Comparative Ambiguity Aversion for Smooth Utility Functions

Chiaki Hara*

April 15, 2021

Abstract

In this paper, we define a measure of ambiguity aversion for a twice continuously differentiable utility function that represents an uncertain-averse preference relation over the set of acts on monetary consequences. The measure is determined by the Hessian of the utility function and the subjective probability implicitly defined by it, and allows us to compare ambiguity aversion represented by two utility functions even when they have different risk attitudes and different functional forms that have been axiomatized in the literature. Our measure of ambiguity aversion is characterized for some commonly used ambiguity-averse utility functions, such as smooth ambiguity models, variational utilities, and confidence functions. Implications on the numerical analysis on experimental findings, portfolio choice, and asset pricing are also explored.

JEL Classification Codes: C38, D81, G11.

Keywords: Expected utility functions, uncertainty aversion, risk aversion, ambiguity aversion.

1 Introduction

In this paper, we introduce a measure of ambiguity aversion or, equivalently, a new "more-ambiguity-averse-than" relation over twice differentiable uncertainty-averse utility functions (or, equivalently, preference relations). The purpose of doing so is two-fold. First, we wish to compare ambiguity aversion between two utility functions having different functional forms. Second, and more importantly, we wish to compare ambiguity aversion between two utility functions having different risk attitudes.

*This research is funded by the Open Research Area (ORA) for the Social Sciences "Ambiguity in Dynamic Environments." I am grateful to Yusuke Osaki and Koji Shirai for their insightful comments.

It is important, especially when applying uncertainty-averse utility functions to a quantitative analysis, to introduce a measure of ambiguity aversion on a broad class of uncertainty-averse utility functions that does not presume or imply a common functional form or a common risk attitude for two preference relations under study. The reason is can be best seen by following up the development of macroeconomic models after Mehra and Prescott (1985). Since they showed that the traditional representative-agent model needs an unreasonably high coefficient (much higher than is experimentally or introspectively inferred) of relative risk aversion to explain the historical average of equity premiums, the subsequent studies has introduced market incompleteness, transaction costs, or model uncertainty (or, synonymously, ambiguity) to explain the equity premiums with a reduced coefficient of relative risk aversion. Moreover, the goodness of a model has often been judged based on to what extent the coefficient can be reduced while still explaining the historical average of the equity premium,[1] and the race for a good model in this sense has shaped our understanding of macroeconomics.

There is now a growing literature in macroeconomics and finance in which the representative agent is assumed to be ambiguity-averse and attempts are made to explain an anomaly, such as the high equity premium and the nonzero alphas in stock returns. Some of them do succeed in doing so, but they vary in the specification of the form of utility functions or the source in which ambiguity is perceived. Moreover, ever since Ellsberg's (1961) seminal contribution, the analysis is often focused on whether the introduction of ambiguity aversion can explain an otherwise unexplainable phenomenon (such as the subjects' tendency to prefer a bet on an urn of balls whose color composition is known to a bet on an urn of balls whose color composition is unknown), paying little attention to how large or strong the tendency of ambiguity aversion needs to be to do. Without a measure of ambiguity aversion that is applicable to a broad class of uncertainty-averse utility functions, we cannot tell whether one model is superior to another in terms of the needed size of ambiguity aversion. In other words, the course of development in macroeconomics with reference to risk aversion cannot be achieved with reference to ambiguity aversion when a widely applicable measure of ambiguity aversion is unavailable.

The purpose of this paper is precisely to introduce a measure of ambiguity aversion that is applicable to a broad class of uncertain-averse utility functions. With such a measure, we can say that a utility function is more ambiguity-averse than another even when they have different functional forms, and, thus, judge which specification would be superior to explain a given phenomenon and whether the required size of ambiguity

[1]As quoted by Kocherlakota (1996), Lucas (1994, Section 4.1) claimed, in incomplete asset markets, that any attempt to solve the equity premium puzzle that requires a constant coefficient of relative risk aversion greater than 2.5 is unlikely to be widely viewed as a resolution that does not depend on a high coefficient.

aversion is reasonable or plausible.

As we will assume that the state space is finite (S in number), the consequences are monetary amounts (which fall into the setting of Savage, while they would be lotteries in the setting of Anscombe and Aumann), and the utility functions are twice continuously differentiable, the measure can be defined within the framework of classical demand theory. Specifically, the measure of ambiguity aversion is taken as the ratio, minus one, of the quadratic form determined by the Hessian of an uncertain-averse utility function at a constant act to that of the ambiguity-neutral utility function having the same risk attitudes as the ambiguity-averse utility function. Geometrically, at each point on the diagonal of $\boldsymbol{R}^S$, the indifference curve of the ambiguity-averse utility function lies above the indifference curve of the ambiguity-neutral utility function; equivalently, the former is tangent to the latter at the point on the diagonal but the former is more convex than the latter. The difference between the two are of the second-order at the point on the diagonal, and the ratio of the second-order difference to the curvature of the former is taken as the measure of ambiguity aversion. The measure is zero if the uncertainty-averse utility function is, in fact, ambiguity-neutral, and the measure is larger the larger the difference between the two indifference curves. It is a measure of ambiguity aversion in the small in the sense of Pratt (1964, Section 3), because it deals with deviations from a constant act having zero mean and small variances (under the probability implicit at the constant act). It is different from his measure of risk aversion in that we, first, specify a direction of deviations from the constant act and, then, quantify the fraction of the curvature of the indifference curve that can be attributed to ambiguity aversion.

While the measure of ambiguity aversion is defined in terms of the Hessian of an uncertain-averse utility function, it can be interpreted in terms of the ambiguity premium and also of asset returns. We will, in fact, show that the measure is approximately equal to the ratio of the ambiguity premium to the risk premium. When the decision maker is the representative consumer whose utility gradient at the aggregate consumption coincides with the state prices, the ratio of the part of the equity premium attributable to ambiguity aversion to that attributable to risk aversion will turn out to be approximately equal to the measure of ambiguity aversion. While defined solely in terms of utility functions, our measure, therefore, has a sound decision-theoretic foundation as well as a useful asset-pricing implication.

As we impose the twice differentiability on uncertainty-averse utility functions, we accommodate the smooth ambiguity model of Klibanoff, Marinacci, and Mukerji (2005) and some classes of variational preferences of Maccheroni, Marinacci, and Rustichini (2006), including relative entropy of Hansen and Sargent (2001), and of confidence functions of Chateauneuf and Faro (2009), but we exclude some important ambiguity-averse utility functions, such as maximin expected utilities of Gilboa and Schmeidler (1989) and Choquet expected utilities of Schmeidler (1989). We extend our measure of ambiguity aversion to such non-differentiable utility functions using second directional one-sided derivatives instead of the Hessian matrices. This extension complements the comparison of behavioral implications of ambiguity aversion between two decision makers when they have the same first-order evaluation of small deviations from a constant but different preferences due to the difference in second-order evaluations.

The notion of comparing ambiguity aversion has been discussed in the decision-theoretic literature. The study is motivated by the desire to compare ambiguity aversion between two decision makers in much the same way as Arrow-Pratt measure of risk aversion allows us to do so for comparative risk aversion. Mentioning the difficulty of disentangling the effect of ambiguity aversion from that of risk aversion in observed choices, Ghirardato and Marinacci (2002, Definition 7 and Section 6.1) defined, in a general framework (encompassing the Anscombe-Aumann setting), one decision maker as more ambiguity-averse than another if they share the same risk attitude and the former prefers an unambiguous act to an ambiguous one whenever the latter does. Klibanoff, Marinacci, and Mukerji (2005, Definition 5 and Theorem 2) introduced the smooth ambiguity model, which we elaborate on in Section 8.1, and defined one decision maker as more ambiguity-averse than another in a way that implies the identical risk attitude (which is assumed by Ghirardato and Marinacci (2002, Definition 7)) and showed, in their smooth model, that one decision maker is more ambiguity-averse than another if and only if they share the same risk attitude and the function, often denoted by ϕ_i, that transforms their utility function for unambiguous acts (risks) to the utility function for second-order acts (ambiguities) is more concave for the former than for the latter. They also introduced a notion of constant ambiguity aversion with respect to variations of expected utilities under different probabilities (first-order beliefs), and showed that it is equivalent to the function ϕ_i exhibiting constant absolute risk aversion. Guetlein (2016, Definitions 3 and 5) introduced, based on what she termed subjectively risk-free acts, a notion of an at-least-as-risk-averse-as relation (a more-risk-averse-than relation) that is well suited to the smooth ambiguity model as it retain the same ambiguity attitudes. She showed (Proposition 1 of her paper) that if a decision maker is more risk-averse than another in her sense, then the ϕ_i functions are different between the two, that is, the ambiguity attitudes in the sense of KMM are different between the two. Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (forthcoming, Definition 2 and Proposition 4) introduced, based on the more-ambiguity-averse-than relation of Ghirardato and Marinacci (2002), notions of constant, increasing, and decreasing ambiguity aversion with respect to a shift, common across all states, in consumption levels, and showed that any one of these notions implies that the decision maker exhibits constant absolute risk aversion. This strand of literature points to the limited applicability of the notion of the more-ambiguity-averse-than relation as it significantly restricts the risk attitudes under consideration.

On the other hand, Wang (2019) introduced a notion of the more-ambiguity-averse-than relation that allows the risk attitudes under study to be different, by inferring ambiguity attitudes, without assuming identical risk attitudes, from the decision maker's preferences between unambiguous acts and ambiguous binary acts that may lead only to the best and worst consequences. His notion is general enough to cover maximin and Choquet expected utilities (for which the impact of ambiguity aversion on choices is of first order) as well as smooth ambiguity models and some classes of variational preferences and confidence functions (for which the impact of ambiguity aversion on choices is of second order). Even when there is no best or worst outcome (as in the case where the feasible consumption levels constitute an open interval of $\boldsymbol{R}$), the more-ambiguity-averse-than relation can be defined but it depends on the choice of two consequences that define the binary acts. In a special class of the smooth ambiguity model, Hara and Honda (forthcoming) considered the optimal investment allocation over several stocks and a risk-free bond, and showed that while the total investment into stocks depend on the risk attitudes, the proportions of the investment allocated to each stock within the total amount invested into the stocks is determined solely by the ambiguity attitudes. Hara (2020) introduced a more-ambiguity-averse-than relation within the smooth ambiguity model using the function, different from the function ϕ_i, that transforms the marginal utility functions for unambiguous acts to the marginal utility functions for second-order acts. These papers can be understood as attempts to lay foundations to and derive implications from comparative ambiguity aversion without assuming identical risk attitudes.

Recent experimental studies on ambiguity aversion include Halevy (2007), Bossaerts, Ghirardato, Guarnaschelli, and Zame (2010), Ahn, Choi, Gale, and Kariv (2014), and Attanasi, Gollier, Montesano, and Pace (2014). While their prime interest is on the identification of phenomena whose explanation requires ambiguity aversion (such as preference for portfolios with unambiguous returns to those with ambiguous ones), they also looked into the forms of ambiguity-averse utility functions and the magnitude of ambiguity aversion that are need to explain their experimental findings. Empirical studies include Ju and Miao (2012), Chen, Ju, and Miao (2014), Jahan-Parvar and Liu (2014), Gallant, Jahan-Parvar, and Liu (2019), Thimme and Völkert (2015), Backus, Ferriere, and Zin (2015), Collard, Mukerji, Sheppard, and Tallon (2018), and Altug, Collard, Çakmakli, Mukerji, and Özsöylev (2018). The types of the exercises they conducted are estimation or calibration of ambiguity aversion, along with risk aversion, intertemporal elasticity of substitution and subjective time discount rates, for specific forms of utility functions and consumption processes, to explain the historically observed equity premiums, optimal portfolios, and their dynamic behaviors.

The rest of this paper is organized as follows. Section 2 give the setup of the paper. Section 3 defines a measure of ambiguity aversion. Section 4 relates our measure of ambiguity aversion to the ambiguity premium. Section 5 relates our measure of ambiguity aversion to asset returns. Section 6 extends our measure of ambiguity aversion to non-differentiable utility functions. Section 7 gives a numerical exercise on our measure based on the results in Sections 4 and 5. Section 8 shows how the measure of ambiguity aversion can be rewritten in some special classes of the smooth ambiguity model, variational utilities, and confidence functions. Section 9 gives a conclusion and suggest directions of future research. The appendix gathers preliminary results and proofs.

