$\Delta = \{\delta : \delta(z) \in \Delta(z) \text{ for all } z \in \mathcal{Z}\} \quad (10.3)$$

for given constraint sets $\Delta(z)$, we may solve

Problem 11.2

$$\max_{\delta(z) \in \Delta(z)} \tilde{U}[\delta(z)] \quad (10.4)$$

Finally, observe that

$$d\bar{\pi}(\theta | z) \stackrel{\text{def}}{=} \left[\frac{\ell_1(z | \theta)}{\int_{\Theta} \ell_1(z | \theta) d\pi(\theta)} \right] d\pi(\theta)$$

is the Bayesian posterior distribution for θ . Equivalently, we may solve the conditional Bayesian problem:

Problem 11.3

$$\max_{\delta(z) \in \Delta(z)} \int_{\Theta} \left(\int_{\mathcal{W}} U[\Psi(\delta(z), w, z), \theta] \ell_2(w | z, \theta) d\tau_2(w) \right) d\bar{\pi}(\theta | z), \quad (10.5)$$

since in forming the objective of conditional problem, [\(10.5\)](#), we divided the objective for the conditional problem, [\(10.4\)](#) by a function of z alone.

This outcome justifies a distinct role for a statistician to produce a posterior distribution for the parameter vector based on empirical evidence and a decision-maker that takes this posterior as input and makes decisions.

10.4.2. A simple application reconsidered I

For illustration purposes, consider the example given in Section [A simple application](#). In this example, $x = z$ and there is no w contribution. Impose prior probabilities, $\pi(\theta_1)$, and $\pi(\theta_2) = 1 - \pi(\theta_1)$ on the two models. Compute the Bayesian posterior probabilities for each value of θ ,

$$\bar{\pi}(\theta_1 | z) = \frac{\ell(z | \theta_1)\pi(\theta_1)}{\ell(z | \theta_1)\pi(\theta_1) + \ell(z | \theta_2)\pi(\theta_2)}$$

$$\bar{\pi}(\theta_2 | z) = \frac{\ell(z | \theta_2)\pi(\theta_2)}{\ell(z | \theta_1)\pi(\theta_1) + \ell(z | \theta_2)\pi(\theta_2)}$$

Consider the conditional problem. If the decision maker chooses model one, then the conditional expected utility is $v_1\bar{\pi}(\theta_1 | z)$ and similarly for choosing model two. Thus the Bayesian decision maker computes:

$$\max \{v_1\bar{\pi}(\theta_1 | z), v_2\bar{\pi}(\theta_2 | z)\}$$

and chooses a model in accordance with this maximization. This maximization is equivalent to

$$\max \{v_1\pi(\theta_1)\ell(z | \theta_1), v_2\pi(\theta_2)\ell(z | \theta_2)\}$$

expressed in terms of the prior, likelihood, and utility contributions. Taking logarithms and rearranging, we see that model θ_1 is selected if

$$\log \ell(z | \theta_2) - \log \ell(z | \theta_1) \leq \log v_1 - \log v_2 + \log \pi(\theta_1) - \log \pi(\theta_2) \quad (10.6)$$

If the right side of this inequality is zero, say because prior probabilities are the same across models and utility weights are also the same, then the decision rule says to maximize the log likelihood. More generally, both prior probabilities and utility weights come into play.

Notice that this decision rule is a threshold rule where we use the posterior probabilities to partition the \mathcal{Z} space. The subset \mathcal{Z}_1 contains all z such that inequality (10.6) is satisfied with a weak inequality. We arbitrarily include the indifference values in \mathcal{Z}_1 .

The Bayesian solution to the decision problem is posed assuming full confidence in a subjective prior distribution. In many problems, including ones with multiple sources of uncertainty, such confidence may well not be warranted. Such a concern might well have been the motivation behind Savage's remark:

... if I knew of any good way to make a mathematical model of these phenomena [vagueness and indecision], I would adopt it, but I despair of finding one. One of the consequences of vagueness is that we are able to elicit precise probabilities by self-interrogation in some situations but not in others.

— Personal communication from L. J. Savage to Karl Popper, 1957

10.5. An extreme response

Having assumed full confidence in a prior, suppose we now go to the other extreme and avoid imposing a prior altogether. Construct preferences by computing the conditional (on θ) expected utilities, the risk function, for each $\theta \in \Theta$. Then γ_2 is preferred to γ_1 if the conditional expected utility of the former exceeds that of the latter for all $\theta \in \Theta$. This, however, only implies a partial ordering among prize rules. Many such rules cannot be ranked. This partial ordering gives rise to a construct called **admissibility**, where an admissible $\delta \in \Delta$ cannot be dominated in the sense of this partial order.

10.5.1. Constructing admissible decision rules

One way to construct an admissible decision rule is to impose a prior and solve the resulting Bayesian decision problem. We give two situations in which this result necessarily applies, but there are other settings where this result is known to hold.

Proposition 10.1

If an *ex ante* Bayesian decision problem, (10.1), has a unique solution, then this Bayesian solution is admissible.

Proof. Let $\tilde{\delta}$ be a decision rule that weakly dominates a Bayesian decision rule, δ , in the sense that

$$\bar{U}(\delta | \theta) \leq \bar{U}(\tilde{\delta} | \theta)$$

for all $\theta \in \Theta$. The $\tilde{\delta}$ must also solve the *ex ante* Bayesian decision problem. Since the solution to the *ex ante* decision problem is unique, $\tilde{\delta} = \delta$.

Proposition 10.2

Suppose Θ has a countable number of elements. If a prior distribution $d\pi$ assigns positive probability to each element of Θ , then a decision rule that solves the Bayesian decision problem (10.1) is admissible.

Proof. Let $\tilde{\delta}$ be a decision rule that weakly dominates a Bayesian decision rule, δ , in the sense that

$$\bar{U}(\delta | \theta) \leq \bar{U}(\tilde{\delta} | \theta) \tag{10.7}$$

for $\theta \in \Theta$. Suppose that the prior, $d\pi$, used in constructing the decision rule, δ , assigns strictly positive probability to each value of $\theta \in \Theta$. Use this prior to form expectations of both sides of inequality (10.7),

$$\int_{\Theta} \bar{U}(\delta | \theta) d\pi(\theta) \leq \int_{\Theta} \bar{U}(\tilde{\delta} | \theta) d\pi(\theta)$$

But this latter inequality must hold with equality. Since each element of Θ has strictly positive prior probability, inequality (10.7) must also hold with equality. Therefore, δ must be admissible.

Remark 10.1

While we are primarily interested in the use of alternative subjective priors as a way to construct admissible decision rules, sufficient conditions have been derived under which we can find priors that give Bayesian justifications for all admissible decision rules. Such results come under the heading of **Complete class theorems**. See, for instance, [LeCam, 1955], [Ferguson, 1967], and [Brown, 1981].

10.5.2. A simple application reconsidered II

For illustration purposes, we again consider the model selection example. Consider a threshold decision rule of the form:

$$\delta_r(z) = \begin{cases} 1, & \log \ell(z | \theta_2) - \log \ell(z | \theta_1) \leq r \\ 0, & \log \ell(z | \theta_2) - \log \ell(z | \theta_1) > r. \end{cases} \quad (10.8)$$

Form the likelihood ratio and its logarithm:

$$m = \frac{\ell(z | \theta_2)}{\ell(z | \theta_1)}$$

$$y = \log m.$$

Let $f(y | \theta)$ be the implied density of the log-likelihood ratio under model θ relative to Lebesgue measure. We assume such a density representation as an example that can emerge for continuous data. Observe that the two densities are related:

$$f(y | \theta_2) = \exp(y) f(y | \theta_1)$$

from our use of log-likelihood ratios when constructing m and y . For a decision rule, δ_r , with threshold r , compute the two risks:

$$\begin{aligned}\bar{u}_1(r) &\stackrel{\text{def}}{=} \bar{U}(\delta_r | \theta_1) = v_1 \int_{-\infty}^r f(y | \theta_1) dy \\ \bar{u}_2(r) &\stackrel{\text{def}}{=} \bar{U}(\delta_r | \theta_2) = v_2 \int_r^{+\infty} \exp(y) f(y | \theta_1) dy\end{aligned}$$

where we include the multiplication by $\exp(y)$ in the second row to change the computation to use the probabilities under model θ_2 . This gives us a curve of feasible risks as a function of the threshold. The slope of this curve is given by

$$-\frac{v_2}{v_1} \exp(r).$$

We compute the second-order derivative of the curve as

$$\frac{\frac{d}{dr} \left[-\frac{v_2}{v_1} \exp(r) \right]}{v_1 f(r | \theta_1)} = -\frac{v_2 \exp(r)}{(v_1)^2 f(r | \theta_1)} < 0.$$

Hence the curve is concave.

Using prior probabilities to weight the two risks gives:

$$\pi(\theta_1) \bar{u}_1(r) + \pi(\theta_2) \bar{u}_2(r)$$

which we seek to maximize. The maximum is achieved when

$$-\frac{\pi(\theta_1)}{\pi(\theta_2)} = -\frac{v_2}{v_1} \exp(r).$$

Multiplying both sides of the equation by minus one and taking logarithms gives:

$$\log \pi(\theta_1) - \log \pi(\theta_2) = \log v_2 - \log v_1 + r. \quad (10.9)$$

Not surprisingly, by solving for r , we obtain formula [\(10.6\)](#), which we derived previously. Here we are instead interested in solving for the prior probabilities for given values for the utility weights v_1 and v_2 and likelihood ratio threshold r . This computation follows from substituting $\pi(\theta_2) = 1 - \pi(\theta_1)$ into formula [\(10.9\)](#). By varying r , we trace out a family of admissible decision rules that coincide with different choices of Bayesian solutions with different prior probabilities.

We illustrate this computation in Figures [Fig. 10.1](#) and [Fig. 10.2](#). Both figures report the upper boundary of the set of feasible risks for alternative decision rules. The risks along the boundary are attainable with admissible decision rules. The utility weights, v_1 and v_2 are both set to one in Figure [Fig. 10.1](#), and v_2 is

set to .5 in Figure [Fig. 10.2](#). Thus, the upper envelope of risks is flatter in Figure [Fig. 10.2](#) than in Figure [Fig. 10.1](#). The flatter curve implies prior probabilities that are close to being the same.