2 Setup

The state space is $\{1, 2, \ldots, S\}$, where $S \geq 2$. As it is finite, the σ-field is the power set and every random variable can be regarded as a vector in $\boldsymbol{R}^S$. Denote by e the vector in $\boldsymbol{R}^S$ whose coordinates are all equal to one. It represents a profile of state-independent consumption or utility levels.

Let T be a nonempty open interval of $\boldsymbol{R}$. It is the set of all possible consumption levels. Let $v : T \to \boldsymbol{R}$. It is referred to the Bernoulli utility function by Mas-Colell, Whinston, and Green (1995), and interpreted as representing attitudes towards pure risk (uncertainty involving no ambiguity). Let $I : v(T)^S \to \boldsymbol{R}$. For each $f : S \to T$, or $f \in T^S$, denote by $v \circ f$ the vector in $v(T)^S$ whose s-th coordinate is equal to $v(f(s))$. Define $V : T^S \to \boldsymbol{R}$ by letting $V(f) = I(v \circ f)$ for every $f \in T^S$. As it is quasi-concave in all applications we consider, following Cerreira-Vioglio, Maccheroni, Marinacci, and Montrucchio (2011, CVMMM thereafter), we refer to it as an uncertainty-averse utility function (over the set T^S of all Savage acts on the state space S). When an act is a scalar multiple of e, we call it as a constant act.

In the literature, the domain of the uncertainty-averse utility function has often been taken as the set of all Anscombe-Aumann acts, that is, the set of all mappings from S to the probabilities on T, and, under the expected utility hypothesis regarding these probabilities, define the uncertainty-averse utility function V so that $V(\pi) = I(u \circ \pi)$ where $u(\pi(s)) = \int_T v(x)\, \mathrm{d}\pi(s)(x)$ and $\pi : s \mapsto \pi(s)$ is a mapping of S into the set of all probabilities on T. While the (larger) domain of Anscombe-Aumann acts is necessary to axiomatize various functional forms of the uncertainty-averse utility functions, we opt for the (smaller) domain of Savage acts, as the latter is sufficient to introduce our measure of ambiguity aversion.

It is theoretically more sound, in this setup, to use a preference (binary) relation on T^S as the primitive concept. But we opt for using the utility function V because that simplifies the subsequent analysis and a sufficient set of conditions on the preference relation to be representable I was given by CVMMM (2011) when the domain of V is extended to the set of all Anscombe-Aumann acts. The use of the Anscombe-Aumann acts enabled them to derive the decision maker's risk attitudes, which is represented by v in our paper, and state axioms that characterize forms of utility functions under study. In our analysis, we will use a fictitious ambiguity-neutral utility function $\bar{V}$ derived from V, and it will turn out that conditions in terms of Anscombe-Aumann acts (especially lottery acts) can be equivalently stated in terms of $\bar{V}$. This is another reason why we opt for using V.

We assume, throughout this paper, that v is twice continuously differentiable and satisfies $v'' < 0 < v'$. Then $v(T)$ is also a nonempty open interval of $\boldsymbol{R}$.

Following CVMMM (2011) and others in the literature, we assume throughout this paper that I is normalized, that is, $I(ye) = y$ for every $y \in v(T)$. It guarantees the uniqueness of the Hessian of I in a sense that will be made precise in the next section.

We impose the following differential conditions on I.

Assumption 1 Twice continuously differentiable I is twice continuously differentiable.

Differentiably strictly increasing $\nabla I(g) \in \boldsymbol{R}^S_{++}$ for every $g \in v(T)^S$.

The twice continuous differentiability excludes many well-known uncertain-averse utility functions of interest, most notably the maximin utility functions of Gilboa and Schmeidler (1985), but accommodates some other ones of interest, such as the smooth ambiguity-averse utility functions of Klibanoff, Marinacci, and Mukerji (2005), some of variational preferences of Maccheroni, Marinacci, and Rustichini (2006), such as the relative entropy used by Hansen and Sargent (2001), and some of those based on confidence functions of Chateauneuf and Faro (2009). While the twice continuous differentiability is imposed on the entire domain $v(T)^S$, for our analysis, it is sufficient to impose it on any open subset of $v(T)^S$ that contains all constant acts. The differentiable strict increasingness implies that the utility function is strictly increasing in consumption levels at every state.

The final assumption is written in a differential form and also has a significant behavioral restriction.

Assumption 2 Common support along the diagonal There is a $p \in \boldsymbol{R}^S_{++}$ such that $p \cdot e = 1$ and $p^\top = \nabla I(ye)$ for every $y \in v(T)$.

The constant supportability along the diagonal means that there is a unique subjective probability that the decision maker holds, up to the first-order approximation, whenever his act is constant (state-independent). When I is differentiable, the condition is equivalent to Axiom 7, Translation Invariance at Certainty, of Rigotti, Shannon, and Strzalecki (2008), as their Proposition 8 showed.

3 Measure of ambiguity aversion

In Section 2, we used a pair of a Bernoulli utility function $v : T \to \boldsymbol{R}$ and a function $I : v(T)^S \to \boldsymbol{R}$ to define an uncertainty-averse utility function $V : T^S \to \boldsymbol{R}$ on the set T^S of all acts. The function I also defines a unique probability p^I on the state space S that decision maker has whenever his act is independent of states. To compare with V, we define the utility function $\bar{V}$ by the pair $(v, \bar{I})$, where $\bar{I} : v(T)^S \to \boldsymbol{R}$ is defined by $\bar{I}(y) = p^I \cdot y$ for every $y \in v(T)^S$. Then $\bar{V}$ is the expected utility function under the subjective probability p^I. It depends on the Bernoulli utility function v that represents the decision maker's risk attitudes and his subjective probability at the constant acts, but nothing else. On some occasions, we consider the ambiguity-averse decision maker's subjective probability implicit in his choice of an act. Note that for each $f \in T^S$,

$$\frac{1}{\nabla I(f)e} \nabla I(f). \tag{1}$$

can be considered as a probability on the state space S as its coordinates are positive and sum up to one. We call it the decision maker's implicit subjective probability at an act f, because it is proportional to the vector of weights given by V to an increase in the (Bernoulli) utility level at each state. Ghirardato and Siniscalchi (2012, Theorems 6 and 7) showed, in a much more general framework, that it measures the decision maker's local subjective probabilities around f.

For each $x \in T$, we define $H^x : \boldsymbol{R}^S \setminus \{0\} \to \boldsymbol{R}$ by letting

$$H^x(z) = \frac{z^\top \left(\nabla^2 V(xe) - \nabla^2 \bar{V}(xe)\right) z}{z^\top \nabla^2 \bar{V}(xe) z} \tag{2}$$

for every $z \in \boldsymbol{R}^S \setminus \{0\}$. That is, H^x can be obtained by dividing the difference between two quadratic forms $\nabla^2 V(xe)$ and $\nabla^2 \bar{V}(xe)$ by the latter. This is the measure of ambiguity aversion that we propose in this paper. Roughly, it is equal to the proportional increases in the quadratic form from the ambiguity-neutral utility function $\bar{V}$ to the ambiguity-averse utility function V. It is homogeneous of degree zero in z.

To obtain other useful expressions of the measure $H^x(z)$, for each $w \in \boldsymbol{R}^S$, denote by

$[w]$ the $S \times S$ matrix of which the s-th diagonal entry is equal to $w(s)$ and the off-diagonal entries are all equal to zero. It is proved in Appendix A that

$$\nabla V(xe) = \nabla \bar{V}(xe) = v'(x)(p^I)^\top, \tag{3}$$

$$\nabla^2 \bar{V}(xe) = v''(x)[p^I], \tag{4}$$

$$\nabla^2 V(xe) - \nabla^2 \bar{V}(xe) = (v'(x))^2 \nabla^2 I(v(x)e). \tag{5}$$

In words, V and $\bar{V}$ have the same gradient at every state-independent act but their Hessian matrices differ by the Hessian of I multiplied by the squared marginal utility, because the Hessian of $\bar{I}$ is zero. By plugging (4), and (5) into (2), we obtain

$$H^x(z) = \frac{-\dfrac{z^\top \left(v'(x)\nabla^2 I(v(x)e)\right) z}{\|z\|^2_{p^I}}}{-\dfrac{v''(x)}{v'(x)}}, \tag{6}$$

where $\|z\|^2_{p^I} = \sum p^I_s z^2_s$, the squared L^2-norm of z with respect to the probability p^I. This expression clarifies what $H^x(z)$ represents. It is the quadratic form of the Hessian $\nabla^2 I(v(x)e)$ of V at each constant act restricted on the unit sphere in the L^2-norm with respect to p^I, multiplied by the marginal utility v' and divided by the Arrow-Pratt measure of absolute risk aversion of the Bernoulli utility function v at x. The denominator is, of course, invariant to the affine transformations of v. Corollary 7 in Appendix A shows that the numerator is also invariant as long as I is normalized. Thus, the value of $H^z(x)$ is uniquely determined by the risk attitudes and the preference relation on the entire set T^S of acts.

Since I is normalized, $\nabla^2 I(v(x)e)e = 0$. Thus, (6) can further be rewritten as

$$H^x(z) = \frac{-\dfrac{(z-ke)^\top \left(v'(x)\nabla^2 I(v(x)e)\right)(z-ke)}{\|z\|^2_{p^I}}}{-\dfrac{v''(x)}{v'(x)}},$$

for every $k \in \boldsymbol{R}$. In particular, if $k = (z \cdot e)S^{-1}$, then $z - ke$ coincides with the orthogonal projection of z onto the hyperplane $e^\perp$ with normal vector e, and if $k = E^{p^I}[z]$, then $z - ke$ has zero mean under p^I. To identify $H^x(z)$ for all $z \in \boldsymbol{R}^S \setminus \{0\}$, therefore, it is sufficient to find them for all z orthogonal to e or with zero mean.

Since $\|z\|_{p^I}^2 = \left(E^{p^I}[z]\right)^2 + \mathrm{Var}^{p^I}[z]$, (6) can also be rewritten as

$$H^x(z) = \frac{-\dfrac{z^\top \left(v'(x)\nabla^2 I(v(x)e)\right) z}{\left(E^{p^I}[z]\right)^2 + \mathrm{Var}^{p^I}[z]}}{-\dfrac{v''(x)}{v'(x)}}.$$

In many places of the subsequent analysis, we consider $H^x(z)$ only for a zero-mean z, that is, $E^{p^I}[z] = 0$. Then

$$H^x(z) = \frac{-\dfrac{z^\top \left(v'(x)\nabla^2 I(v(x)e)\right) z}{\mathrm{Var}^{p^I}[z]}}{-\dfrac{v''(x)}{v'(x)}}.$$

Moreover, then, $\left(E^{p^I}[z+ke]\right)^2 + \mathrm{Var}^{p^I}[z+ke] = k^2 + \mathrm{Var}^{p^I}[z]$. Thus, for every $x \in T$, the maximum $\max_{z\in \boldsymbol{R}^S\setminus\{0\}} H^x(z)$ is attained by a z with $E^{p^I}[z] = 0$.

Since the gradient $\nabla I(v \circ f)$ measures the decision maker's subjective probability implicit at the act f, the Hessian $\nabla^2 I(v \circ f)$ is the first-order approximation of the change in his subjective probabilities in response to a change in acts, which is often referred to in the literature as the probability tilting. Hence, $H^x(z)$ measures the probability tilting when small uncertainties are introduced to his state-independent act xe. As tje subjective probabilities are changed via Bernoulli utility levels, and, thus, depends not only on the size of pure ambiguity aversion (which is determined by the nonlinearity of I but not by any property of v) but also on the size of the change in Bernoulli utility levels (which is determined by the curvature of v but not by any property of I). To pick up the pure ambiguity aversion, therefore, we divide, in (6), the quadratic form defined by the Hessian of I by the Arrow-Pratt measure of absolute risk aversion of v.

This last fact enables us to compare ambiguity aversion of two decision makers even when they have different risk attitudes, as long as both satisfy the conditions of Assumptions 1 and 2. Formally, consider two utility functions V_1 and V_2 that are derived from (v_1, I_1) and (v_2, I_2). Define $H_1^{x_1}$ and $H_2^{x_2}$ for V_1 and V_2 as in (2). We can define an at-least-as-ambiguity-averse-as relation between V_1 and V_2 in the following manner.

Definition 1 Assume that v_1, I_1, V_1, v_2, I_2, V_2 are as above. Assume that $T_1 = T_2$ and $p^{I_1} = p^{I_2}$. Denote them by T and p. Let $x \in T$ and $z \in \boldsymbol{R}^S$. We say that V_1 *is at least as ambiguity-averse as* V_2 *at* x *in the direction of* z if $H_1^x(z) \geq H_2^x(z)$. We say that V_1 *is more ambiguity-averse than* V_2 *at* x *in the direction of* z if $H_1^x(z) > H_2^x(z)$.

We also say that V_1 is at most as ambiguity-averse as V_2 at x in the direction of z if V_2 is at least as ambiguity-averse as V_1 at x in the direction of z. Since $\nabla^2 I_n(v(x)e)e = 0$, $H_n^x(ye) = 0$ for all n, x and $y \in \boldsymbol{R}$, and if V_1 is more ambiguity-averse than V_2 at x in the direction of z, then z is not a scalar multiple of e.

Since the definition is concerned with the gradients and Hessians of the utility functions, but not with the other higher-order derivatives, it is possible that two utility functions are different and, yet, are equally ambiguity-averse. We will give such an example in Proposition 6.

It will be illustrative to present two polar cases of comparison between V_1 and V_2. First, if $v_1 = v_2$, then the denominator of (6) is common between V_1 and V_2. If, in addition, $p^{I_1} = p^{I_2}$, then the ranking of the measures of ambiguity aversion of V_1 and V_2 are determined solely by the Hessians of I_1 and I_2 at the state-independent acts. Second, let $x \in T$ and assume that $v_1(x) = v_2(x)$ and $v_1'(x) = v_2'(x)$. This assumption, imposed at a single consumption level x, is without loss of generality as some affine transformation of either one of the two Bernoulli utility functions can have it satisfied. We also assume that $I_1 = I_2$. Then the numerator of (6) is common between V_1 and V_2. Thus, if it is strictly positive, then $H_1^x(z) \geq H_1^x(z)$ if and only if $-v_1''(x) \leq -v_2''(x)$, that is, $-v_1''(x)/v'(x) \leq -v_2''(x)/v_2'(x)$. That is, when both V_1 and V_2 are uncertainty-averse, while sharing the same function I, V_1 at least as ambiguity-averse as V_2 if and only if v_2 is at least as risk-averse as v_1. If, on the other hand, the Hessian of the common function I is zero, then I is linear, V_1 and V_2 are ambiguity-neutral, and $H_1^x(z) = H_2^x(z) = 0$ for every z. The lesson to be drawn from this case is that the ambiguity aversion is determined not only by the Hessian of I but also the risk aversion.

4 Choice condition

Although our measure of ambiguity aversion, $H^x(z)$, is defined in terms of the Hessians of utility functions, it can also be stated in terms of the choices that the decision maker would make between uncertain and constant acts. In this section, we show that it approximates the so-called ambiguity equivalents when the uncertainties are small.

Let $x \in T$ and $z \in T^S$ and suppose that $p^I \cdot z = E^{p^I}[z] = 0$. Since T^S is open, $xe + \varepsilon z \in \boldsymbol{R}^S$ for a sufficiently small $\varepsilon > 0$. For such an ε, define $\kappa^x(\varepsilon, z)$ by

$$V\left((x - \kappa^x(\varepsilon, z))\, e\right) - V(xe + \varepsilon z) = 0. \tag{7}$$

Then $\kappa^x(\varepsilon, z)$ is the uncertainty premium of the act $xe + \varepsilon z$ with respect to the utility function V. Since V is strictly increasing and strictly concave around the set of constant acts (which is true if I is quasi-concave by (50)), $(x - \varepsilon\|z\|_\infty)e \in T^S$ for a sufficiently small $\varepsilon > 0$ and $V((x - \varepsilon\|z\|_\infty)e) \leq V(xe + \varepsilon z) \leq V(xe)$, where $\|z\|_\infty = \max_s z(s)$ and the second weak inequality holds as a strict inequality unless $z = 0$. Thus, for a sufficiently small $\varepsilon > 0$, $\kappa^x(\varepsilon, z)$ is well defined (exists and is unique), nonnegative, and strictly positive unless $\varepsilon z = 0$. Define $\bar{\kappa}^x(\varepsilon, z)$ analogously for the ambiguity-neutral utility function $\bar{V}$. As it does not involve ambiguity aversion, we will refer to it as the risk premium, and refer to the difference $\kappa^x(\varepsilon, z) - \bar{\kappa}^x(\varepsilon, z)$ as the ambiguity premium.

The following theorem relates our measure of ambiguity aversion to the ratio of the ambiguity premium to the risk premium.

Theorem 1 *For every $x \in T$ and $z \in \boldsymbol{R}^S$ with $z \neq 0$ and $p^I \cdot z = E^{p^I}[z] = 0$,*

$$\frac{\kappa^x(\varepsilon, z) - \bar{\kappa}^x(\varepsilon, z)}{\bar{\kappa}^x(\varepsilon, z)} \to H^x(z)$$

as $\varepsilon \to 0$.

That is, for a small randomness εz, our measure of ambiguity aversion, $H^x(z)$, is approximately equal to the ratio of the ambiguity premium $\kappa^x(\varepsilon, z) - \bar{\kappa}^x(\varepsilon, z)$ to the risk premium $\bar{\kappa}^x(\varepsilon, z)$.

Theorem 1 provides us with a more choice-based characterization of our measure of ambiguity aversion. To give such a characterization, we need the following definition.

Definition 2 Assume that v_1, I_1, V_1, v_2, I_2, V_2 are as above. Assume that $T_1 = T_2$ and $p^{I_1} = p^{I_2}$. Denote them by T and p. Let $x \in T$ and $z \in \boldsymbol{R}^S$, and suppose that $p \cdot z = E^p[z] = 0$. We say that V_1 *is at least as ambiguity-averse in choice as* V_2 *at* x *in the direction of* z if there is an $\bar{\varepsilon} > 0$ such that for every $\varepsilon \in (0, \bar{\varepsilon})$, there are a $\bar{k}_1 \in \boldsymbol{R}$, a $\bar{k}_2 \in \boldsymbol{R}$, $h_1 \in \boldsymbol{R}$, and $h_2 \in \boldsymbol{R}$ such that $h_1 \geq h_2$ and

$$V_1\left(\left(x - \bar{k}_1 - h_1\bar{k}_1\right)e\right) \geq V_1\left(xe + \varepsilon z\right), \tag{8}$$

$$\bar{V}_1\left(\left(x - \bar{k}_1\right)e\right) \leq \bar{V}_1\left(xe + \varepsilon z\right), \tag{9}$$

$$V_2\left(\left(x - \bar{k}_2 - h_2\bar{k}_2\right)e\right) \leq V_2\left(xe + \varepsilon z\right), \tag{10}$$

$$\bar{V}_2\left(\left(x - \bar{k}_2\right)e\right) \geq \bar{V}_2\left(xe + \varepsilon z\right). \tag{11}$$

We say that V_1 *is more ambiguity-averse in choice than* V_2 *at* x *in the direction of* z if there are an $\bar{\varepsilon} > 0$ and a $\delta > 0$ such that for every $\varepsilon \in (0, \bar{\varepsilon})$, there are a $\bar{k}_1 \in \boldsymbol{R}$, a $\bar{k}_2 \in \boldsymbol{R}$, $h_1 \in \boldsymbol{R}$, and $h_2 \in \boldsymbol{R}$ such that $h_1 \geq h_2 + \delta$ and (8), (9), (10), and (11) hold.

The first inequality means that the uncertainty premium of $xe + \varepsilon z$ for V_1 is at least $\bar{k}_1 + h_1\bar{k}_1$. The second inequality means that the risk premium of $xe + \varepsilon z$ for V_1 is at most

$\bar{k}_1$. Thus the uncertainty premium is higher than the risk premium by at least h_1 times the risk premium. The third inequality means that the uncertainty premium of $xe + \varepsilon z$ for V_2 is at most $\bar{k}_2 + h_2\bar{k}_2$. The last inequality means that the risk premium of $xe + \varepsilon z$ for V_2 is at least $\bar{k}_2$. Thus the uncertainty premium is higher than the risk premium by at most h_2 times the risk premium. Since $h_1 \geq h_2$, the four inequalities together imply that the proportional increase from the risk premium to the uncertainty premium for V_1 is at least as large as that for V_2. We say that V_1 is more ambiguity-averse than V_2 if the proportional increase for V_1 is higher than for V_2. In this case, the four weak inequalities can all be replaced by strict inequalities, by taking h_1, $\bar{k}_2$, and δ smaller and h_2 and $\bar{k}_1$ larger whenever necessary.

To offer the menus $\{\left(x - \bar{k}_1 - h_1\bar{k}_1\right) e, xe+\varepsilon\}$, $\{\left(x - \bar{k}_1\right) e, xe+\varepsilon z\}$, $\{\left(x - \bar{k}_2 - h_2\bar{k}_2\right) e, xe+\varepsilon\}$, and $\{\left(x - \bar{k}_2\right) e, xe + \varepsilon z\}$ to the decision makers, it is necessary to know their risk attitudes, represented by v_1 and v_2, because $\bar{V}_1$ and $\bar{V}_2$ are constructed from v_1 and v_2. Ghirardato and Marinacci (2002, Section 6.1) already pointed out the complexity of defining a more-ambiguity-averse-than relation with heterogeneous risk attitudes, and proposed a two-step elicitation, in which the first step is devoted to identifying risk attitudes only and the second step involves the use of ambiguous lotteries. Such a two-step elicitation would also be necessary to offer the menus in Definition 2.

The following corollary gives the relationship between the utility-based definition (Definition 1) and the choice-based definition (Definition 2).

Corollary 1 *Assume that v_1, I_1, V_1, v_2, I_2, V_2 are as above. Assume that $T_1 = T_2$ and $p^{I_1} = p^{I_2}$. Denote them by T and p. Let $x \in T$ and $z \in \boldsymbol{R}^S$, and suppose that $z \neq 0$ and $p \cdot z = E^p[z] = 0$.*

1. *V_1 is at least as ambiguity-averse in choice as V_2 at x in the direction of z only if V_1 is at least as ambiguity-averse as V_2 at x in the direction of z .*

2. *V_1 is more ambiguity-averse in choice than V_2 at x in the direction of z if and only if V_1 is more ambiguity-averse than V_2 at x in the direction of z .*

We now compare our definition of the at-least-as-ambiguity-averse-as relation in choice (Definition 2) with those used in the literature, including Ghirardato and Marinacci (2002), Klibanoff, Marinacci, and Mukerji (2005), Macherroni, Marinacci, and Rustichini (2006), Chateauneuf and Faro (2009), and CVMMM (2011). For this purpose, first, we give almost equivalent conditions of Definition 2 directly in terms of the uncertainty and risk premiums.

Proposition 1 *Assume that v_1, I_1, V_1, v_2, I_2, V_2 are as above. Assume that $T_1 = T_2$ and $p^{I_1} = p^{I_2}$. Denote them by T and p. Let $x \in T$ and $z \in \boldsymbol{R}^S$, and suppose that $p \cdot z = E^p[z] = 0$.*

1. *V_1 is at least as ambiguity-averse in choice as V_2 at x in the direction of z if and only if there is an $\bar{\varepsilon} > 0$ such that*

$$\frac{\kappa_1^x(\varepsilon, z) - \bar{\kappa}_1^x(\varepsilon, z)}{\bar{\kappa}_1^x(\varepsilon, z)} \geq \frac{\kappa_2^x(\varepsilon, z) - \bar{\kappa}_2^x(\varepsilon, z)}{\bar{\kappa}_2^x(\varepsilon, z)}. \tag{12}$$

   for every $\varepsilon \in (0, \bar{\varepsilon})$.

2. *V_1 is more ambiguity-averse in choice than V_2 at x in the direction of z if and only if there are an $\bar{\varepsilon} > 0$ and a $\delta > 0$ such that*

$$\frac{\kappa_1^x(\varepsilon, z) - \bar{\kappa}_1^x(\varepsilon, z)}{\bar{\kappa}_1^x(\varepsilon, z)} \geq \frac{\kappa_2^x(\varepsilon, z) - \bar{\kappa}_2^x(\varepsilon, z)}{\bar{\kappa}_2^x(\varepsilon, z)} + \delta. \tag{13}$$

   for every $\varepsilon \in (0, \bar{\varepsilon})$.