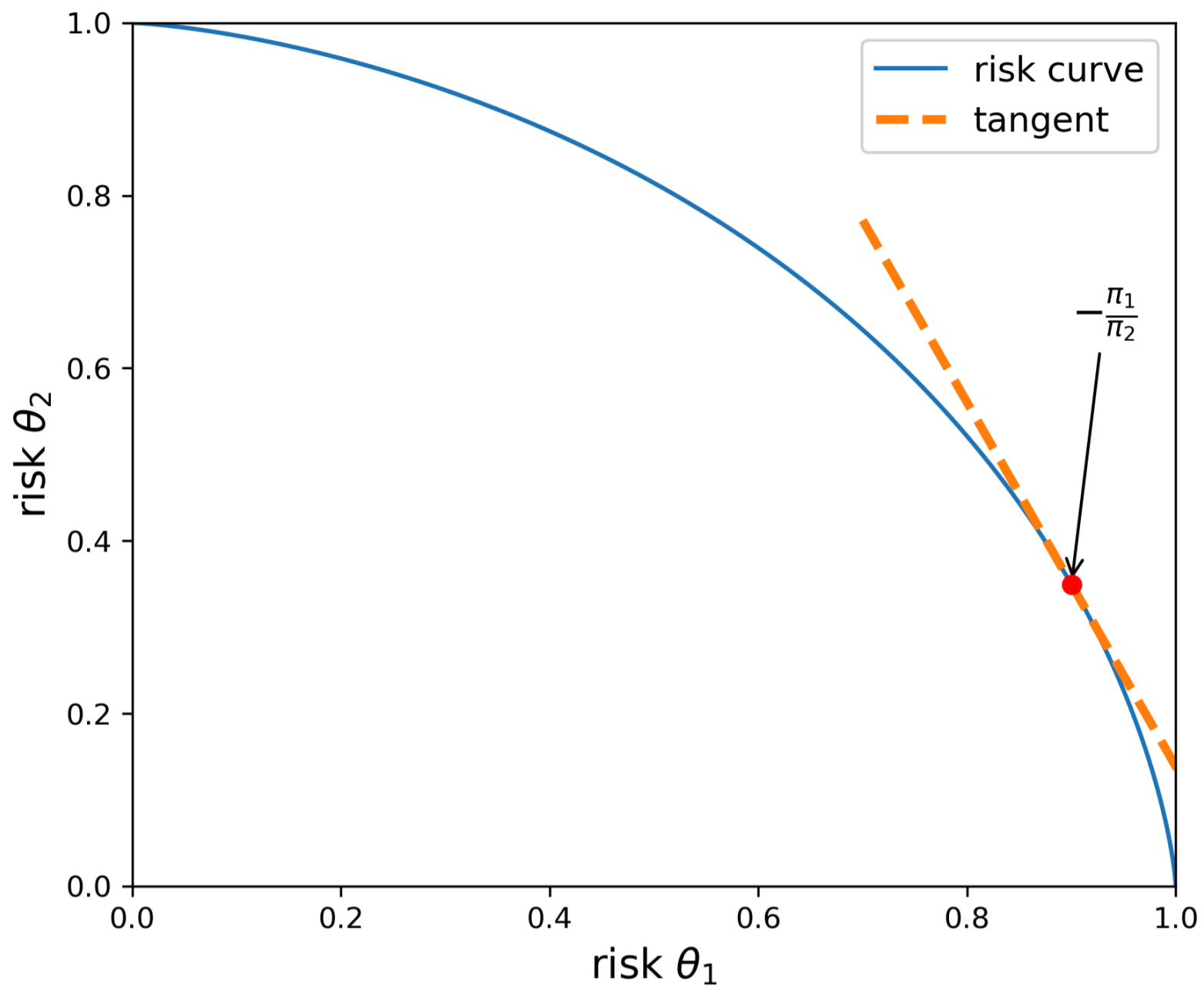


Fig. 10.1 The blue curve gives the upper boundary of the feasible set of risks. The utility function parameters are given by $v_1 = v_2 = 1$. When $\bar{u}_r(\theta_1) = .9$, the implied prior is given by $\pi(\theta_1) = .68$ and $\pi(\theta_2) = .32$, as implied by the slope of the tangent line.

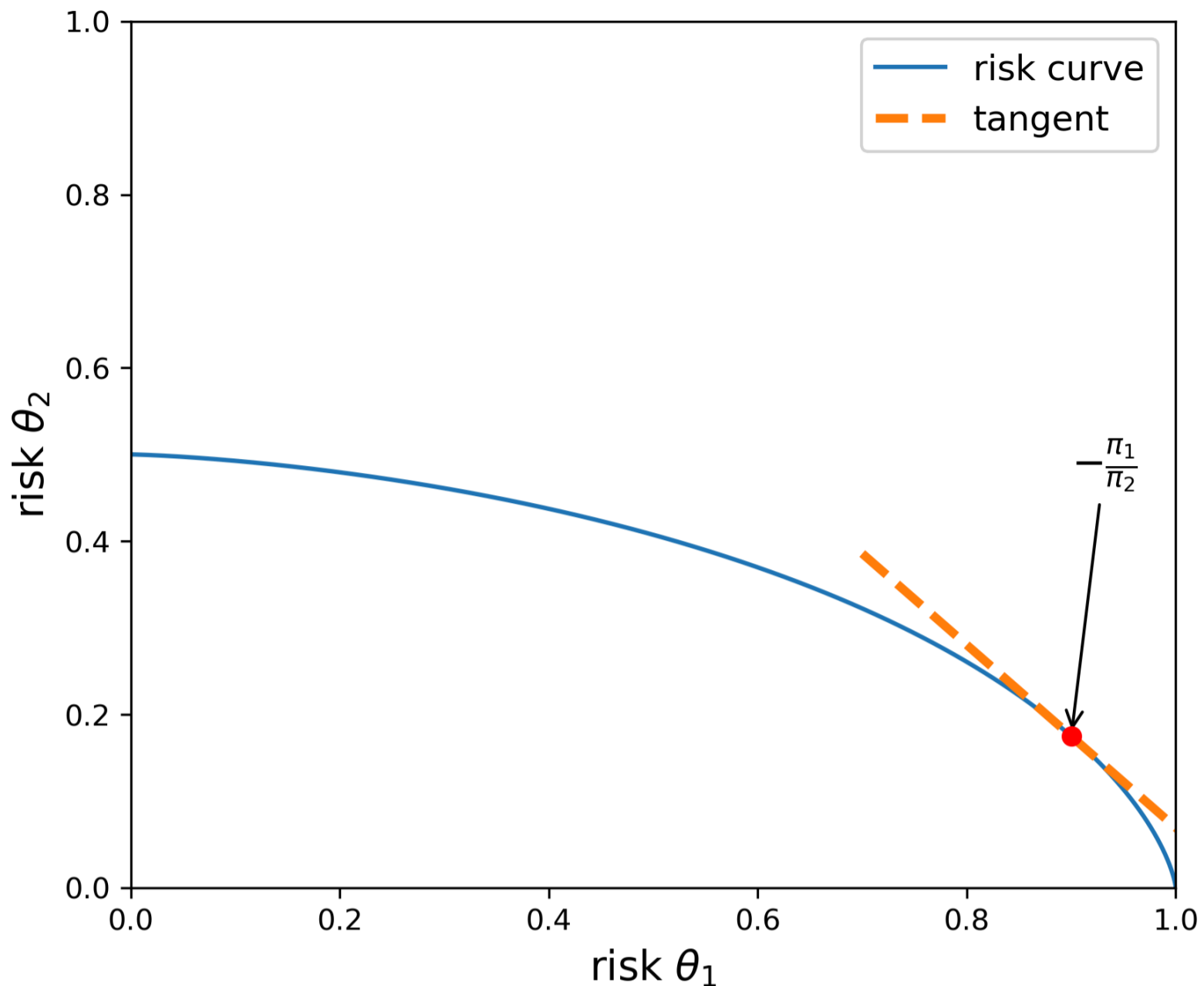


Fig. 10.2 The blue curve gives the upper boundary of the feasible set of risks. The utility function parameters are given by $v_1 = 1$, $v_2 = .5$. When $\bar{u}_r(\theta_1) = .9$, the implied prior is given by $\pi(\theta_1) = .51$ and $\pi(\theta_2) = .49$, as implied by the slope of the tangent line.

10.6. Divergences

To investigate prior sensitivity, we seek a convenient way to represent a family of alternative priors. We start with a baseline prior $d\pi(\theta)$. Consider alternative priors of the form $d\pi(\theta) = n(\theta)d\pi(\theta)$ for $n \geq 0$ satisfying:

$$\int_{\Theta} n(\theta)d\pi(\theta) = 1.$$

Call this collection \mathcal{N} . Thus the n 's in \mathcal{N} are expressed as densities relative to a baseline prior distribution.

Introduce a convex function ϕ to construct a divergence between a probability distribution represented by $n \neq 1$ and the baseline probability. Restrict ϕ to satisfy $\phi(1) = 0$ and $\phi''(1) = 1$ (normalization). As a measure of divergence, form

$$\int_{\Theta} \phi[n(\theta)] d\pi(\theta) \geq 0.$$

Of course, many such divergences can be constructed. Three interesting ones use the convex functions:

- $\phi(n) = -\log n$
- $\phi(n) = \frac{n^2-1}{2}$
- $\phi(n) = n \log n$

The divergence implied from the third choice is commonly used in applied probability theory and information theory. It is called **Kullback-Leibler divergence** or **relative entropy**. More generally, divergences for alternative choices of ϕ are called **ϕ -divergences** or **f-divergences**.

10.7. Robust Bayesian preferences and ambiguity aversion

Since the Bayesian ranking of prize rules depends on the prior distribution, we now explore how to proceed if the decision maker does not have full confidence in a specific prior. This leads naturally to an investigation of prior sensitivity. A decision or policy problem provides us with an answer to the question: sensitive to what. One way to investigate prior sensitivity is to approach it from the perspective of robustness. A robust decision rule then becomes one that performs well under alternative priors of interest. To obtain robustness guarantees, we are naturally led to minimization, providing us with a lower bound on performance. As we will see, prior robustness has very close ties to preferences that display ambiguity aversion. Just as risk aversion induces a form of caution in the presence of uncertain outcomes, ambiguity aversion induces a caution because of the lack of confidence in a single prior.

We explore prior robustness by using a version of **variational preferences** ([\[Maccheroni et al., 2006\]](#))

$$\min_{n \in \mathcal{N}} \int_{\Theta} \left(\int_{\mathcal{X}} U[\gamma(x), \theta] \ell(x|\theta) d\tau(x) \right) n(\theta) d\pi(\theta) + \xi_1 \int_{\Theta} \phi_1[n(\theta)] d\pi(\theta)$$

for $\xi_1 > 0$ and a convex function ϕ_1 such that $\phi_1(1) = 0$ and $\phi_1''(1) = 1$. The penalty parameter ξ_1 reflects the degree of ambiguity aversion. An arbitrarily large value of ξ_1 approximates subjective expected utility. Relatively small values of ξ_1 induce relatively large degrees of ambiguity aversion.

Remark 10.2

Axiomatic developments of decision theory in the presence of risk typically do not tell us the functional form for the utility function. Doing that requires additional inputs. An analogous observation applies to the axiomatic development of variational preferences by [\[Maccheroni et](#)

[al., 2006](#)]. Their axioms do not inform as to how to capture the cost associated with search over alternative priors.