To compare Definition 2 with those in the literature, second, we need to fill in a gap in formulation between this and other papers in the literature. The gap lies in the fact that utility functions in other papers are defined mostly on a set of Anscombe-Aumann acts,[2] where the consequences are objective lotteries, while utility functions are defined, in their terminology, on a set of Savage acts, where the consequences are deterministic consumption levels. We will fill it in by using the fictitious utility function $\bar{V}$, which is ambiguity-neutral and has the same risk attitudes as the original utility function V. We will later show that using the fictitious utility function $\bar{V}$ serves essentially for the same purpose as expanding the domain of the original utility function V to a set of Anscombe-Aumann acts.

Definition 3 Assume that V_1 and V_2 are as above. Assume that $T_1 = T_2$ and denote them by T. We say that V_1 *is at least as ambiguity-averse as V_2 in the ranking sense* if $V_2(f) \geq \bar{V}_2(g)$ whenever $f \in T^S$, $g \in T^S$, and $V_1(f) \geq \bar{V}_1(g)$.

When evaluating the act f, both decision makers use the original utility functions V_n, thereby perceiving ambiguity, but when evaluating the act g, they use the fictitious utility functions $\bar{V}_n$, thereby ignoring ambiguity and behaving as if a subjective expected utility maximizer under p^{I_1}. The definition, thus, means that the first decision maker is at least as ambiguity-averse as the second decision maker if, whenever the former finds

[2]An exception is Klibanoff, Marinacci, and Mukerji (2005), who assumed that there are events to which the objective (unique) probabilities are given.

a possibly ambiguous act f at least as preferable as a hypothetically unambiguous act g, the second does so as well. If we were to extend the domain of V_n to the set of all Anscombe-Aumann acts, then $\bar{V}_n(g)$ would be equal to the utility level of the lottery act that gives, at every state, the distribution function of g under the probability p^{I_n}.

Proposition 2 *Assume that V_1 and V_2 are as above. Assume that $T_1 = T_2$, which we denote by T.*

1. *If V_1 is at least as ambiguity-averse as V_2 in the ranking sense (Definition 3), then $p^{I_1} = p^{I_2}$ and for every $x \in T$, every $z \in \boldsymbol{R}^S$ with $E^{p^{I_1}}[z] = E^{p^{I_2}}[z] = 0$, and every $\varepsilon > 0$ with $xe + \varepsilon z \in T^S$,*

$$\kappa_1^x(\varepsilon, z) \geq \kappa_2^x(\varepsilon, z), \tag{14}$$

$$\bar{\kappa}_1^x(\varepsilon, z) \leq \bar{\kappa}_2^x(\varepsilon, z). \tag{15}$$

2. *Suppose that $p^{I_1} = p^{I_2}$. If both* (14) *and* (15) *hold for every $x \in T$ and every $z \in \boldsymbol{R}^S$ with $E^{p^{I_1}}[z] = E^{p^{I_2}}[z] = 0$, every $\varepsilon > 0$ with $xe + \varepsilon z \in T^S$, then V_1 is at least as ambiguity-averse as V_2 in the ranking sense (Definition 3).*

In the first part of the above proposition, it follows from Definition 3 that $p^{I_1} = p^{I_2}$. A similar result was obtained for KMM utility functions by Wang (2019, Theorem 6). In its second part, it is indeed necessary to assume that $p^{I_1} = p^{I_2}$. Once this is assumed, it is sufficient to assume that either $E^{p^{I_1}}[z] = 0$ or $E^{p^{I_2}}[z] = 0$. Combining the two parts, we can say that (14) and (15) constitute an equivalent condition for V_1 to be at least as ambiguity-averse as V_2 in the ranking sense. Note that (14) and (15) are valid for all values of ε that satisfy $xe + \varepsilon z \in T^S$, while (12) is valid for smaller values of ε that, as the proof of this proposition shows, are determined by the size of the residuals of Taylor series expansion.

The inequalities (12), (14), and (15) show that the at-least-as-ambiguity-averse-as relation in the ranking sense implies the at-least-as-ambiguity-averse-as relation in choice.

Corollary 2 *Assume that V_1 and V_2 are as above. Assume that $T_1 = T_2$, which we denote by T. If V_1 is at least as ambiguity-averse as V_2 in the ranking sense (Definition 3), then $p^{I_1} = p^{I_2}$ and, for every $x \in T$ and every $z \in T$ with $E^{p^{I_1}}[z] = E^{p^{I_2}}[z] = 0$, V_1 is at least as ambiguity-averse in choice as V_2 (Definition 2) at x in the direction of z.*

This corollary shows that the standard at-least-as-ambiguity-averse-as relation in the ranking sense implies our at-least-as-ambiguity-averse-as relation in choice. As a prerequisite for the latter relation, it also shows that the two utility functions share the same probability implicit at the constant acts.

Our definition of the more-ambiguity-averse-than relation (Definition 1) is local in nature, while the standard one (Definition 3) is global. The following example shows, in fact, that the former does not imply the latter.

Example 1 Let $S = 2$. For each $n = 1, 2$, define $b_n : \boldsymbol{R} \to \boldsymbol{R}$ by letting

$$b_1(y) = y^2,$$
$$b_2(y) = y^4 + \frac{y^2}{2}$$

for every $y \in \boldsymbol{R}$. Define $I_n : \boldsymbol{R}^2 \to \boldsymbol{R}$ by letting

$$I_n(g) = \frac{1}{2}(g(1) + g(2)) - b_n(g(1) - g(2))$$

for every $g = (g(1), g(2)) \in \boldsymbol{R}^2$. Let $v : T \to \boldsymbol{R}$ be a Bernoulli utility function such that $v(T) \supset [1, 3]$. Define $V_n : T^2 \to \boldsymbol{R}$ by letting $V_n(f) = I_n(v \circ f)$ for every $f \in T^2$. Write $\underline{x} = v^{-1}(2)$ and $\overline{x} = v^{-1}(3)$.

The two utility functions V_1 and V_2 represent variational preferences of Maccheroni, Marinacci, and Rustichini (2006), which we will discuss in more detail in Section 8.3. This follows from the constant additivity of I_1 and I_2 and Theorem 1 (iii) of CVMMM. Indeed, by Theorem 3 of Maccheroni, Marinacci, and Rustichini (2006), if we let $\Delta = \{p = (p(1), p(2)) \in \boldsymbol{R}^2_+ \mid p(1) + p(2) = 1\}$ and define $c_n : \Delta \to \boldsymbol{R}$ by letting

$$c_n(p) = \sup_{g \in \boldsymbol{R}^2} (I(g) - p \cdot g)$$

for every $p \in \Delta$, then

$$I_n(g) = \inf_{p \in \Delta} (p \cdot g + c_n(p))$$

for every $g \in \boldsymbol{R}^2$. For example,

$$c_1(p) = \frac{1}{4}\left(p(1) - \frac{1}{2}\right)^2$$

for every $p \in \Delta$.

Proposition 3 *In Example 1,*

1. $p^{I_n} = (1/2, 1/2)$ *for each* n.

2. *For every $x \in T$ and $z \in \boldsymbol{R}^2$, V_1 is at least as ambiguity-averse as V_2 at x in the direction of z. For every $x \in T$, V_1 is more ambiguity-averse than V_2 at x in the direction of* $(1, -1)$.

3. *There is an $x \in T$ such that $V_1(\overline{x}, \underline{x}) > \bar{V}_1(x, x)$ and $V_2(\overline{x}, \underline{x}) < \bar{V}_2(x, x)$.*

Part 1 of this proposition implies that the two utility functions V_1 and V_2 share the same probability implicit at the constant acts. As shown in Corollary 2, it is a necessary condition for the at-least-as-ambiguity-averse-as relation in the ranking sense. Part 2 says that V_1 is at least as ambiguity-averse as V_2 according to our definition, regardless of the consumption level at the constant act and the direction of zero-mean uncertainty. By Corollary 2, it also implies that V_2 is not at least as ambiguity-averse as V_1 in the ranking sense. Part 3 implies that V_1 is not at least as ambiguity-averse as V_2 in the ranking sense.

5 Pricing implications

Our measure of ambiguity aversion, $H^x(z)$, was defined as the difference in Hessians between the ambiguity-averse utility function V and the ambiguity-neutral utility function $\bar{V}$. In this section, we show that it also measure the difference in the valuations, or prices, of an act between the two utility functions. This result shows that while our measure of ambiguity aversion was defined in terms of utility functions, it can in fact be given an interpretation in terms of choices. Recall that we ask under what prices of acts a given acts is optimal for an ambiguity-averse decision maker. It may be optimal for an ambiguity-neutral decision maker but only under different prices. The difference between these two sets of prices arises from ambiguity aversion, and is shown, by our result, can be quantified by our measure of ambiguity aversion.

The prices, or valuations, in this exercise can be considered as being dual to the behavioral conditions that are standard in decision theory. That is, the so-called axiomatizations in characterizing an ambiguity-averse preference relation consists of a set of axioms that can impose observable restrictions on choices that the decision maker would make. As a typical axiom is stated in the form of which one of two acts or alternatives is preferred, it can be interpreted as specifying a choice that the decision maker having the preference relation that conforms to the axiom would necessarily make. On the other hand, by identifying the prices under which an act is optimal for the ambiguity-averse decision maker, we are, in fact, specifying the set of affordable acts from which the decision maker would necessarily choose the given act. When there is a difference in prices between two utility functions in which an act is optimal for both utility functions, the two utility functions may well be different. In the exercise of this section, we quantify the difference in prices by our measure of ambiguity aversion.

Let $x \in T$ and $z \in \boldsymbol{R}^S$ and suppose that $p^I \cdot z = E^{p^I}[z] = 0$. Since T^S is open, $xe + \varepsilon z \in T^S$ for a sufficiently small $\varepsilon > 0$. For such an ε, define

$$\pi^x(\varepsilon, z) = \frac{\nabla V(xe + \varepsilon z)z}{\nabla V(xe + \varepsilon z)e}. \tag{16}$$

This is the relative price of a zero-mean act z with respect to the constant act e when these two acts (and possibly other acts) can be bought and sold in any amount under constant (linear) prices and the act $xe + \varepsilon z$ is optimal for a decision maker having the utility function V. Define $\bar{\pi}^x(\varepsilon, z)$ analogously for $\bar{V}$. It is the relative price of z with respect to e when both of them can bought and sold in any amount under constant prices and $xe + \varepsilon z$ is optimal for a decision maker having the ambiguity-neutral utility function $\bar{V}$. Our prime interest is the difference $\pi^x(1, z) - \bar{\pi}^x(1, z)$. That is, when a decision maker consumes act $xe + z$, we ask which price he attaches to the zero-mean part z of his act and how much of his evaluation is due to his ambiguity aversion. Since any act f can be written in the form $xe + z$ with $x = E^{p^I}[f]$ and $z = f - E^{p^I}[f]e$, our use of the act $xe + z$ in no way restricts the acts at which the decision maker's evaluation is assessed via $\pi^x(1, z)$. Since $\pi^x(0, z) = (p^I \cdot z)/(p^I \cdot e) = 0$ and $\bar{\pi}^x(0, z) = (p^I \cdot z)/(p^I \cdot e) = 0$, there is no difference in evaluations between the ambiguity-averse decision maker and the ambiguity-neutral decision maker when they consume a constant act. Since our measure of ambiguity aversion, $H^x(z)$, is concerned with a small ambiguity, we are led to study the difference $\pi^x(\varepsilon, z) - \bar{\pi}^x(\varepsilon, z)$ when ε is not equal, but close, to zero. The theorem below is a limit result on the fraction of the price for the nonzero part z of the act $xe + \varepsilon z$ that can be attributed ambiguity aversion to the fictitious price $\bar{\pi}^x(\varepsilon, z)$ in much the same way as Theorem 1, which is concerned with the the uncertainty premium.

Theorem 2 *Let $x \in T$ and $z \in \boldsymbol{R}^S$ and suppose that $z \neq 0$ and $z \neq 0$ and $p^I \cdot z = E^{p^I}[z] = 0$. Then,*

$$\frac{\pi^x(\varepsilon, z) - \bar{\pi}^x(\varepsilon, z)}{\bar{\pi}^x(\varepsilon, z)} \to H^x(z)$$

as $\varepsilon \to 0$.

To understand this theorem, recall that $H^x(z) > 0$ if $\nabla^2 I(xe)$ is negative definite on the hyperplane with normal e, which is true for all ambiguity-averse utility functions we consider in this paper. On the other hand, $\bar{\pi}^x(\varepsilon, z) < 0$ for a small $\varepsilon > 0$, which can be seen from (62) in its proof. Thus, Theorem 2 implies that $\pi^x(\varepsilon, z) < \bar{\pi}^x(\varepsilon, z)$ for a small

ε, that is, the relative price of the act $xe + \varepsilon z$ with respect to the constant act e is lower under V than under $\bar{V}$, and the difference is, in proportion, approximately equal to the measure of ambiguity aversion, $H^x(z)$ when ε is small.

Theorem 2 relates the measure $H^x(z)$ of ambiguity aversion the price of the nonzero-part z of the act (consumption plan) $xe+\varepsilon z$ that can be attributed to ambiguity aversion. In the finance and macroeconomics literature, however, the pricing implications are often stated in terms of expected returns, rather than prices. In the following, we relate the measure $H^x(z)$ of ambiguity aversion to the expected return on the portfolio that pays off $xe + \varepsilon z$.

To do so, define

$$\rho^x(\varepsilon, z) = -\frac{\pi^x(\varepsilon, z)}{x + \pi^x(\varepsilon, z)}.$$

Since $E^{p^I}[xe + z] = x$ and the price of $xe + z$ with respect to the constant act (the payoff of the risk-free bond) e is equal to $xe + \pi^x(\varepsilon, z)$, the expected return on the portfolio that pays off $xe + z$ is equal to

$$\frac{x}{x + \pi^x(\varepsilon, z)} - 1.$$

This is nothing but $\rho^x(\varepsilon, z)$. Since the risk-free rate is zero, $\rho^x(\varepsilon, z)$ is also equal to the equity premium if $xe + z$ is equal to the aggregate consumption. Define $\bar{\rho}^x(\varepsilon, z)$ analogously for $\bar{\pi}^x(\varepsilon, z)$ in place of $\pi^x(\varepsilon, z)$. Then the expected return on the portfolio that pays off $xe + z$ would be equal to $\bar{\rho}^x(\varepsilon, z)$ if the decision maker were ambiguity-neutral. Thus, the difference $\rho^x(\varepsilon, z) - \bar{\rho}^x(\varepsilon, z)$ is the expected return on the portfolio that can be attributed to ambiguity aversion. The theorem below gives a limit result on the fraction of this part to the fictitious expected return $\bar{\rho}^x(\varepsilon, z)$, in much the same way as Theorem 2 give a limit result on the faction of the price for the nonzero part z that can be attributed to ambiguity aversion.

Theorem 3 *For every $x \in T$ and $z \in \boldsymbol{R}^S$ with $z \neq 0$ and $p^I \cdot z = E^{p^I}[z] = 0$,*

$$\frac{\rho^x(\varepsilon, z) - \bar{\rho}^x(\varepsilon, z)}{\bar{\rho}^x(\varepsilon, z)} \to H^x(z)$$

as $\varepsilon \to 0$.

There is another way to relate our measure of ambiguity aversion to expected returns. These expected returns are evaluated not under p^I but under the tilted or distorted probability, which has been used to assess the impact of ambiguity aversion on optimal portfolio choice and asset pricing in terms of biases of subjective probabilities that the (fictitious) ambiguity-neutral decision maker would need to have and, yet find the same act optimal. The literature includes Hansen (2007, Section IV.B), Ju and Miao (2012, Section 3.2), Chen, Ju, and Mia (2014, Section 3.3), Jahan-Parvar and Liu (2014, Section 1.3), Gallant, Jahan-Parvar, and Liu (forthcoming, Section 1.1), Collard, Mukerji, Sheppard, and Tallon (2018, Section 3.2), and Altug, Collard, Çakmakli, Mukerji, and Özsöylev (2018, Section 2.4). Here, we quantify the tilted probability in terms of the expected return under that probability and show that it can be approximately represented by our measure of ambiguity aversion for the representative consumer.

To do so, define

$$\tilde{\rho}^x(\varepsilon, z) = \frac{\dfrac{\nabla I(v \circ (xe + \varepsilon z))z}{\nabla I(v \circ (xe + \varepsilon z))e} - \pi^x(\varepsilon, z)}{x + \pi^x(\varepsilon, z)}$$

Since the denominator of the fraction on the right-hand side is equal to the price of the portfolio that pays off the act $xe + z$ with respect to the constant act (the payoff of the risk-free bond) e, and its numerator is equal to the expected value of the act $xe + z$ evaluated by the decision maker's subjective probability implicit at the act $xe + \varepsilon z$, the expected return on the portfolio that pays off the act $xe + z$, where the expected return is evaluated by the decision maker's probability implicit at the act $xe + \varepsilon z$, is equal to

$$\frac{x + \dfrac{\nabla I(v \circ (xe + \varepsilon z))z}{\nabla I(v \circ (xe + \varepsilon z))e}}{x + \pi^x(\varepsilon, z)} - 1.$$

This is nothing but $\tilde{\rho}^x(\varepsilon, z)$. Since the risk-free rate is zero, $\tilde{\rho}^x(\varepsilon, z)$ is also equal to the equity premium, under the subjective probability implicit at $xe + \varepsilon z$, if $xe + z$ is equal to the aggregate consumption. Thus, $\rho^x(\varepsilon, z) - \tilde{\rho}^x(\varepsilon, z)$ is, therefore, the reduction in the expected rate of return on the same portfolio that is due to the use of the subjective probability, which is tilted towards states s in which $z(s)$ is negative. The theorem below gave a limit result on the fraction of this reduction to the fictitious equity premium $\tilde{\rho}^x(\varepsilon, z)$ in much the same way as Theorem 3 did. But they are different in that Theorem 3 takes, as a reference point, a fictitious expected return $\bar{\rho}^x(\varepsilon, z)$ that would arise if the decision maker were ambiguity-neutral (but the expected payoff is evaluated under p^I), while the theorem below takes, as a reference point, the expected return under the titlted probability (but the price of the portfolio is fixed at the original level $x + \pi^x(\varepsilon, z)$).

Theorem 4 *For every $x \in T$ and $z \in \boldsymbol{R}^S$ with $z \neq 0$ and $p^I \cdot z = E^{p^I}[z] = 0$,*

$$\frac{\rho^x(\varepsilon, z) - \tilde{\rho}^x(\varepsilon, z)}{\tilde{\rho}^x(\varepsilon, z)} \to H^x(z)$$

as $\varepsilon \to 0$.

Given the expected return $\rho^x(\varepsilon, z)$ evaluated under p^I, the fraction in the theorem is higher the lower lower the expected return $\tilde{\rho}^x(\varepsilon, z)$ evaluated under the decision maker's subjective probability implicit at $xe + \varepsilon z$. Since the change is made by tilting probabilities toward states with less favorable outcomes, which is, in turn, caused by ambiguity aversion, the fraction measures the impact of ambiguity aversion. The theorem shows, indeed, that the fraction is very close to our measure of ambiguity aversion when the uncertainty in the act $xe + \varepsilon z$ is small.

The expected return $\tilde{\rho}^x(\varepsilon, z)$ and, accordingly, Theorem 4, could be interpreted in terms of a fictitious decision maker who is ambiguity-neutral and has the same subjective probability as that of the original ambiguity-averse decision maker implicit at $xe + \varepsilon z$. For such a decision maker, $xe + \varepsilon z$ is optimal because the gradient of his utility function coincides with the gradient $\nabla V(xe + \varepsilon z)$ of the original ambiguity-averse decision maker's utility function. Thus, $\tilde{\rho}^x(\varepsilon, x)$ would be the expected return on the portfolio that pays off $xe + z$ in the ambiguity-neutral representative-consumer economy. This interpretation is, in fact, more standard in the asset-pricing literature where the impact of ambiguity aversion is evaluated in terms of the tilting of the probability (a deviation from the natural or physical probability) while assuming ambiguity-neutrality for the representative consumer.

6 Extension to the non-differentiable case

While our measure of ambiguity aversion, $H^x(z)$ defined in (2), requires the utility function V to be twice-differentiable, this requirement eliminates, most notably, maximin expected utility of Gilboa and Schmeidler (1989) and Choquet expected utility of Schmeidler (1989). As they have significant behavioral implications, such as the inertia in portfolio choice, we attempt, in this section, to extend the definition of the measure of ambiguity aversion to such non-differentiable utility functions. Mathematically, while we gave the definition (2) in terms of the Hessian matrix, we will give the extended definition in terms of the second directional derivatives from the right, as they exist under much milder conditions. The fact that the extended definition is determined by the second derivatives indicates, however, that its role in behavioral implications is subordinate to the first-order ambiguity aversion in maximin expected utility and Choquet expected utility. Indeed, our extended measure can only compare ambiguity aversion between two decision makers who share the same first-order ambiguity attitudes. If they do not, then the ambiguity aversion is more pronounced in choice behavior of the decision maker who has, say, a larger set of probabilities implicit at constant acts, even when the measure of ambiguity aversion of the other decision maker is larger.

Formally, for each $x \in T$ and $z \in \boldsymbol{R}^S$, write $\gamma^x(\varepsilon, z) = V(xe + \varepsilon z) - V(xe)$. It is a function well defined for every ε close to 0. We assume that it is right-differentiable at every ε that is non-negative and sufficiently close to 0. Denote the right derivatives by

$$\frac{\partial\, \gamma^x}{\partial_+ \varepsilon}(\varepsilon, z).$$

It is a function of ε. We assume that it is right-differentiable at 0 and denote the second right-derivative by

$$\frac{\partial^2\, \gamma^x}{\partial_+ \varepsilon^2}(0, z).$$

If V is twice differentiable, then so is $\gamma^x(\cdot, z)$ and

$$\frac{\partial\, \gamma^x}{\partial \varepsilon}(0, z) = \nabla V(xe) z, \tag{17}$$

$$\frac{\partial^2\, \gamma^x}{\partial \varepsilon^2}(0, z) = z^\top \nabla^2 V(xe) z. \tag{18}$$

To define the measure of ambiguity aversion based on $\gamma^x(\cdot, z)$, we also need to define an ambiguity-neutral utility function $\bar{V}$. For this purpose, denote by $\Delta^x(z)$ the set of all $p \in \Delta$ be such that $v'(x)p$ is a supergradient of V at xe, that is, $V(xe+w) \leq V(xe)+(v'(x)p)\cdot w$ for every $w \in \boldsymbol{R}^S$ with $xe + w \in T^S$, and that $(v'(x)p) \cdot z = \partial\, \gamma^x(0, z)/\partial_+\varepsilon$. This is the set of all probabilities that are implicit at the constant act xe when evaluating the small deviations in the direction of z. The following lemma shows that $\Delta^x(z)$ is nonempty under weak conditions.

Lemma 1 *Assume that V is non-decreasing and concave. Then, for every $x \in T$ and every $z \in \boldsymbol{R}^S$, $\partial\, \gamma^x(0, z)/\partial_+\varepsilon$ exists and $\Delta^x(z) \neq \varnothing$.*

Let $p \in \Delta^x(z)$. Define $\bar{I}_p : \boldsymbol{R}^S \to \boldsymbol{R}$ by letting $\bar{I}_p(g) = p \cdot g$ for every $g \in \boldsymbol{R}^S$ an define $\bar{V}_p : T^S \to \boldsymbol{R}$ by letting $\bar{V}_p(f) = \bar{I}(v \circ f)$ for every $f \in T^S$. Define $\bar{\gamma}^x_p(\varepsilon, z)$ for $\bar{V}_p$ in place of V. Then $\bar{V}_p$ and $\bar{\gamma}^x_p(\cdot, z)$ are twice differentiable and, by (17) and (18),

$$\frac{\partial \bar{\gamma}^x_p}{\partial \varepsilon}(0, z) = \nabla \bar{V}_p(xe)z = v'(x)(p \cdot z) = v'(x)E^p[z] = \frac{\partial\, \gamma^x}{\partial_+ \varepsilon}(0, z), \tag{19}$$

$$\frac{\partial^2 \bar{\gamma}^x_p}{\partial \varepsilon^2}(0, z) = z^\top \nabla^2 \bar{V}_p(xe)z = v''(x)\|z\|^2_p = v''(x)\left((E^p[z])^2 + \mathrm{Var}^p[z]\right). \tag{20}$$

Thus, the (first) derivative $\partial \bar{\gamma}^x_p(0, z)/\partial \varepsilon$ does not depend on the the choice of $p \in \Delta^x(z)$ but the second derivative $\partial^2 \bar{\gamma}^x_p(0, z)/\partial \varepsilon^2$ does so through the squared norm $\|z\|^2_p$. This is not surprising because the second derivative of an expected utility from $xe + \varepsilon z$ depends on the squared norm $\|z\|^2_p$, although the first derivative depends only on the mean $E^p[z]$.