Remark 10.3

The variational preferences of [Maccheroni et al., 2006](#) also include preferences with a constraint on priors:

$$\int_{\Theta} \phi_1[n(\theta)] d\pi(\theta) \leq \kappa.$$

The more restrictive axiomatic formulation of [Gilboa and Schmeidler, 1989](#) supports a representation with a constraint on the set of priors. In this case we use standard Karush-Kuhn-Tucker multipliers to model the preference relation:

$$\begin{aligned} \max_{\xi_1 \geq 0} \min_{n \in \mathcal{N}} \int_{\Theta} \left(\int_{\mathcal{X}} U[\gamma(x), \theta] \ell(x|\theta) d\tau(x) \right) n(\theta) d\pi(\theta) \\ + \xi_1 \left[\int_{\Theta} \phi_1[n(\theta)] d\pi(\theta) - \kappa \right]. \end{aligned}$$

10.7.1. Preferences with relative entropy divergence

Suppose we use $n \log n$ to construct our divergence measure. Recall the construction of the risk function

$$\bar{U}(\gamma | \theta) = \int_{\mathcal{X}} U[\gamma(x), \theta] \ell(x | \theta) d\tau(x)$$

Solve the Lagrangian:

$$\min_n \int_{\Theta} \bar{U}(\gamma | \theta) n(\theta) d\pi(\theta) + \xi_1 \int_{\Theta} \log n(\theta) n(\theta) d\pi(\theta) + \lambda \int_{\Theta} [n(\theta) - 1] d\pi(\theta)$$

This problem separates in terms of the choice of $n(\theta)$, and can be solved θ by θ . The first-order conditions are:

$$\bar{U}(\gamma | \theta) + \xi_1 + \xi_1 \log n(\theta) + \lambda = 0.$$

Solving for $\log n(\theta)$:

$$\log n(\theta) = -\frac{1}{\xi_1} \bar{U}(\gamma | \theta) - 1 - \frac{\lambda}{\xi_1}$$

Thus

$$n(\theta) \propto \exp \left[-\frac{1}{\xi_1} \bar{U}(\gamma | \theta) \right].$$

Imposing the integral constraint on n gives the solution:

$$n^*(\theta) = \frac{\exp \left[-\frac{1}{\xi_1} \bar{U}(\gamma | \theta) \right]}{\int_{\Theta} \exp \left[-\frac{1}{\xi_1} \bar{U}(\gamma | \theta) \right] d\pi(\theta)},$$

provided that the denominator is finite. This solution induces what is known as **exponential tilting**. The baseline probabilities are tilted towards lower values of $\bar{U}(\gamma | \theta)$. Plugging back into the minimization problem gives:

$$-\xi_1 \log \int_{\Theta} \exp \left[-\frac{1}{\xi_1} \bar{U}(\gamma | \theta) \right] d\pi(\theta). \quad (10.10)$$

This minimized objective is known to be a special case of **smooth ambiguity** preferences initially proposed by [\[Klibanoff et al., 2005\]](#), although these authors provide a different motivation for their ambiguity adjustment. The connection we articulate opens the door to a more direct link to challenges familiar to statisticians and econometricians wrestling with how to analyze and interpret data. Indeed [\[Cerrei-Vioglio et al., 2013\]](#) also adopt a robust statistics perspective when exploring smooth ambiguity aversion preferences. They use constructs and distinctions of the type we explored in [Chapter 1](#) in characterizing what is and is not learnable from the Law of Large Numbers.

10.7.2. Robust Bayesian decision problem

We extend Decision Problem [\(10.1\)](#) to include prior robustness by introducing a special case of a two-player, zero-sum game:

Game 11.4

$$\begin{aligned} \max_{\delta \in \Delta} \min_{n \in \mathcal{N}} \int_{\Theta} \left(\int_{\mathcal{X}} U(\Psi[\delta(z), x], \theta) \ell(x|\theta) d\tau(x) \right) n(\theta) d\pi(\theta) \\ + \xi_1 \int_{\Theta} \phi_1[n(\theta)] d\pi(\theta). \end{aligned} \quad (10.11)$$

Notice that in this formulation, the minimization depends on the choice of the prize rule γ and hence on the decision rule δ that constrains γ . This is to be expected as the prior with the most adverse consequences for the expected utility should plausibly depend on the potential course of action under consideration.

For a variety of reasons, it is of interest to investigate a related problem in which the order of extremization is exchanged:

Game 11.5

$$\begin{aligned} \min_{n \in \mathcal{N}} \max_{\delta \in \Delta} \int_{\Theta} \left(\int_{\mathcal{X}} U(\Psi[\delta(z), x], \theta) \ell(x|\theta) d\tau(x) \right) n(\theta) d\pi(\theta) \\ + \xi_1 \int_{\Theta} \phi_1[n(\theta)] d\pi(\theta). \end{aligned} \quad (10.12)$$

Notice that for a given n , the inner problem is essentially just a version of the Bayesian problem [\(10.1\)](#). The penalty term

$$\xi_1 \int_{\Theta} \phi_1[n(\theta)] d\pi(\theta)$$

is additively separable and does not depend on δ . In this formulation, we solve a Bayesian problem for each possible prior, and then minimize over the priors taking account of the penalty term. Provided that the outer minimization problem over n has a solution, n^* , the implied decision rule, δ_{n^*} , solves a Bayesian decision problem. As we know from section [Constructing admissible decision rules](#), this opens the door to verifying admissibility.

The two decision games, [\(10.11\)](#) and [\(10.12\)](#), essentially have the same solution under a Minimax Theorem. That is the implied value functions are the same and δ_{n^*} , from Game [\(10.12\)](#) solves Game [\(10.11\)](#) and gives the robust decision rule under prior ambiguity. This result holds under a variety of sufficient conditions. Notice that the objective for Game [\(10.11\)](#) is convex in n . A well known result due to [\[Fan, 1952\]](#) verifies the Minimax Theorem when the objective satisfies a generalization of concavity with a

convex constraint set Δ . While we cannot always justify this exchange, there are other sufficient conditions that are also applicable.

A robust Bayesian advocate along the lines of [Good, 1952], views the solution, say n^* , from Game (10.12) as a choice of a prior to be evaluated subjectively. It is often referred to as a “worst-case prior”, but an object that is of interest in its own right. For an application of this idea in economics see [Chamberlain, 2000].^[6] Typically, $n^*(\theta)d\pi(\theta)$ can be computed as part of an algorithm for finding a robust decision rule, and is arguably worth serious inspection. While we could just view robustness considerations as a way to select a prior, the (penalized) worst-case solutions can instead be viewed as a device to implement robustness. While they are worthy of inspection, just as with the baseline prior probabilities, the worst-case probabilities are not intended to be a specification of a fully confident input of subjective probabilities and are dependent on the utility function and constraints imposed on the decision problem.

10.7.2.1. A simple application reconsidered III

We again use the model selection example to illustrate ambiguity aversion in the presence of a relative entropy cost of a prior deviating from the baseline. Since the Minimax Theorem applies, we focus our attention on admissible decision rules parameterized by thresholds of the form (10.8). With this simplification, we use formula (10.10) and solve the scalar maximization problem:

$$\max_r -\xi_1 \log \left(\exp \left[-\frac{1}{\xi_1} \bar{u}_1(r) \right] \pi(\theta_1) + \exp \left[-\frac{1}{\xi_1} \bar{u}_2(r) \right] \pi(\theta_2) \right)$$

This maximization problem may be solved easily using a numerical method. Nevertheless, the first-order conditions are:

$$\begin{aligned} & \frac{\exp \left[-\frac{1}{\xi_1} \bar{u}_1(r) \right] \pi(\theta_1)}{\exp \left[-\frac{1}{\xi_1} \bar{u}_1(r) \right] \pi(\theta_1) + \exp \left[-\frac{1}{\xi_1} \bar{u}_2(r) \right] \pi(\theta_2)} f(r | \theta_1) \\ & - \frac{\exp \left[-\frac{1}{\xi_1} \bar{u}_2(r) \right] \pi(\theta_2)}{\exp \left[-\frac{1}{\xi_1} \bar{u}_1(r) \right] \pi(\theta_1) + \exp \left[-\frac{1}{\xi_1} \bar{u}_2(r) \right] \pi(\theta_2)} \exp(r) f(r | \theta_1) \\ & = 0. \end{aligned}$$

Notice the exponential tilting adjustment to the prior probabilities. The resulting threshold decision rule is admissible, which can be verified using these altered probabilities. The first-order conditions can be further simplified to

$$-\frac{\bar{u}_1(\mathbf{r})}{\xi_1} + \frac{\bar{u}_2(\mathbf{r})}{\xi_1} + \log \pi(\theta_1) - \log \pi(\theta_2) + \log v_1 - \log v_2 = \mathbf{r}.$$

We consider two limiting cases. When $\xi_1 \rightarrow \infty$, the first two terms drop out and the first-order conditions are the ones for maximizing to the subjective expected utility:

$$\bar{u}_1(\mathbf{r})\pi(\theta_1) + \bar{u}_2(\mathbf{r})\pi(\theta_2).$$

To investigate the solution when $\xi_1 \rightarrow 0$, we first multiply [\(10.13\)](#) by ξ_1 in advance of taking the limit. The result

$$-\bar{u}_1(\mathbf{r}) + \bar{u}_2(\mathbf{r}) = 0.$$

Thus the robustly optimal \mathbf{r} is determined by making the two possible risks are equal when finding the optimal \mathbf{r} .

For a more general decision rule, δ form the two risks:

$$\frac{\bar{U}(\delta | \theta_1)}{\bar{U}(\delta | \theta_2)}.$$

The minimization in this limiting case assigns probability one to lowest risk with an objective:

$$\min\{\bar{U}(\delta | \theta_1), \bar{U}(\delta | \theta_2)\}$$

Graphically, the objective for any point to the left of the 45 degree line from the origin equals the outcome of a vertically downward movement to that same line. Analogously, the objective for any point to the right of the 45 degree line from the origin equals the outcome of a horizontally leftward movement to that same line. Thus the maximizing δ is obtained at the intersection of the 45 degree line and the boundary of the risk set. Along the 45 degree line, the choice of prior is inconsequential because the two risks are the same. Nevertheless, the “worst-case” prior is determined by slope of the risk curve at the intersection point. Recall that we defined this prior after exchanging orders of extremization. Figures [Fig. 10.3](#) and [Fig. 10.4](#) illustrate this outcome for the two risk curves on display in Figures [Fig. 10.1](#) and [Fig. 10.2](#).