Define

$$H^x_p(z) = \frac{\dfrac{\partial^2\, \gamma^x}{\partial_+ \varepsilon^2}(0, z) - \dfrac{\partial^2 \bar{\gamma}^x_p}{\partial \varepsilon^2}(0, z)}{\dfrac{\partial^2 \bar{\gamma}^x_p}{\partial \varepsilon^2}(0, z)} \tag{21}$$

If V is twice differentiable, then $\Delta^x(z)$ is a singleton, that is, $\Delta^x(z) = \{v'(x)p\}$ and, by (17), (18), (19), and (20), $H^x_p(z)$ coincides with $H^x(z)$ defined in (2). When V is not differentiable, however, $\Delta^x(z)$ consists of more than one elements and $H^x_p(z)$ may depend on the choice of $p \in \Delta^x(z)$. Of these multiple values of $H^x_p(z)$, we use the infimum as our extended measure of ambiguity aversion, that is, for each $x \in T$ and $z \in \boldsymbol{R}^S \setminus \{0\}$, we define

$$H^x(z) = \inf_{p \in \Delta^x(z)} H^x_p(z). \tag{22}$$

Since

$$H^x_p(z) = \frac{\left|\dfrac{\partial^2\, \gamma^x}{\partial_+ \varepsilon^2}(0, z)\right|}{|v''(x)|} \frac{1}{\|z\|^2_p} - 1,$$

$$H^x(z) = \frac{\left|\dfrac{\partial^2\, \gamma^x}{\partial_+ \varepsilon^2}(0, z)\right|}{|v''(x)|} \frac{1}{\sup\limits_{p \in \Delta^x(z)} \|z\|^2_p} - 1. \tag{23}$$

That is, $H^x(z)$ is determined by the probability p that is implicit at the constant act xe when evaluating the small deviations in the direction of z and, moreover, attains the largest squared norm $\|z\|^2_p$ among all such p's.

We need to justify the use of the infimum (rather than the supremum, for example) of the $H_p^x(z)$ in (22), or, equivalently, the use of the supremum (rather than the infimum, for example) of the $\|z\|_p^2$ in (23). To do so, note first that if p is a supergradient of I at a constant act xe, then $\bar{V}_p \geq V$ because v and I are concave. Thus, when we choose a supergradient p so that $\bar{V}_p$ approximates V around xe in the direction of z, the smaller the value of $\bar{V}_p(xe + \varepsilon z)$, the better the approximation. Obtaining the best approximation is, indeed, important because $H_p^x(z)$ in (21) would otherwise include the discrepancy between the probability implicit at xe for deviations in the direction of z and a (erroneously chosen) probability p for which $\bar{V}_p$ is not the best approximation of V. The following lemma implies that taking the supremum of $\|z\|_p^2$ over p is the right thing to do because the supremum is attained when $\bar{V}_p$ is the best approximation of V.

Lemma 2 *Let $x \in T$ and $z \in \boldsymbol{R}^S$. Let p and q be supergradients of I at $v(x)e$. If there are sequences $(p_n)_n$, $(q_n)_n$, and $(\varepsilon_n)_n$ such that p_n and q_n are supergradients of V at xe, $\varepsilon_n > 0$, $p_n \cdot z = q_n \cdot z$, and $\bar{V}_{p_n}(xe + \varepsilon_n z) \leq \bar{V}_{q_n}(xe + \varepsilon_n z)$ for every n, and that $p_n \to p$, $q_n \to q$, and $\varepsilon_n \to 0$ as $n \to \infty$, then $\|z\|_p^2 \geq \|z\|_q^2$.*

In the conditions of this lemma, the equality $p_n \cdot z = q_n \cdot z$ means that there is no difference in the expected values of the deviation z under p_n and q_n. By (19), the equality holds whenever $p_n \in \Delta^x(z)$ and $q_n \in \Delta^x(z)$.

By definition, $\gamma^x(0, z) = \bar{\gamma}_p^x(0, z) = 0$ for every $p \in \Delta^x(z)$. By (19),

$$\frac{\partial \bar{\gamma}_p^x}{\partial \varepsilon}(0, z) = \frac{\partial \gamma^x}{\partial_+ \varepsilon}(0, z) = 0.$$

By applying L'Hôpital's rule twice, we obtain

$$\frac{\gamma^x(\varepsilon, z) - \bar{\gamma}_p^x(\varepsilon, z)}{\bar{\gamma}_p^x(\varepsilon, z)} \to \frac{\dfrac{\partial^2 \gamma^x}{\partial_+ \varepsilon^2}(0, z) - \dfrac{\partial^2 \bar{\gamma}_p^x}{\partial \varepsilon^2}(0, z)}{\dfrac{\partial^2 \bar{\gamma}_p^x}{\partial \varepsilon^2}(0, z)}$$

as $\varepsilon \to 0+$. Thus

$$H^x(z) = \inf_{p \in \Delta^x(z)} \lim_{\varepsilon \to 0+} \frac{\gamma^x(\varepsilon, z) - \bar{\gamma}_p^x(\varepsilon, z)}{\bar{\gamma}_p^x(\varepsilon, z)}.$$

This gives an expression of the extended measure of ambiguity aversion as the limit of changes in utility levels from the constant act xe caused by the small deviations in the direction of z.

As an example, consider the case where there is a finite subset Δ^* of Δ such that

$I(g) = \min_{p\in\Delta^*} E^p[z]$. This falls, of course, into the case of maximin expected utilities of Gilboa and Schmeidler (1989), and V is not differentiable. Then, for every sufficiently small $\varepsilon \geq 0$,

$$\frac{\partial\,\gamma^x}{\partial_+\varepsilon}(\varepsilon, z) = \min_{p\in\Delta^*} \sum_s p(s)z(s)v'(x + \varepsilon z(s)).$$

Thus,

$$\frac{\partial\,\gamma^x}{\partial_+\varepsilon}(0, z) = v'(x) \min_{p\in\Delta^*} p \cdot z$$

and

$$\frac{\partial^2\,\gamma^x}{\partial_+\varepsilon^2}(0, z) = v''(x) \max_{\substack{p\in\Delta^* \\ p\cdot z = \max\{q\cdot z | q\in\Delta^*\}}} \|z\|_p^2.$$

By (23), $H^x(z) = 0$. In some cases, there is a $p \in \Delta^*$ such that $p \cdot z = \max\{q \cdot z \mid q \in \Delta^*\}$ and, yet $\|z\|_p^2$ is less than the maximum squared norm. For such a p, $H_p^x(z) > H^x(z)$.

To show the applicability of this extension to the non-differentiable case, we now extend Theorem 1 as the extension implies that the results in Section 4 on the at-least-as-ambiguity-as relation in choice and in the ranking sense can also be extended to the non-differentiable case.

As in the proof of Theorem 1, consider the function $(k, \varepsilon) \mapsto V((x-k)e) - V(xe+\varepsilon z)$. The value is equal to $v(x - k) - (\gamma^x(\varepsilon, z) + v(x))$, its partial derivative with respect to k is equal to $-v'(x - k)$, and the right partial derivative with respect to ε is equal to $-\partial\,\gamma^x(\varepsilon, z)/\partial_+\varepsilon$. Define $\kappa^x(\varepsilon, z)$ by $V((x - \kappa^x(\varepsilon, z))e) - V(xe + \varepsilon z) = 0$. By the implicit function theorem,

$$\frac{\partial\kappa^x}{\partial_+\varepsilon}(\varepsilon, z) = -\frac{1}{v'(x)} \frac{\partial\,\gamma^x}{\partial_+\varepsilon}(\varepsilon, z).$$

Since $\partial\,\gamma^x(\varepsilon, z)/\partial_+\varepsilon$ is assumed to be a function of ε that is differentiable at $\varepsilon = 0$,

$$\frac{\partial^2\kappa^x}{\partial_+\varepsilon^2}(\varepsilon, z) = -\frac{1}{v'(x)} \frac{\partial^2\,\gamma^x}{\partial_+\varepsilon^2}(0, z).$$

Let $p \in \Delta^x(z)$ and define $\bar{I}_p$ and $\bar{V}_p$ as before, and $\bar{\gamma}_p^x(\cdot, z)$ and $\bar{\kappa}_p^x(\cdot, z)$ just like $\gamma^x(\cdot, z)$ and $\kappa^x(\cdot, z)$ but using $\bar{V}_p$ in place of V. We can similarly show that

$$\frac{\partial \bar{\kappa}_p^x}{\partial \varepsilon}(0,z) = -\frac{1}{v'(x)}\frac{\partial \bar{\gamma}_p^x}{\partial \varepsilon}(0,z),$$
$$\frac{\partial^2 \bar{\kappa}_p^x}{\partial \varepsilon^2}(0,z) = -\frac{1}{v'(x)}\frac{\partial^2 \bar{\gamma}_p^x}{\partial \varepsilon^2}(0,z).$$

Thus,

$$\frac{\dfrac{\partial^2 \kappa^x}{\partial_+ \varepsilon^2}(0,z) - \dfrac{\partial^2 \bar{\kappa}_p^x}{\partial \varepsilon^2}(0,z)}{\dfrac{\partial^2 \bar{\kappa}_p^x}{\partial \varepsilon^2}(0,z)} = \frac{\dfrac{\partial^2 \gamma^x}{\partial_+ \varepsilon^2}(0,z) - \dfrac{\partial^2 \bar{\gamma}_p^x}{\partial \varepsilon^2}(0,z)}{\dfrac{\partial^2 \bar{\gamma}_p^x}{\partial \varepsilon^2}(0,z)} = H_p^x(z).$$

If $\partial \gamma^x(0,z)/\partial_+\varepsilon = 0$, then, by (19), $\partial \bar{\gamma}_p^x(0,z)/\partial \varepsilon = 0$ and we can apply L'ôpital's rule twice to show that

$$\frac{\gamma^x(\varepsilon,z) - \bar{\gamma}_p^x(\varepsilon,z)}{\bar{\gamma}_p^x(\varepsilon,z)} \to H_p^x(z)$$

as $\varepsilon \to 0+$. This is an extension of Theorem 1 to the non-differentiable case. In particular, if $\|p\|_p^2 = \max_{q \in \Delta^x(z)} \|z\|_q^2$, then $H_p^x(z) = H^x(z)$.

We now move on to compare two ambiguity-averse utility functions that need not be differentiable. Suppose that T is an open interval of $\boldsymbol{R}$ that represents the consumption levels. For each $n = 1, 2$, let $v_n : T \to \boldsymbol{R}$ be a twice differentiable Bernoulli utility function that satisfies $v_n'' < 0 < v_n'$, let $I_n : v_n(T)^S \to \boldsymbol{R}$ be a concave and normalized function, and define $V_n : T \to \boldsymbol{R}$ by letting $V_n(f) = I_n(v_n \circ f)$ for every $f \in T^S$.

For each $x \in T$, $z \in \boldsymbol{R}^S$, and sufficiently small $\varepsilon \geq 0$, define $\gamma_n^x(\varepsilon, z) = V_n(xe + \varepsilon z) - V_n(xe)$. We assume that $\gamma_n^x(\cdot, z)$ is right-differentiable on some non-degenerate interval of $\boldsymbol{R}_+$ that contains 0, and the first-derivative is right-differentiable at 0. Denote by $\Delta_n^x(z)$ the set of all $p \in \Delta$ be such that $v_n'(x)p$ is a supergradient of V_n at xe and that $(v_n'(x)p) \cdot z = \partial \gamma_n^x(0,z)/\partial_+\varepsilon$.

For each $n = 1, 2$ and $p \in \Delta_n^x(z)$, define $\bar{I}_{n,p_n} : \boldsymbol{R}^S \to \boldsymbol{R}$ by letting $\bar{I}_{n,p}(g) = p \cdot g$ for every $g \in \boldsymbol{R}^S$ and $\bar{V}_{n,p} : T^S \to \boldsymbol{R}$ by letting $\bar{V}_{n,p}(f) = \bar{I}_{n,p}(v_n \circ f)$ for every $f \in T^S$. Write $\bar{\gamma}_{n,p}^x(\varepsilon, z) = \bar{V}_{n,p}(xe + \varepsilon z) - \bar{V}_{n,p}(xe)$. We can then $H_{n,p}^x(z)$ as in (21). If $p^n \in \Delta_n^x(z)$ and $\|z\|_{p^n}^2 \geq \|z\|_p^2$ for every $p \in \Delta_n^x(z)$, then $H_{n,p^n}^x(z) = H_n^x(z)$, where $H_n^x(z)$ is defined as in (22).