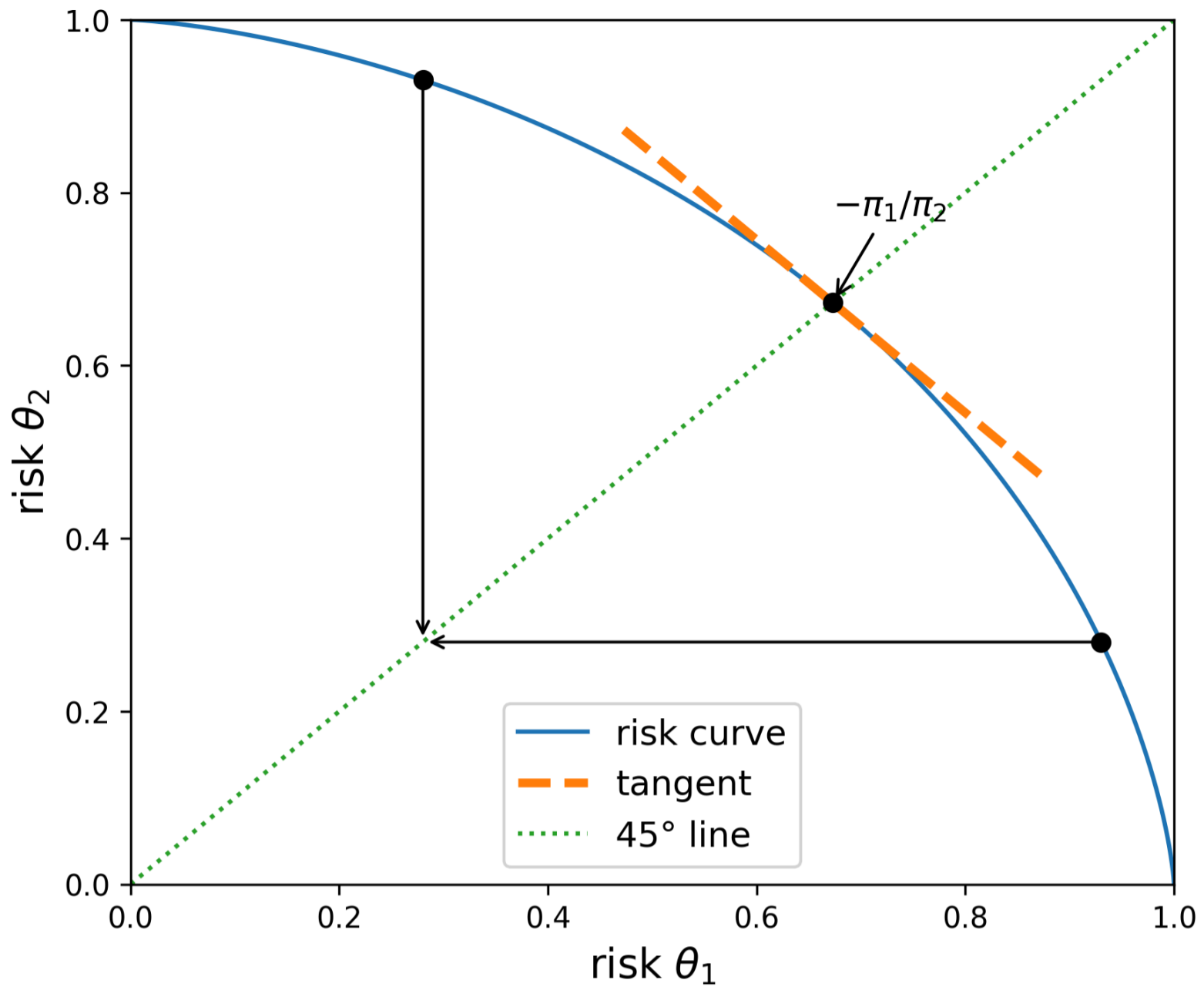


Fig. 10.3 The blue curve gives the upper boundary of the feasible set of risks. The utility function parameters are given by $v_1 = v_2 = 1$. The implied worst-case prior is given by $\pi(\theta_1) = .5$ and $\pi(\theta_2) = .5$, as implied by the slope of the tangent line at the intersection with the 45 degree line from the origin.

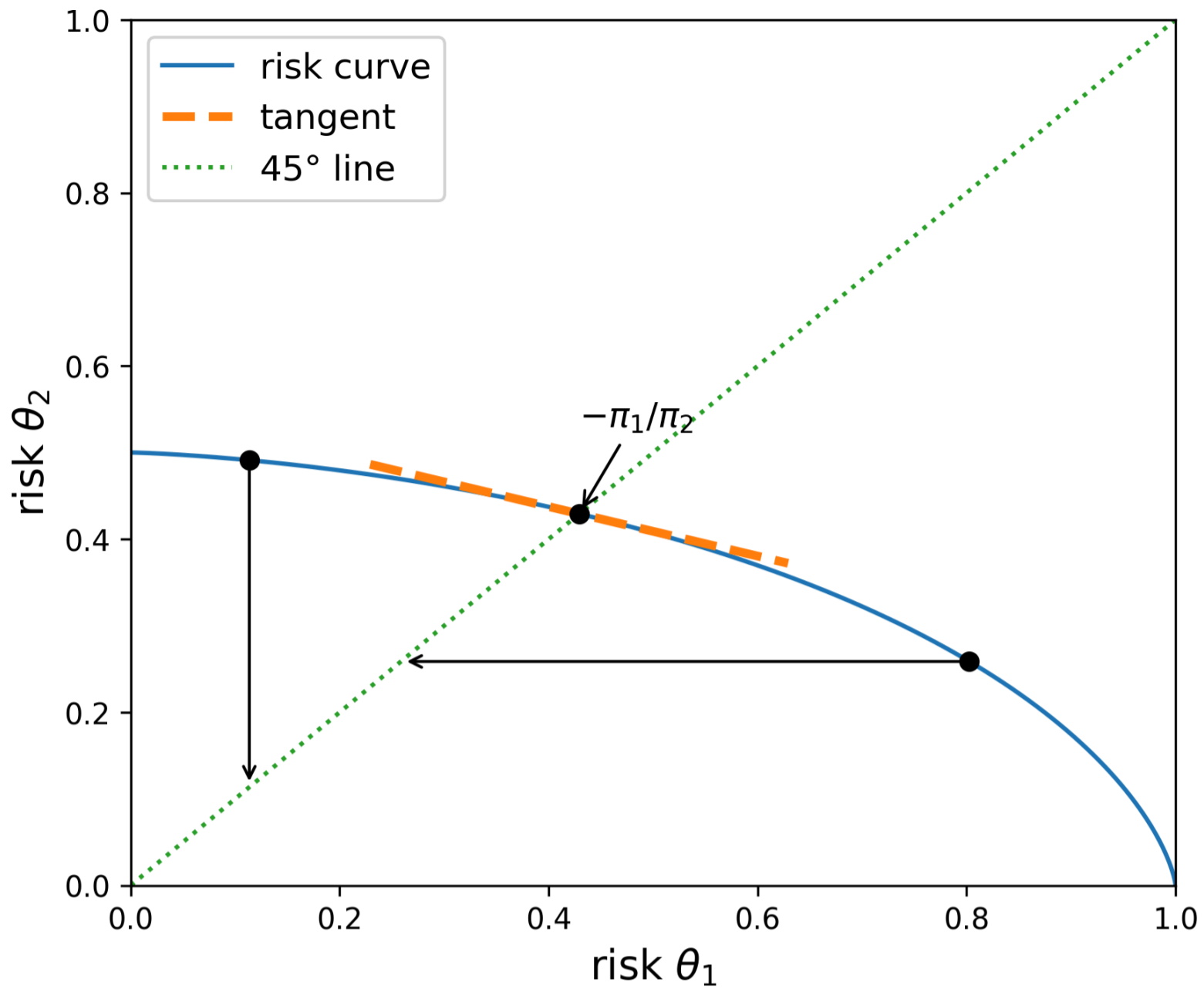


Fig. 10.4 The blue curve gives the upper boundary of the feasible set of risks. The utility function parameters are given by $v_1 = 1$ and $v_2 = .5$. The implied worst-case prior is given by $\pi(\theta_1) = .22$ and $\pi(\theta_2) = .78$, as implied by the slope of the tangent line at the intersection with the 45 degree line from the origin.

For positive values of ξ_1 , the implied worst-case priors are between the baseline π ($\xi = \infty$) and the worst-case without restricting the prior probabilities ($\xi = 0$). Observe that the worst-case priors depend on the utility weights, v_1 and v_2 . See Figures [Fig. 10.5](#) and [Fig. 10.6](#) as illustrations.

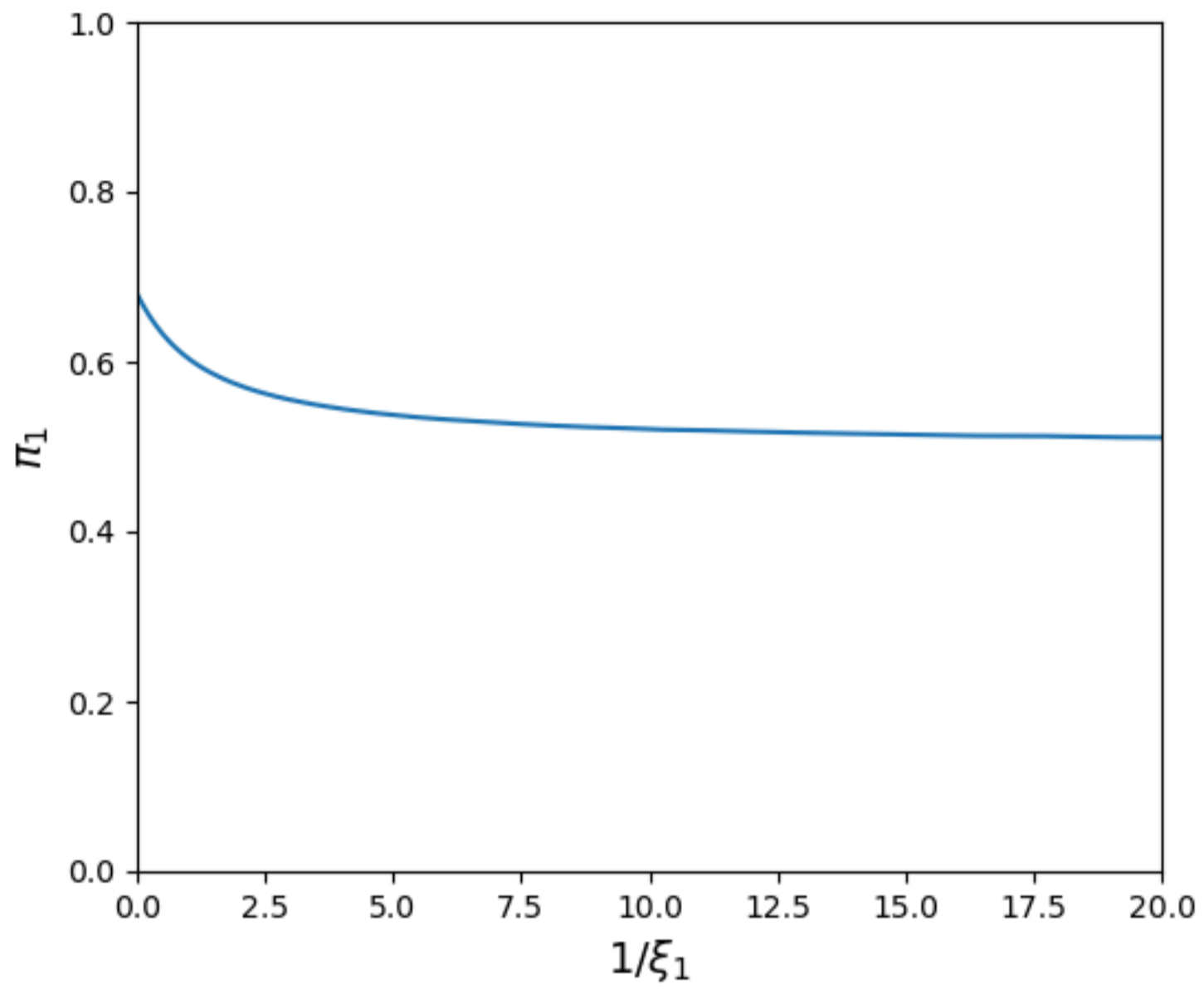


Fig. 10.5 Minimizing prior probabilities for θ_1 as a function of $1/\xi_1$ when the baseline prior probabilities are $\pi(\theta_1) = .68$ and $\pi(\theta_2) = .32$ and the utility parameters are $v_1 = v_2 = 1$.

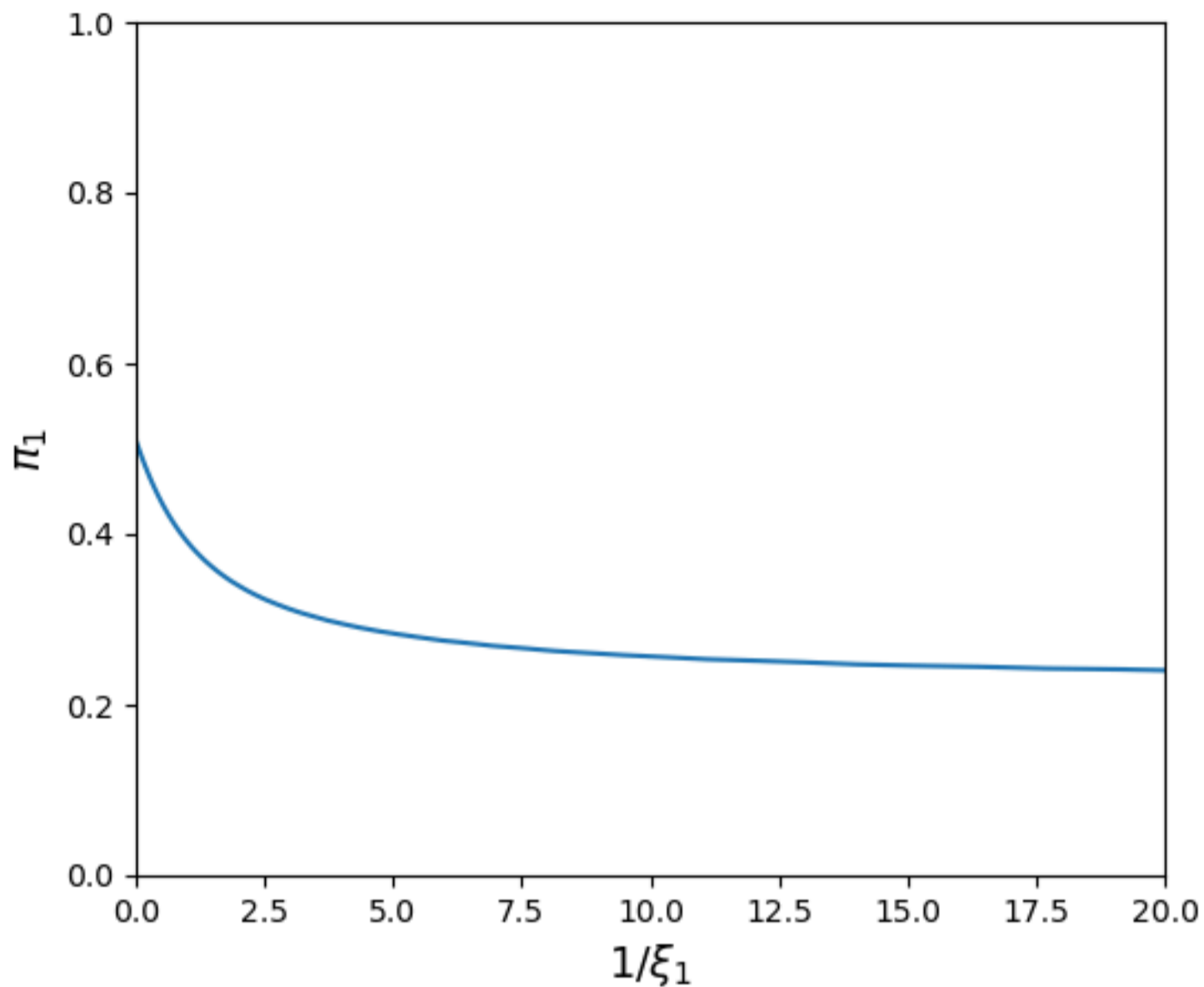


Fig. 10.6 Minimizing prior probabilities for θ_1 as a function of $1/\xi_1$ when the baseline prior probabilities are $\pi(\theta_1) = .51$ and $\pi(\theta_2) = .49$ and the utility parameters are $v_1 = 1$ and $v_2 = .5$.

The model selection example dramatically understates the potential value of ambiguity aversion as a way to study prior sensitivity. In typical applied problems, the subjective probabilities are imposed on a much richer collection of alternative models including families of models indexed by unknown parameters. In such problems the outcome is more subtle since the minimization isolates dimensions along which prior sensitivity has the most adverse impacts on the decision problem and are perhaps most worthy of further consideration. This can be especially important in problems where baseline priors are imposed “as a matter of computational convenience.”

10.8. Using uncertainty aversion to represent concerns about model misspecification

Two prominent statisticians remarked on how model misspecification is pervasive:

“Since all models are wrong, the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.” — Box (1976).

“... it does not seem helpful just to say that all models are wrong. The very word ‘model’ implies simplification and idealization. The idea that complex physical, biological or sociological systems can be exactly described by a few formulae is patently absurd. The construction of idealized representations that capture important stable aspects of such systems is, however, a vital part of general scientific analysis and statistical models, especially substantive ones ...” — Cox (1995).

Other scholars have made similar remarks. Robust control theorists have suggested one way to address this challenge, an approach that we build on in the discussion that follows. Motivated by such sentiments, [Cerrea-Vioglio *et al.*, 2025] extended decision theory axioms to accommodate misspecification concerns. To proceed, we focus our analysis of model misspecification by conditioning θ . Our decision maker will no longer have full confidence in the probability specification $\ell(x | \theta)d\tau(x)$. In other words our treatment of this probability as a source of risk or say as a “roulette wheel” lottery in the sense of [Anscombe and Aumann, 1963]. Instead of exploring sensitivity to probability specifications over a parameter space Θ , our focus shifts to exploring sensitivity to alternative probability specifications over the space \mathcal{X} for which $\ell(x | \theta)d\tau(x)$ merely serves as baseline. Think of this baseline as implied by a model of interest. We then consider alternative probability assignments to entertain potential model misspecification.

10.8.1. Basic approach

To focus on the misspecification of a specific model, we fix θ , but vary the likelihood function as a way to investigate likelihood sensitivity. We then replace ℓ with $m\ell$ satisfying:

$$\int_{\mathcal{X}} m(x | \theta)\ell(x | \theta)d\tau(x) = 1,$$

and denote the set of all such density ratios m as \mathcal{M} .

Observe that $m(x | \theta)$ can be viewed as the ratio of two densities. Consider two alternative probability densities (with respect to $d\tau(x)$) for shock/signal probabilities: $\ell(x | \theta)$ and $\tilde{\ell}(x | \theta)$ and let:

$$m(x | \theta) = \frac{\tilde{\ell}(x | \theta)}{\ell(x | \theta)}$$

where we assume $\ell(x | \theta) > 0$ for $x \in \mathcal{X}$. Then by construction:

$$m(x | \theta)\ell(x | \theta) = \tilde{\ell}(x | \theta),$$

and

$$\int_{\mathcal{X}} m(x | \theta) \ell(x | \theta) d\tau(x) = \int_{\mathcal{X}} \tilde{\ell}(x | \theta) d\tau(x) = 1.$$

We use density ratios to capture alternative models as inputs into divergence measures. Let ϕ_2 be a convex function such that $\phi_2(1) = 0$ and $\phi_2''(1) = 1$ (normalization). Instead of specifying a divergence between probabilities over the parameter space, Θ , we now impose the divergence over the \mathcal{X} space:

$$\int_{\mathcal{X}} \phi_2 [m(x | \theta)] \ell(x | \theta) d\tau(x).$$

We use this divergence to limit or constrain our search over alternative probability models. In this approach we deliberately avoid imposing a prior distribution over the space of densities (with respect to $d\tau(x)$).

Preferences for model robustness rank alternative prize rules, γ , by solving:

$$\min_{m \in \mathcal{M}} \int_{\mathcal{X}} (U[\gamma(x), \theta] m(x | \theta) + \xi_2 \phi_2 [m(x | \theta)]) \ell(x | \theta) d\tau(x). \quad (10.14)$$

for a penalty parameter $\xi_2 > 0$. The parameter, ξ_2 , dictates the strength of the restraint in an exploration of possible model misspecifications.

10.8.2. Relative entropy divergence

This approach to model misspecification has direct links to *robust control theory* in the case of relative entropy divergence. Suppose that $\phi_2(m) = m \log m$ (relative entropy). Then by imitating previous computations, we find that the outcome of the minimization in [\(10.14\)](#) is

$$-\xi_2 \log \int_{\mathcal{X}} \exp \left[- \left(\frac{1}{\xi_2} \right) U[\gamma(x), \theta] \right] \ell(x | \theta) d\tau(x)$$

Remark 10.4

Robust control emerged from the study of optimization of dynamical systems. The use of the relative entropy divergence showed up prominently in [\[Jacobson, 1973\]](#) and later in [\[Whittle, 1981\]](#), [\[Petersen et al., 2000\]](#) and many other related papers as a response to the excessive simplicity of assuming shocks to dynamical systems that were iid and mean zero with normal distributions. [\[Hansen and Sargent, 1995\]](#) and [\[Hansen and Sargent, 2001\]](#) showed how to reformulate the insights from robust control theory to apply to dynamical economic systems with

recursive formulations and [Hansen et al., 1999] used these ideas in an initial empirical investigation.

When we use relative entropy as a measure of divergence, we have the ability to factor likelihoods in convenient ways. Recall the partitioning $x = (w, z)$, where the decision rule can only depend on z and the prize rule on the entire x vector. As in (10.2), factor $\ell(\cdot | \theta)$ and τ as:

$$\begin{aligned} d\tau(x) &= d\tau_2(w)d\tau_1(z) \\ \ell(x | \theta) &= \ell_2(w | z, \theta)\ell_1(z | \theta). \end{aligned}$$

Add to this a factorization of $m(\cdot | \theta)$,

$$\begin{aligned} m(x | \theta) &= m_2(w | z, \theta)m_1(z | \theta) \\ \int_{\mathcal{W}} m_2(w | \theta, z)\ell_2(w | z, \theta)d\tau_2(w) &= 1 \\ \int_{\mathcal{Z}} m_1(z | \theta)\ell_1(z | \theta)d\tau_1(z) &= 1 \end{aligned} \tag{10.15}$$

Let \mathcal{M}_1 denote the set of $m_1(z | \theta) \geq 0$ that satisfy the relevant integral constraint in (10.15), and similarly let \mathcal{M}_2 be the set of $m_2(w | z, \theta) \geq 0$ that satisfy the relevant integral constraint.

Using these factorizations, the relative entropy may be written as:

$$\begin{aligned} \int_{\mathcal{Z}} \left(\int_{\mathcal{W}} \log[m_2(w | z, \theta)]m_2(w | z, \theta)d\tau_2(w) + \log[m_1(z | \theta)] \right) \\ \times m_1(z | \theta)\ell_1(z | \theta)d\tau_1(z) \end{aligned} \tag{10.16}$$

where the first contribution features the potential misspecification of the conditional probability over \mathcal{W} and the second term features the potential misspecification of the probability over \mathcal{Z} .