Suppose that $\partial \gamma_n^x(0,z)/\partial_+\varepsilon = 0$ for each n. Then $p^n \cdot z = 0$ for each n. We can then extend Definition 2 to this non-differentiable case using $\bar{V}_{1,p_1}$ and $\bar{V}_{2,p_2}$ in place of $\bar{V}_1$ and $\bar{V}_1$. We can also extend Corollary 1 to this non-differentiable case. These extensions allow us to translate the definition of ambiguity aversion in terms of the second directional derivatives from the right to the choice behavior, but a caveat is in order. It is that if p^1 and p^2 cannot taken to be equal, then the two ambiguity-neutral utility functions $\bar{V}_{1,p^1}$ and $\bar{V}_{2,p^2}$ are expected utility functions under different probabilities, and the difference in ambiguity attitudes between the two is, in part, due to the difference in probabilities implicit at the constant act.

7 Numerical exercise

In this section, we conduct a simple numerical exercise to derive our measure $H^x(z)$ of ambiguity aversion and show how large it must be to explain observed choices while keeping the Arrow-Pratt measure of ambiguity aversion reasonably low. We use the data reported in Mehra and Prescott (1985), because our result will then be easily comparable with those in the literature on the equity premium puzzle. We give a simple calibration exercise, only to illustrate how the limit results in Sections 4 and 5 can be used, without paying attention to estimation errors. The measure of ambiguity aversion that we derive should therefore be understood only as being close to the true measure.

Consider a decision maker having a utility function V defined by a function I that is normalized and satisfies Assumptions 1 and 2. Let p^I be his subjective probability implicit at the constant acts. In our analysis, we will assume that it coincides with the natural (physical) probability on the state space S. Denote the decision maker's consumption by $xe + z$, where $x \in T$ and $E^{p^I}[z] = 0$. That is, the consumption has mean x under the natural probability, and z is its zero-mean part.

Assume that the aggregate consumption $xe + z$ follows the log-normal distribution with parameters μ and σ^2 under p^I. Our analysis has so far assumed that the state space is finite, but any lognormal distribution requires a continuum of states. Yet, in the following, we shall proceed as if all our results were valid even in the case of infinitely many states.[3] We write $c = \ln(xe + z)$ and treat c as a random variable that follows the normal distribution with mean μ and variance σ^2. As for the Bernoulli utility function, we let $v(x) = (x^{1-\theta} - 1)/(1 - \theta)$ for every $x > 0$, that is, v has a constant coefficient θ of relative risk aversion.

We now express some key parameters in terms of μ, σ^2, and θ. First, the mean of

[3]As will be mentioned in the conclusion, extending our analysis to the case of infinitely many states is a topic of future research.

$\exp(c)$ is equal to $\exp(\mu + \sigma^2/2)$ and

$$\begin{aligned}\bar{V}(\exp(c)) &= \frac{1}{1-\theta}\left(E[\exp((1-\theta)c)] - 1\right) \\ &= \frac{1}{1-\theta}\left(\exp\left(\mu(1-\theta) + \frac{\sigma^2}{2}(1-\theta)^2\right) - 1\right).\end{aligned}$$

Thus, the risk premium $\bar{\kappa}^x(\varepsilon, z)$ satisfies

$$\exp\left(\mu(1-\theta) + \frac{\sigma^2}{2}(1-\theta)^2\right) = \exp\left((1-\theta)\left(\exp\left(\mu + \frac{\sigma^2}{2}\right) - \bar{\kappa}^x(1,z)\right)\right),$$

that is,

$$\bar{\kappa}^x(1,z) = \exp\left(\mu + \frac{\sigma^2}{2}\right) - \exp\left(\mu + \frac{\sigma^2}{2}(1-\theta)\right).$$

Thus, its ratio to the expected consumption level is given by

$$\frac{\bar{\kappa}^x(1,z)}{x} = \frac{\exp\left(\mu + \frac{\sigma^2}{2}\right) - \exp\left(\mu + \frac{\sigma^2}{2}(1-\theta)\right)}{\exp\left(\mu + \frac{\sigma^2}{2}\right)} = 1 - \exp\left(-\frac{\sigma^2\theta}{2}\right) \tag{24}$$

As for asset pricing, by the first-order condition for utility maximization, if the decision maker were ambiguity-neutral, the pricing kernel (also known as the state-price deflator or the stochastic discount factor) is a scalar multiple of $v'(\exp(c))$. Since $E[v'(\exp(c))] = \exp(-\mu\theta + \sigma^2\theta^2/2)$ and the price of the risk-free bond is taken to be one, it can be taken as

$$\frac{1}{E[v'(\exp(c))]}v'(\exp(c)) = \exp\left(\mu\theta - \frac{\sigma^2}{2}\theta^2\right)\exp(-\theta c).$$

The mean of $\exp(c)$ is equal to $\exp(\mu + \sigma^2/2)$ and its price is equal to

$$\begin{aligned}&\exp\left(\mu\theta - \frac{\sigma^2}{2}\theta^2\right)E[\exp(-\theta c)\exp(c)] \\ &= \exp\left(\mu\theta - \frac{\sigma^2}{2}\theta^2\right)\exp\left(\mu(1-\theta) + \frac{\sigma^2}{2}(1-\theta)^2\right) \\ &= \exp\left(\mu + \frac{\sigma^2}{2}(1-2\theta)\right)\end{aligned}$$

Thus the expected return on a portfolio that pays off $\exp(c)$ is given by

$$\bar{\rho}^x(1,z) = \exp\left(\mu + \frac{\sigma^2}{2}\right)\exp\left(-\mu - \frac{\sigma^2}{2}(1-2\theta)\right) - 1 = \exp(\sigma^2\theta) - 1. \tag{25}$$

Note that the equity premium is determined solely by the variance of the logarithm of the consumption, σ^2. Its mean μ is irrelevant for the equity premium.

As for the values of μ and σ^2, we follow the data reported by Mehra and Prescott (1985) on the annual (real) per capita consumption growth rate. Their sample mean and the sample standard deviation are 0.0183 and 0.0357. Let C be the current aggregate consumption, then the consumption growth rate is $C^{-1}\exp(c) = \exp(c - \ln C)$, and $c - \ln C$ follows the normal distribution with mean $\mu - \ln C$ and variance σ^2. Since the mean and variance of $\exp(c - \ln C)$ are equal to $\exp(\mu - \ln C + \sigma^2/2)$ and $(\exp(\sigma^2) - 1)\exp(2(\mu - \ln C) + \sigma^2)$, we set the values of μ and σ^2 so that

$$\begin{aligned}\exp\left(\mu - \ln C + \frac{\sigma^2}{2}\right) &= a,\\ (\exp(\sigma^2) - 1)\exp(2(\mu - \ln C) + \sigma^2) &= b^2,\end{aligned}$$

where $a = 1 + 0.0183$ and $b = 0.0357$. By solving these equations for μ and σ^2 in terms of a and b, we obtain

$$\begin{aligned}\mu - \ln C = \ln a - \frac{1}{2}\ln\frac{a^2 + b^2}{a^2} &= 0.0175,\\ \sigma^2 = \ln\frac{a^2 + b^2}{a^2} &= 0.0012.\end{aligned}$$

Because of this choice of parameter values, in the context of asset pricing, we can think of the decision maker as the representative consumer, his consumption $\exp(c)$ as the aggregate consumption, and a portfolio that pays off $\exp(c)$ as the market portfolio.

Having laid out the setting for our numerical exercise, we now explain how Theorems 1 and 3 can be used to infer the measure of ambiguity aversion. Theorem 4 will be used to assess how reasonable or plausible the derived measure of ambiguity aversion is.

As Chen, Ju, and Miao (2014) did for their calibration exercise on the optimal portfolio, the ambiguity premium, $\kappa^x(1,z) - \bar{\kappa}^x(1,z)$, is often taken to lie between ten and twenty percents of the expected consumption levels. To obtain a conservative estimate of the measure of ambiguity aversion, we suppose that the decision maker's ambiguity premium is ten percents of the expected consumption levels and see how the measure of ambiguity aversion is related to the coefficient θ of relative risk aversion. That is, assume that

$$\frac{\kappa^x(1,z)-\bar{\kappa}^x(1,z)}{x}=0.1.$$

According to Theorem 1, if $\varepsilon = 1$ may be deemed small, then

$$\frac{\kappa^x(1,z)-\bar{\kappa}^x(1,z)}{\bar{\kappa}^x(1,z)}\approx H^x(z).$$

Thus, we let

$$\frac{\bar{\kappa}^x(1,z)}{x}H^x(z)=0.1,$$

that is,

$$H^x(z)=\frac{0.1x}{\bar{\kappa}^x(1,z)}. \tag{26}$$

Then, by (24),

$$H^x(z)=\frac{0.1}{1-\exp(-0.006\theta)}$$

In Table 1, we give the measure of ambiguity aversion that corresponds to each of some commonly used coefficients of relative risk aversion.

For the logarithmic Bernoulli utility function, $\theta = 1$ and $H^x(z)$ exceeds 166. This means that the ambiguity premium is more than 166 times the risk premium. As the coefficient θ of constant relative risk aversion goes up, the measure $H^x(z)$ of ambiguity aversion goes down, because, then, the risk premium goes up while the ambiguity premium stays at the same level (ten percent of the expected consumption). Lucas (1994,

θ	$H^x(z)$
1	166.72
2.5	66.72
5	33.38
10	16.72
25	6.72
50	3.38
175	1.00

Table 1: Relationship between the coefficients of constant risk aversion and the measure of ambiguity aversion that generate the ambiguity premium equal to ten percents of the expected consumption level.

Section 4.1) claimed that in any attempt to solve the equity premium puzzle by appealing to incomplete asset markets, the coefficient of constant relative risk aversion should be at most 2.5, because, otherwise, the proposed solution would be regarded as a consequence of high coefficients of relative risk aversion. If we stick to this view, the measure of ambiguity aversion must exceed 66, which means that the ambiguity premium is more than 66 times the risk premium. Mehra and Prescott (1985) gave consideration to the coefficient of constant relative risk aversion up to ten, but the measure of ambiguity aversion still needs to exceed sixteen. The ambiguity premium is equal to the risk premium when the coefficient of relative risk aversion is approximately equal to 175. A casual observation we can make on this numerical exercise is that since the risk premium is rather small for coefficients of constant relative risk that may be deed reasonable or plausible, the measure of ambiguity aversion must be so large that the ambiguity premium is at least sixteen times the risk premium.

We now explore an implication of Theorem 3. Note first that we can state he equity premium puzzle of Mehra and Prescott (1985) in our notation as follows: If the representative consumer were ambiguity-neutral, then by (25), $\exp(0.0012\theta) - 1 = 0.0618$, that is, $\theta = 49.97$. This coefficient is much higher than is the coefficients inferred from the findings of laboratory experiments or introspection. We now ask whether we can reconcile, with this empirical equity premium, a more plausible, much lower, coefficient of constant relative risk aversion by introducing ambiguity aversion. Recall from Theorem 3 that if we deem $\varepsilon = 1$ deem small, then

$$\frac{\rho^x(1,z) - \bar{\rho}^x(1,z)}{\bar{\rho}^x(1,z)} \approx H^x(z).$$

We let $\rho^x(1,z) = 0.0618$, $\bar{\rho}^x(1,z)$ be as in (25), and

$$H^x(z) = \frac{\rho^x(1,z) - \bar{\rho}^x(1,z)}{\bar{\rho}^x(1,z)} = \frac{0.0618 - (\exp(0.0012\theta) - 1)}{\exp(0.0012\theta) - 1}$$

From this, we can derive the relationship between the constant coefficient θ of risk aversion on the one hand and the measure $H^x(z)$ of ambiguity aversion on the other that jointly generate the historical equity premium, 6.18%. In Table 2, we give the value of the measure of ambiguity aversion that corresponds to each of some commonly used coefficients of relative risk aversion.

For the logarithmic Bernoulli utility function, $\theta = 1$ and $H^x(z)$ exceeds fifty. This means that the part of the equity premium that can be attributed to the ambiguity aversion is more than fifty times the part that can be attributed to risk aversion. According to Lucas (1994, Section 4.1), the coefficient of relative risk aversion should be at most 2.5.