Rewrite the expected utility function in (10.14) with an inner integral:

$$\begin{aligned} \int_{\mathcal{Z}} \left[\int_{\mathcal{W}} U(\Psi[\delta(z), w, z], \theta)m_2(w | z, \theta)\ell_2(w | z, \theta)d\tau_2(w) \right] \\ \times m_1(z | \theta)\ell_1(z | \theta)d\tau_1(z) \end{aligned} \tag{10.17}$$

Notice that both the formulas (10.16) and (10.17) scale linearly in $m_1(z | \theta)$ and that the additional relative entropy term depends only on z . Thus we can use a conditional objective when solving for robust decision rule δ :

Game 11.6

$$\begin{aligned}
& \max_{\delta(z) \in \Delta(z)} \min_{m_2 \in \mathcal{M}_2} \\
& \int_{\mathcal{W}} U(\Psi[\delta(z), w, z], \theta) m_2(w | z, \theta) \ell_2(w | z, \theta) d\tau_2(w) \\
& + \xi_2 \int_{\mathcal{W}} \log[m_2(w | z, \theta)] m_2(w | z, \theta) \ell_2(w | z, \theta) d\tau_2(w)
\end{aligned} \tag{10.18}$$

where constraint set Δ satisfies the separability constraint [\(10.3\)](#).

Notice that the minimizing over m_1 does not play a role in constructing a robustly optimal decision in this setup since the decision rule can depend what value of z is realized.

Remark 10.5

Rather than conditioning on a parameter value, [\[Chamberlain, 2020\]](#) argues for exploring the potential misspecification of the so called predictive density. Given a baseline probability density over \mathcal{X} conditioned on a parameter vector and a baseline prior over the family Θ of potential models, form a density over \mathcal{X}

$$\bar{\ell}(x) \stackrel{\text{def}}{=} \int_{\Theta} \ell(x | \theta) d\pi(\theta).$$

Use this single density over \mathcal{X} as the basis for a robustness adjustment. In what follows we explore approaches that treat potential prior misspecification and model misspecification as distinct concerns.

Remark 10.6

It is useful to compare two approaches to robustness that we have taken. One possibility is that the decision maker starts with a baseline prior over parameter vectors and considers consequences of misspecifying that prior. Under this approach, the decision maker takes as given the parameterized family of densities, $\ell(x | \theta)$ that imply a corresponding parameterized family of probabilities over the set of parameter Θ 's as correctly specified. As an alternative, instead of parameterizing a family of densities, suppose we start one such density and explore its potential misspecification.

Our setup allows the parameter space to be infinite dimensional. Consider a prior that is consistent with a “nonparametric Bayesian approach to decision making. Specifying a prior over the infinite-dimensional space exposes a challenge associated with “nonparametric Bayesian” methods. The prior probability 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. See [[Freedman, 1963](#)], [[Sims, 1971](#)], and [[Diaconis and Freedman, 1986](#)] for formal elaborations. 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 the potential misspecification of a of a probability distribution on \mathcal{X} .

10.9. Robust Bayes with model misspecification

So far we have explored two misspecification concerns in separate formulations. We now explore the potential misspecification of $\ell(x | \theta)d\tau(x)$ simultaneously with a prior π over the parameters.

It requires a different decision theory perspective as developed recently in [[Cerrei-Vioglio et al., 2025](#)]. Their contribution was motivated by earlier contributions [[Hansen and Sargent, 2007](#)] and [[Hansen and Sargent, 2022](#)] that proposed representations without formal axiomatic defenses. The latter paper presents a class of continuous-time diffusion models for which the robustness distinction between likelihood and prior potential misspecification is captured by featuring a set of “structured” models that could each be misspecified.

10.9.1. A modified decision theory formulation

Combining concerns for likelihood and prior misspecification pushes the decision problem outside the Savage and Anscombe-Aumann frameworks. [[Cerrei-Vioglio et al., 2025](#)] recently proposed an extension of the Anscombe-Aumann formulation. Their advance was motivated by earlier contributions [[Hansen and Sargent, 2007](#)] and [[Hansen and Sargent, 2022](#)] that proposed representations of decision problems, but they did not provide an axiomatic defenses for these representations. The latter paper presents a class of continuous-time diffusion models for which the robustness distinction between model and prior potential misspecification is captured by featuring a set of “structured” models, each of which could each be misspecified.

[[Cerrei-Vioglio et al., 2025](#)] incorporate a two-preference framework proposed previously by [[Gilboa et al., 2010](#)]. One of the preferences are referred to as “mental preferences.” These preferences are incomplete and only provide a partial ordering induced by *unanimity* over all *structured models*. We saw an analogous approach in our treatment of admissibility. Whereas we defined admissibility in terms of risk functions, we allow the mental preferences to incorporate model misspecification concerns of the type that we analyzed in the previous section. We then construct “behavioral preferences” that are *complete* and used in actual decision making by entertaining *ambiguity aversion* over structured probability models. [[Cerrei-Vioglio et al., 2025](#)] provide a formal axiomatic defense for proceeding in this fashion. What follows are two approaches that can accommodate within this extended decision theory.

10.9.2. Approach one

Form a convex, compact constraint set of prior probabilities, $\mathcal{N}_o \subset \mathcal{N}$. Represent preferences over γ using:

$$\begin{aligned} \min_{n \in \mathcal{N}_o} \min_{m \in \mathcal{M}} \int_{\Theta} \left(\int_{\mathcal{X}} U[\gamma(x), \theta] m(x | \theta) \ell(x | \theta) d\tau_o(x) \right) n(\theta) d\pi(\theta) \\ + \xi_2 \int_{\Theta} \left(\int_{\mathcal{X}} \phi_2[m(x | \theta)] \ell(x | \theta) d\tau(x) \right) n(\theta) d\pi(\theta) \end{aligned}$$

This is a special case of what [[Gilboa and Schmeidler, 1989](#)] call “max-min expected utility” with a nonunique prior.

10.9.3. Approach two

Represent preferences over γ using:

$$\begin{aligned} \min_{n \in \mathcal{N}} \min_{m \in \mathcal{M}} \int_{\Theta} \left(\int_{\mathcal{X}} U[\gamma(x), \theta] m(x | \theta) \ell(x | \theta) d\tau(x) \right) n(\theta) d\pi(\theta) \\ + \xi_2 \int_{\Theta} \left(\int_{\mathcal{X}} \phi_2[m(x | \theta)] \ell(x | \theta) d\tau(x) \right) n(\theta) d\pi(\theta) \\ + \xi_1 \int_{\Theta} \phi_1[n(\theta)] d\pi(\theta) \end{aligned}$$

Note that this approach uses a scaled version of a *joint divergence* over (m, n) , as reflected in the term

$$\xi_2 \int_{\Theta} \left(\int_{\mathcal{X}} \phi_2[m(x | \theta)] \ell(x | \theta) d\tau(x) \right) n(\theta) d\pi(\theta) + \xi_1 \int_{\Theta} \phi_1[n(\theta)] d\pi(\theta)$$

Either can be used when formulating decision problems. For instance, the associated decision problem for the second approach is:

Game 11.7

$$\begin{aligned} \max_{\delta \in \Delta} \min_{n \in \mathcal{N}} \min_{m \in \mathcal{M}} \int_{\Theta} \left[\int_{\mathcal{X}} U(\Psi[\delta(z), x], \theta) m(x | \theta) \ell(x | \theta) d\tau(x) \right] n(\theta) d\pi(\theta) \\ + \xi_2 \int_{\Theta} \left(\int_{\mathcal{X}} \phi_2[m(x | \theta)] \ell(x | \theta) d\tau(x) \right) n(\theta) d\pi(\theta) \\ + \xi_1 \int_{\Theta} \phi_1[n(\theta)] d\pi(\theta). \end{aligned}$$

Remark 10.7

Consider the special case in which $\xi_1 = \xi_2 = \xi$, and use this common parameter to scale a relative entropy divergence. Then the combined robustness penalty contribution to preferences is measured by ξ multiplying the relative entropy of the joint distribution $m(x | \theta)n(\theta)\ell(x | \theta)d\tau(x)d\pi(\theta)$ relative to the baseline $\ell(x | \theta)d\tau(x)d\pi(\theta)$ and given by:

$$\xi \int_{\Theta} \left(\int_{\mathcal{X}} \log[m(x | \theta)]m(x | \theta)\ell(x | \theta)d\tau(x) \right) n(\theta)d\pi(\theta) + \int_{\Theta} \log[n(\theta)]n(\theta)d\pi(\theta)$$

Joint densities can be factored in alternative ways. In solving the robust decision problem, a different factorization is more convenient that conditions on the data available for making decisions. As we have done previously, form three contributions of the joint density under the baseline:

$$\begin{aligned} \ell(x | \theta)d\tau(x)d\pi(\theta) &= \ell_2(w | z, \theta)\ell_1(z | \theta)d\tau_2(w)d\tau_1(z)d\pi(\theta) \\ &= \ell_2(w | z, \theta)d\tau_2(w)d\bar{\pi}(\theta | z) \left[\int_{\Theta} \ell_1(z | \theta)d\pi(\theta) \right] d\tau_1(z). \end{aligned}$$

Notice that the last term in the factorization depends only on z , whereas the decision rule conditions on z . We now explore likelihood misspecification using $m_2(w | z, \theta)$ satisfying:

$$\int_{\mathcal{W}} m_2(w | z, \theta)\ell_2(w | z, \theta)d\tau_2(w) = 1$$

and posterior misspecification using $\bar{n}(\theta | z)$ where

$$\int_{\Theta} \bar{n}(\theta | z)d\bar{\pi}(\theta | z) = 1.$$

Thus to compute δ we again have a type of separation where we alter the posterior implied by the baseline probability specifications and the conditional density for $\ell_2((w | z, \theta)$. We may alter the density

$$\int_{\Theta} \ell_1(z | \theta)d\pi(\theta)$$

in an analogous way, but this exploration will have no impact on the robustly optimal choice of δ .

The logic behind this conditioning is entirely analogous to the argument we provided for the conditional version of the Bayesian decision problem. The objective for the conditional robust game is:

💡 Game 11.8

$$\begin{aligned} & \max_{\delta(z) \in \Delta(z)} \min_{\bar{n} \in \bar{\mathcal{N}}} \min_{m_2 \in \mathcal{M}_2} \\ & \int_{\Theta} \left[\int_{\mathcal{W}} U(\Psi[\delta(z), w, z], \theta) m_2(w | z, \theta) \ell_2(w | z, \theta) d\tau_2(w) \right] \bar{n}(\theta | z) d\bar{\pi}(\theta | z) \\ & + \xi \int_{\Theta} \left(\int_{\mathcal{W}} \log[m_2(w | z, \theta)] m_2(w | z, \theta) \ell_2(w | z, \theta) d\tau_2(w) \right) \bar{n}(\theta | z) d\bar{\pi}(\theta | z) \\ & + \xi \int_{\Theta} \log[\bar{n}(\theta | z)] \bar{n}(\theta | z) d\bar{\pi}(\theta | z) \end{aligned}$$

where constraint set Δ satisfies the separability constraint [\(10.3\)](#).

If the utility function does not depend directly on θ , we may perform the robustness analysis on a conditional version starting with the predictive density over \mathcal{W}

$$\int_{\Theta} \ell_2(w | x, \theta) d\bar{\pi}(\theta | z),$$

which is a conditional version of the proposal in [\[Chamberlain, 2020\]](#).

While $\xi_1 = \xi_2$ opens the door to computing robustly optimal decision rules via conditioning, we find we find this penalty equalization to be too constraining for many problems of interest. For example, [\[Hansen and Miao, 2018\]](#) describe continuous environments for which is important that $\xi_1 \neq \xi_2$.

10.10. A dynamic robust decision problem under commitment

So far, we have studied static decision problems. This formulation can accommodate dynamic problems by allowing for decision rules that depend on histories of data available up until the date of the decision. To convert to a static problem we pose the corresponding two-player game as one with a date zero “commitment” to robustly optimal decision rules at the initial date.

Recall from [Chapter 1](#) that we use an ergodic decomposition to identify a family of statistical models that are dynamic in nature, along with probabilities across models that are necessarily subjective as they are

not revealed by data. The example incorporates this perspective.

Consider the following example of a real investment problem with a single stochastic option for transferring goods from one period to another. This problem could be a planner's problem supporting a competitive equilibrium outcome associated with a stochastic growth model with a single capital good. Introduce an exogenous stochastic technology process that has an impact on the growth rate of capital. We design this stochastic technology process to capture what a previous literature in macro-finance has referred to as "long-run risk." For instance, see [\[Bansal and Yaron, 2004\]](#).^[7]

We extend this formulation by introducing a parameterized family of stochastic technology processes. This gives rise to a family of "structured models". The investor's exploration of the entire family of these processes reflects uncertainty among possible structured models. We also allow the investor to entertain misspecification concerns for each of the structured models

The exogenous (system) state vector Z_t used to capture fluctuations in the technological opportunities and the bi-variate, standard, normaly distribe shock vector W_{t+1} . We build the exogenous technology process from the shocks in a parameter dependent way:

$$Z_{t+1} = \psi(Z_t, W_{t+1}, \theta)$$

for a given initial condition Z_0 .

For instance, in long-run risk modeling one component of Z_t evolves as a first-order autoregression:

$$Z_{t+1}^1 = \mathbf{a}_\theta^1 Z_t^1 + \mathbf{b}_\theta^1 \cdot W_{t+1}$$

and another component is given by:

$$Z_{t+1}^2 = \mathbf{a}_\theta^2 + \mathbf{b}_\theta^2 \cdot W_{t+1}$$

The matrix: $[\mathbf{b}_\theta^1 \quad \mathbf{b}_\theta^2]$ is presumed to be nonsingular for all values of θ .

At each time t the investor observes past and current values $\mathbf{Z}^t = \{Z_0, Z_1, \dots, Z_t\}$ of the technology process, but does not know θ and does not directly observe the random shock vector W_t .

Similarly, we consider a recursive representation of capital evolution given by:

$$K_{t+1} = K_t \varphi(I_t/K_t, Z_{t+1}) \tag{10.19}$$

where φ is used capture adjustment costs for changing the capital stock. Consumption $C_t \geq 0$ and investment $I_t \geq 0$ are constrained by an output relation:

$$C_t + I_t = \alpha K_t$$

for a pre-specified initial condition K_0 . The parameter α captures the productivity of capital. By design this technology is homogeneous of degree one, which opens the door to stochastic growth as assumed in long-run risk models. In what follows we, let the date t decision variable be:

$$D_t = I_t/K_t.$$

Then C_t/K_t is constrained by

$$\frac{C_t}{K_t} = \alpha - D_t.$$

Thus date t consumption, C_t , can be inferred from D_t and K_t . While preferences are defined over consumption processes, we express the date t contribution to the discounted utility as a function of D_t and K_t . The date t decision D_t is constrained to be a function of \mathbf{Z}^t at each date t , reflecting the observational constraint that θ is unknown to the investor in contrast to the history \mathbf{Z}^t of the technology process.^[8] Let \mathfrak{B}_t be the collection of events (sigma algebra) generated by \mathbf{Z}^t and K_0 . We denote the larger sigma algebra that also conditions on knowledge of the parameter to be \mathfrak{B}_t^+ .

Absent robustness concerns the preferences are given by

$$(1 - \beta)E \left[\sum_{t=0}^{\infty} \beta^t v(D_t, K_t) \mid \mathfrak{B}_0^+ \right].$$

We include divergences, one for the prior on the set of possible parameter values, Θ , and the other for the potential misspecification in the dynamics as reflected in the shock distributions. The latter divergence will necessarily be dynamic in nature. We introduce a positive martingale $\{M_t : t \geq 0\}$ for an altered probability. Let $M_0 = 1$ and let M_t be measurable with respect to \mathfrak{B}_t^+ . We use the random variable M_t to alter date t probabilities conditioned on \mathfrak{B}_0^+ when evaluating the discounted expected utility:

$$(1 - \beta)E \left[\sum_{t=0}^{\infty} \beta^t M_t v(D_t, K_t) \mid \mathfrak{B}_0^+ \right].$$

Construct an intertemporal divergence:

$$(1 - \beta)E \left(\sum_{t=0}^{\infty} \beta^t M_{t+1} \log M_{t+1} \mid \mathfrak{B}_0^+ \right)$$

as a measure of divergence and scale this by a penalty parameter, ξ_2 . We purposefully discount the relative entropies in the same way as we discount the utility function, and the computations are conditioned on θ . We then use a second divergence to investigate prior sensitivity over the space of potential parameter vectors θ with a penalty parameter ξ_1 .

In this model, the investor or planner will use available data to learn about θ at each point in time. By contrast, the potential model misspecifications are presumed not to be learnable because the formulation allows for potential departures at future dates to be different from those that might have been present in the past. This model formulation allows for a lack of confidence in both the prior and intertemporal state dynamics.

10.11. A recursive counterpart to a dynamic robust decision problem

10.11.1. Model misspecification concerns

To construct a recursive formulation of preferences, we first deduce a recursive representation of the model misspecification divergence measure by exploiting the multiplicative martingale construction. Let $\{\mathfrak{B}_t : t \geq 0\}$ capture the decision maker information available at the respective dates. In our focus on misspecification concerns, we condition on the parameter θ . Write:

$$\log M_{t+1} = \sum_{j=0}^t \log M_{j+1} - \log M_j$$

and M_0 is initialized to be one. By the Law of Iterated Expectations:

$$\begin{aligned} & E(M_{t+1} \log M_{t+1} \mid \mathfrak{B}_0, \theta) \\ &= E \left[M_{t+1} \left(\sum_{j=0}^t \log M_{j+1} - \log M_j \right) \mid \mathfrak{B}_0^+ \right] \\ &= E \left[\sum_{j=0}^t M_{j+1} (\log M_{j+1} - \log M_j) \mid \mathfrak{B}_0^+ \right] \\ &= E \left[\sum_{j=0}^t M_j \left(\frac{M_{j+1}}{M_j} \right) (\log M_{j+1} - \log M_j) \mid \mathfrak{B}_0^+ \right] \\ &= E \left(\sum_{j=0}^t M_j E \left[\left(\frac{M_{j+1}}{M_j} \right) (\log M_{j+1} - \log M_j) \mid \mathfrak{B}_j^+ \right] \mid \mathfrak{B}_0^+ \right). \end{aligned} \tag{10.20}$$

Using this calculation and applying “summation-by-parts” (implemented by changing the order of summation) gives:

$$\begin{aligned}
& (1 - \beta) E \left(\sum_{t=0}^{\infty} \beta^t M_{t+1} \log M_{t+1} \mid \mathfrak{B}_0^+ \right) \\
&= (1 - \beta) \sum_{t=0}^{\infty} \beta^t E \left(\sum_{j=0}^t M_j E \left[\left(\frac{M_{j+1}}{M_j} \right) (\log M_{j+1} - \log M_j) \mid \mathfrak{B}_j^+ \right] \mid \mathfrak{B}_0^+ \right) \quad (10.21) \\
&= \sum_{t=0}^{\infty} \beta^t E \left(M_t E \left[\left(\frac{M_{t+1}}{M_t} \right) (\log M_{t+1} - \log M_t) \mid \mathfrak{B}_t^+ \right] \mid \mathfrak{B}_0^+ \right)
\end{aligned}$$

In this formula,

$$E \left[\left(\frac{M_{t+1}}{M_t} \right) (\log M_{t+1} - \log M_t) \mid \mathfrak{B}_t^+ \right]$$

is the relative entropy pertinent to the transition probabilities between date t and $t + 1$. Abstracting from parameter uncertainty, given this recursive structure, we may evaluate the objective and solve the decision problem recursively. The recursive and commitment solutions will coincide which opens the door to the use of dynamic programming methods.

Using representation [\(10.21\)](#) as input, we construct a recursive representation for preferences over decision processes. Specifically, an implied recursion for the robust-adjusted continue values

$$\begin{aligned}
V_t(k) = & \min_{\substack{M_{t+1} \geq 0, \\ E(M_{t+1}/M_t \mid \mathfrak{B}_t^+) = 1}} (1 - \beta)v(D_t, k) \\
& + \beta E \left[\left(\frac{M_{t+1}}{M_t} \right) [\xi_2 (\log M_{t+1} - \log M_t) + V_{t+1}[k\psi(D_t, Z_{t+1})]] \mid \mathfrak{B}_t^+ \right]
\end{aligned}$$

where $V_t(k)$ is the current-period continuation value as a function of the possible values for the date t capital stock. The ratio M_{t+1}/M_t is used when computing the conditional expectation of next period's continuation value, V_{t+1} . In forming this recursion, we substituted the right-hand side of capital evolution equation [\(10.19\)](#) for $K_t = k$ into the next period continuation value. Imitating our previous computations, we substitute a solution of the minimization contribution to obtain:

$$V_t(k) = (1 - \beta)v(D_t, k) + \beta \xi_2 \log E \left[\exp \left(-\frac{1}{\xi_2} V_{t+1}[k\psi(D_t, Z_{t+1})] \right) \mid \mathfrak{B}_t^+ \right]. \quad (10.22)$$

Notice that this recursion is the essentially the same as that we introduced in [Chapter 4](#). (See [\(4.8\)](#).) The rationale provided here is the decision-maker concern about model misspecification instead of risk aversion.

With preference recursion, (10.22) (10.22) we can apply dynamic programming methods to compute a robustly-optimal control by solving:

the recursion:

$$V_t^*(k) = \max_{D_t} (1 - \beta)v(D_t, K_t) + \beta\xi_2 \log E \left[\exp \left(-\frac{1}{\xi_2} V_{t+1}^*[k\psi(D_t, Z_{t+1})] \right) \mid \mathfrak{B}_t^+ \right]. \quad (10.23)$$

In this formulation, we have assumed knowledge of the parameter θ . While the maximizing D_t will be \mathfrak{B}_t^+ measurable, it will typically not be \mathfrak{B}_t measurable.

Remark 10.8

Our discounting of relative entropy has important consequences for exploration of potential misspecification. From (10.20), it follows that the sequence

$$\{E(M_t \log M_t \mid \mathfrak{B}_0^+) : t \geq 0\} \quad (10.24)$$

is increasing in t . Observe that

$$\lim_{\beta \uparrow 1} (1 - \beta)E \left(\sum_{t=0}^{\infty} \beta^t M_{t+1} \log M_{t+1} \mid \mathfrak{B}_0^+ \right) \quad (10.25)$$

gives an upper bound on

$$E(M_{s+1} \log M_{s+1} \mid \mathfrak{B}_0^+)$$

for each $s \geq 0$. This follows from monotonicity since

$$\begin{aligned} (1 - \beta)E \left(\sum_{t=0}^{\infty} \beta^t M_{t+1} \log M_{t+1} \mid \mathfrak{B}_0^+ \right) &\geq (1 - \beta) \sum_{t=s}^{\infty} \beta^t E(M_{t+1} \log M_{t+1} \mid \mathfrak{B}_0^+) \\ &= \beta^s E(M_{s+1} \log M_{s+1} \mid \mathfrak{B}_0^+) \end{aligned}$$

Taking limits as $\beta \uparrow 1$ gives the bound of interest.

If the discounted limit in (10.25) is finite, then the increasing sequence (10.24) has a finite limit. It follows from a version of the Martingale Convergence Theorem (see [Barron, 1990]) that there is a limiting random nonnegative variable M_∞ such that $E(M_\infty \mid \mathfrak{B}_0^+) = 1$ and

$$M_t = E(M_\infty \mid \mathfrak{B}_t^+).$$

When the limiting random variable, M_∞ , is well defined, almost sure limits under the probability specification must agree with those under the change implied probability measure induced by M_∞ . If the limiting random variable is strictly positive, then the baseline and the M_∞ imply the same Law of Large Numbers. In this sense, only transient departures from the baseline probability are part of the misspecification exploration. By including discounting, we expand the family of alternative probabilities in a substantively important way.

As an alternative calculation, consider a different discount factor scaling:

$$(1 - \beta) \sum_{t=0}^{\infty} \beta^t E \left(M_t E \left[\left(\frac{M_{t+1}}{M_t} \right) (\log M_{t+1} - \log M_t) \mid \mathfrak{B}_t^+ \right] \mid \mathfrak{A}_0 \right)$$

The limiting version of this measure allows for a substantially larger set of alternative probabilities and results in a limiting characterization that is used in Large Deviation Theory as applied in dynamic settings.

Remark 10.9

To explore potential misspecification [[Eckstein, 2019](#)] and [[Chen et al., 2020](#)] suggest other divergences with convenient recursive structures. A discounted version of their proposal is

$$(1 - \beta) E \left[\sum_{j=0}^t \beta^j E \left[M_j \phi_2 (M_{j+1}/M_j) \mid \mathfrak{B}_j^+ \right] \mid \mathfrak{B}_0^+ \right]$$

for a convex function ϕ_2 where this function equals one when evaluated at one, and its second derivative at one is normalized to be one.

10.11.2. A recursive formulation with parameter ambiguity

Once we include parameter uncertainty, we lose the recursive structure in the commitment version of the decision problem. We now suggest a recursive specification of preferences that includes parameter uncertainty, but with distinct preferences from those used for the commitment problem.

Let $N_t \geq 0$ be a scalar random variable that is \mathfrak{B}_t^+ measurable and satisfies:

$$E(N_t \mid \mathfrak{B}_t) = 1.$$

Recall that we constructed \mathfrak{B}_t^+ to include knowledge of the underlying parameter vector in contrast to \mathfrak{B}_t . A “behind the scenes” recursive learning input will be needed to implement computations based on the smaller sigma algebra \mathfrak{B}_t in what follows. We use N_t to alter baseline date t expectations by altering the posterior probabilities over the possible unknown parameters, but those posteriors typically require recursive computations of the type we illustrated in [Chapter 5](#).

We now incorporate parameter learning into recursion [\(10.22\)](#) for ranking alternative decision processes

$$\begin{aligned} \bar{V}_t(k) = & (1 - \beta)v(D_t, k) \\ & + \min_{N_t \geq 0, E(N_t | \mathfrak{B}_t) = 1} \beta \xi_2 E \left(N_t \log E \left[\exp \left(-\frac{1}{\xi_2} \bar{V}_{t+1}[k\psi(D_t, Z_{t+1})] \right) \mid \mathfrak{B}_t^+ \right] \mid \mathfrak{B}_t \right) \\ & + \beta \xi_1 E[\phi_1(N_t) \mid \mathfrak{B}_t]. \end{aligned}$$

Analogously, we modify recursion [\(10.23\)](#) used for finding the robustly optimal decision processes:

$$\begin{aligned} \bar{V}_t^*(k) = & \max_{D_t} (1 - \beta)v(D_t, k) \\ & + \min_{N_t \geq 0, E(N_t | \mathfrak{B}_t) = 1} \beta \xi_2 E \left(N_t \log E \left[\exp \left(-\frac{1}{\xi_2} \bar{V}_{t+1}^*[k\psi(D_t, Z_{t+1})] \right) \mid \mathfrak{B}_t^+ \right] \mid \mathfrak{B}_t \right) \\ & + \beta \xi_1 E[\phi_1(N_t) \mid \mathfrak{B}_t] \end{aligned}$$

The solution for D_t will be \mathfrak{B}_t measurable, which excludes knowledge of the parameter vector. This formulation gives a dynamic extension of Game 11.6.

Remark 10.10

Consider the special case in which $\xi_1 = \xi_2$ and $\phi_1(n) = n \log n$. In this we are essentially using a single divergence over a joint density. We may imitate previous logic and apply a *single* robust adjustment expressed in terms the the transition probabilities from \mathfrak{B}_t after \mathfrak{B}_{t+1} in place of recursions use in [\(10.22\)](#) and [\(10.23\)](#). The exploration supporting robustness in both transition dynamics and the date t distribution over the parameter, θ , led to use:

$$N_t \left(\frac{M_{t+1}}{M_t} \right)$$

to induce a change in probabilities conditioned on \mathfrak{B}_t . As an alternative to using this factorization, we now explore the ramifications of a different one by writing:

$$N_t \left(\frac{M_{t+1}}{M_t} \right) = \left(\frac{N_t \left(\frac{M_{t+1}}{M_t} \right)}{E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right]} \right) E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right].$$

Observe the composite divergence used in defining preferences can be expressed equivalently with the two different factorizations:

$$\begin{aligned} & E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) (\log N_t + \log M_{t+1} - \log M_t) \mid \mathfrak{B}_t \right] \\ &= E (N_t \log N_t \mid \mathfrak{B}_t) + E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) (\log M_{t+1} - \log M_t) \mid \mathfrak{B}_t \right] \\ &= E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \left(\log N_t + \log M_{t+1} - \log M_t - \log E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right] \right) \mid \mathfrak{B}_t \right] \\ &\quad + E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \log E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right] \mid \mathfrak{B}_t \right] \\ &= E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \left(\log N_t + \log M_{t+1} - \log M_t - \log E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right] \right) \mid \mathfrak{B}_t \right] \\ &\quad + E \left(E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right] \log E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right] \mid \mathfrak{B}_t \right) \end{aligned}$$

The right-hand side of the first equality is what we have used in constructing penalties for model misspecification concerns and ambiguity over alternative models. The right-hand-side of the third equality leads to a simplification in the minimization problem. The continuation-value functions at date $t + 1$, $V_{t+1}(k)$ and $\bar{V}_{t+1}(k)$ are measurable with respect \mathfrak{B}_{t+1} and do not depend on the unknown θ . Thus the term:

$$\frac{N_t \left(\frac{M_{t+1}}{M_t} \right)}{E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right]}$$

will be set equal to one via minimization. In other words,

$$N_t \left(\frac{M_{t+1}}{M_t} \right) = E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right].$$

The relevant choice variable to use in minimization is the single contribution:

$$E \left[N_t \left(\frac{M_{t+1}}{M_t} \right) \mid \mathfrak{B}_{t+1} \right]$$

which, by construction, is \mathcal{B}_{t+1} measurable and has an expectation conditioned on \mathcal{B}_t that is equal to one. This random variable is used to alter the probabilities of \mathcal{B}_{t+1} events conditioned on \mathcal{B}_t . In summary, for the robustness analysis over state dynamics and distributions over parameter, it suffices to focus on the minimization implications of such random variables subject to the corresponding relative entropy penalty. To implement this reduction in practice requires as input the recursive solutions to parameter learning under a baseline specification.

We do not take this argument as a blanket endorsement for the resulting simplification. Quite the contrary, we find it advantageous to explore the distinct roles of model misspecification and prior/posterior misspecification.

10.12. Implications for uncertainty quantification

Uncertainty quantification is a challenge that pervades many scientific disciplines. The methods we describe here open the door to answering the “so what” aspect of uncertainty measurement. So far, we have deliberately explored examples that are low-dimensional to illustrate results. While these are pedagogically revealing, the methods we describe have all the more potency in problems with high-dimensional uncertainty. By including minimization as part of the decision problem, we isolate the uncertainties that are of most relevance to the decision or policy problem. This may pave the way for incorporating sharper prior inputs or for guiding future efforts aimed at providing additional evidence relevant to the decision-making challenge. Furthermore, there may be multiple channels by which uncertainty can impact the decision problem. As an example, consider an economic analysis of climate change. There is uncertainty in i) the global warming implications of increases in carbon emissions, ii) the impact of global warming on economic opportunities, and iii) the prospects for the discovery of new, clean technologies that are economically viable. A direct extension of the methods developed in this chapter provides a (non-additive) decomposition of the channels of uncertainty. By modifying the penalization, uncertainty in each channel could be activated separately in comparison to activating uncertainty in all channels simultaneously. Comparing outcomes of such computations reveals which channel of uncertainty is most consequential to the structuring of a prudent decision rule.^[9]

↓ [Homework For Econ 312](#)

- [1] This chapter draws heavily on a paper with the same title that was recently published in the *Journal of Applied Econometrics* by Hansen and Sargent entitled [Risk, ambiguity and misspecification](#).
- [2] 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.

- [3] We put “behavioral” in quotes to emphasize that most economic models are about agents’ behaviors, including models that impose the rational expectations and common knowledge assumptions that “behavioral” economists want to drop. “Behavioral” economics sometimes means work that is linked more or less informally to psychology.
- [4] 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.
- [5] Statisticians often use a loss function instead of a utility function. One can think of negative of a loss function as our utility function.
- [6] Chamberlain uses a constraint on the family of prior distribution, consistent with the axiomatic formulation of [[Gilboa and Schmeidler, 1989](#)].
- [7] While this contribution and some of the papers they cite, assume an endowment economy, their insights extend directly to a model with production.
- [8] Note that the endogenous state variable, K_t , reveals no new information in addition to current and past values of the technology process. This means that there is no incentive for the investor to experiment in this setting.
- [9] [[Cappelli et al., 2021](#)] develop a related proposal with a similar ambition.