Then the measure of ambiguity aversion needs to be around twenty, which means that the part of the equity premium that can be attributed to the ambiguity aversion is about twenty times the part that can be attributed to risk aversion. Even if the coefficient of constant relative risk aversion is increased to ten, but the measure of ambiguity aversion still needs to exceed four. The ambiguity-attributable part and the risk-attributable part are approximately equal the coefficient of constant relative risk aversion is about 25, and the former can be made zero only if, as we already saw, the coefficient is nearly fifty.

The measures of ambiguity aversion implied by the ambiguity premium (in Table 1) are much larger than those of the equity premium (in Table 2) for all choices of θ. This can also be confirmed by Figure 1. If the subject in experiments (for whom the measures of ambiguity aversion are given in Table 1) and the representative consumer in stock markets (for whom the measures of ambiguity aversion are given in Table 2) share the same ambiguity attitudes in the sense of Ghirardato, Maccheroni, and Marinacci (2004) and KMM, for example,[4] then the difference in the measure of ambiguity aversion indicates that the ambiguity perceived in bets or rewards in experiments is much larger than the ambiguity perceived in stock returns. This is somewhat understandable because the experiments often pose choice questions based on Ellsberg urns of which composition of colors of balls are entirely unknown, while some past return data are available for stocks. In other words, the estimate of the ambiguity premium being ten percents or higher of the expected reward or consumption should better be revised downwards when applied to the analysis of asset markets.[5]

Take the measure 19.57 of ambiguity aversion derived from the equity premium when

θ	$H^x(z)$
1	50.47
2.5	19.57
5	9.27
10	4.12
25	1.03
50	0.00

Table 2: Relationship between the coefficients of constant risk aversion and the measure of ambiguity aversion that generate the historical equity premium of 6.18%

[4]However, we cannot deal with their α-maximin expected utility functions in our framework because they are not differentiable.

[5]The premise of the ten-percent ambiguity premium is used mostly when interpreting experimental findings. The money amounts of bets or rewards in experiments are usually much smaller than the stock returns or dividends. But, for the purpose of comparison, we used the same act (consumption plan) $\exp(c)$ to derive the measures of ambiguity aversion from the ambiguity premium and also from the equity premium. This may well be responsible for the much larger measure of ambiguity aversion derived from the ambiguity premium.

the constant coefficient of relative risk aversion is equal to 2.5. Is it reasonable or plausible? On this, Theorem 3 gives us a hint. The measure is so large that the fraction of the equity premium (6.18%) that can be attributed to risk aversion is less than $0.309\,(= 6.18/(19+1))\%$, and the rest is attributed to ambiguity aversion. Theorem 4 gives us another perspective. The representative consumer's subjective probability implicit at the aggregate consumption is so tilted towards states of relatively low consumption levels that the expected return on the market portfolio has a minuscule expected return of 0.309% or less in the eye of the representative consumer. When interpreted in these manners, the measure 19.57 of ambiguity aversion appears to be excessive. For one, ambiguity aversion alone should not really account for such a large part of the equity premium; if risk aversion fails to account for it, we should probably believe that other economic or financial factors such as incomplete markets and transaction costs also contribute to the high equity premium. For another, the probability tilting is so large that the representative consumer's subjective probability implicit at the aggregate consumption moves away from, and has no coherence to, the natural probability p^I, inducing him to be quantitatively almost irrational. We should then probably believe that such tilted probabilities will be corrected in the long run even when estimations are difficult. Our measure of ambiguity aversion is useful because it tells us that we might be counting too much on ambiguity aversion to explain an otherwise unexplainable phenomenon and should be looking into other economic factor for a solution.

8 Examples

In this section, we show how the measure H^x of ambiguity aversion can be rewritten in some well known subclasses of uncertainty-averse utility functions.

8.1 Smooth ambiguity

Let the state space $\{1, 2, \ldots, S\}$, the nonempty open interval T of consumption levels, and a Bernoulli utility function $v : T \to \boldsymbol{R}$ be as in Section 2. Let $\phi : v(T) \to \boldsymbol{R}$ be twice continuously differentiable and satisfy $\phi'' \leq 0 < \phi'$. Denote by Δ the set of all probabilities on the state space $\{1, 2, \ldots, S\}$. It can be identified with the unit simplex $\{p \in \boldsymbol{R}^S_+ \mid p \cdot e\} = 1$. Let μ be a probability measure on Δ such that $\int_\Delta p \, \mathrm{d}\mu(p) \in \boldsymbol{R}^S_{++}$. Define $I : v(T)^S \to \boldsymbol{R}$ by letting

$$I(z) = \phi^{-1} \left(\int_\Delta \phi \left(p \cdot z \right) \mathrm{d}\mu(p) \right) \tag{27}$$

for every $z \in v(T)^S$. Then define $V : T^S \to \boldsymbol{R}$ by letting $V(f) = I(v \circ f)$ for every $f \in T^S$. This subclass of ambiguity-averse utility functions was axiomatized by Klibanoff, Marinacci, and Mukerji (2005). Denote by E^p and Var^p the expectation operator and the variance operator for the random variables defined on the state space $\{1, 2, \ldots, S\}$ under the probability p. Then $p \cdot z = E^p[z]$ and, hence,

$$\phi(V(f)) = \int_\Delta \phi\left(E^p(v \circ f)\right) \mathrm{d}\mu(p). \tag{28}$$

Thus the decision maker has an subjective expected utility function under the (subjective) probability μ (which they termed the second-order belief) on the set Δ of probabilities on the state space and a Bernoulli utility function ϕ on the expected utility levels $E^p(v \circ f)$ under different probabilities p on the state space. Let $w = \phi \circ v$, then (28) can be rewritten as

$$\phi(V(f)) = \int_\Delta w\left(v^{-1}\left(E^p(v \circ f)\right)\right) \mathrm{d}\mu(p).$$

Thus the decision maker has an subjective expected utility function under the probability μ on Δ and a Bernoulli utility function w on the certainty equivalents $v^{-1}\left(E^p(v \circ f)\right)$ under different probabilities p on the state space. Hence, the difference between ϕ and w is that the former is a Bernoulli utility function over expected utility levels, while the latter is a Bernoulli utility function over certainty equivalents. Denote by $E^{\cdot}[z]$ the random variable $p \mapsto E^p[z]$ defined on the probability space $(\Delta, \mathscr{B}(\Delta), \mu)$. Denote by E^μ and Var^μ be the expectation operator and the variance operator for the random variance defined on Δ under the probability μ. In the subclass of smooth ambiguity-averse utility functions, the measure of ambiguity aversion can be written as follows.

Proposition 4 *If I is defined by* (27)*, then it satisfies all the conditions of Assumptions 1 and 2 and*

$$H^x(z) = \left(\frac{-\dfrac{w''(x)}{w'(x)}}{-\dfrac{v''(x)}{v'(x)}} - 1 \right) \frac{\mathrm{Var}^\mu[E^{\cdot}[z]]}{\left(E^{p^I}[z]\right)^2 + \mathrm{Var}^{p^I}[z]} \tag{29}$$

for every $x \in T$ and $z \in \boldsymbol{R}^S \setminus \{0\}$.

This proposition shows that in the case of the smooth ambiguity utility functions, the measure $H^x(z)$ of ambiguity aversion consists of two terms. The first term of (29) represents how much in proportion the decision maker is more averse to the risk in certainty equivalents under different probabilities on the state space than to the risk that is independent of the choice of such probabilities. This term is independent of the second-order belief μ and also of the choice of z. The second term is dependent on the choice of z. Since its numerator is the variance of the function $E^{\cdot}[z] : p \mapsto E^p[z]$ under the probability μ on Δ, it measures how much uncertainty is perceived in the distribution of z. The denominator is a normalizing factor, as it is equal to the L^2-norm of z under the probability p^I. Since $p^I = \int_\Delta p \, \mathrm{d}\mu(p)$, as proved by Rigotti, Shannon, and Strzalecki (2008, Proposition 5), the second term is determined by the second-order belief μ and independent of the Bernoulli utility function ϕ on the expected utility levels, and represents how much ambiguity is perceived in the direction of the profile z of state-dependent utility levels. Thus, although the definition of the measure of ambiguity aversion is based on the utility function V, when V is a KMM utility function, it consists of, and can be decomposed into, two parts, the first part determined by the (pure) aversion to ambiguity, and the second part determined by the perception of ambiguity.

Consider two uncertainty-averse utility functions V_1 and V_2 in this subclass that are generated by (v_1, ϕ_1, μ_1) and (v_2, ϕ_2, μ_2). Write $w_1 = \phi_1 \circ v_1$ and $w_2 = \phi_2 \circ v_2$. It will be illustrative to present two polar cases of comparison between V_1 and V_2. First, suppose that

$$\frac{-\dfrac{w_1''(x)}{w_1'(x)}}{-\dfrac{v_1''(x)}{v_1'(x)}} = \frac{-\dfrac{w_2''(x)}{w_2'(x)}}{-\dfrac{v_2''(x)}{v_2'(x)}}$$

for every $x \in T$. This means that the ratio of the Arrow-Pratt measures of absolute risk aversion between v_i and w_i is common between V_1 and V_2. Since the first term of (29) is then common, the ranking of the measures of ambiguity aversion of V_1 and V_2 is determined solely by its second term. If, in addition, μ_1 and μ_2 have the same reduced distribution, that is, $\int_\Delta p \, \mathrm{d}\mu_1(p) = \int_\Delta p \, \mathrm{d}\mu_2(p)$, then $p^{I_1} = p^{I_2}$ and the denominator of the second term is common between V_1 and V_2. Thus $H_1^x(z) \geq H_2^x(z)$ if and only if $\mathrm{Var}^{\mu_1}[E^{\cdot}[z]] \geq \mathrm{Var}^{\mu_2}[E^{\cdot}[z]]$, that is, the second-order belief μ_1 is at least dispersed as μ_2 in the direction of the random variable z. As for the other polar case, suppose that $\mu_1 = \mu_2$. Then the second term of (29) is common between V_1 and V_2. If, in addition, $x \in T$ and

$$\frac{-\dfrac{w_1''(x)}{w_1'(x)}}{-\dfrac{v_1''(x)}{v_1'(x)}} \geq \frac{-\dfrac{w_2''(x)}{w_2'(x)}}{-\dfrac{v_2''(x)}{v_2'(x)}}, \tag{30}$$

then $H^x(z) \geq H^x(z)$ for every z. That is, if V_1 and V_2 have the same second-order belief, then the ranking of the measures of ambiguity aversion is determined by how much, in proportion, the Bernoulli utility function w_i on the certainty equivalents under different probabilities on the state space is more risk-averse than the Bernoulli utility function v_i on the unambiguous acts.

The first term of the measure of ambiguity aversion, which involves the Arrow-Pratt measures of absolute risk aversion of v_i and w_i, deserves special attention. KMM used the definition of the more-ambiguity-averse-than relation that builds on lotteries (unambiguous acts) to prove, in their Theorem 2, that V_1 is more ambiguity-averse than V_2 if and only if, according to our notation, v_1 is an affine transformation of v_2 and w_1 is a concave transformation of w_2. Guetlein (2016) introduced a notion of subjectively risk-free acts to define a KMM utility function as being more risk-averse than another KMM utility function to prove, in her Proposition 1, that V_1 is more risk-averse than V_2 if and only if v_1 is a concave transformation of v_2 and w_1 is an affine transformation of w_2. She concluded that when it comes to comparing ϕ_1 and ϕ_2, no such behavioral counterpart is available. We now argue that the first term of the measure of ambiguity aversion on the right-hand side of (29) points to a way to find such a behavioral counterpart. Indeed, define $\psi_1 = w_1' \circ (v_1')^{-1}$ and $\psi_2 = w_2' \circ (v_2')^{-1}$, that is, we define ψ_1 and ψ_2 in the same way as ϕ_1 and ϕ_2 except that we use marginal utility functions in place of utility functions. Hara (2020) showed that (30) holds if and only if the elasticity of ψ_1 is greater than or equal to that of ψ_2, and gave an equivalent condition in terms of proportional increments in certainty premiums when replacing v_1 by w_1 and also v_2 by w_2. We give a generalization of this behavioral condition in Section 4.

As for the proof method of Proposition 4, in principle, we could apply the proof of Proposition 5 in the next subsection, because the KMM utility functions constitute a subclass of uncertainty-averse utility functions of CVMMM. But our proof is more straightforward as it does not go through the function G in their formulation.

The next example gives two uncertainty-averse utility functions that have different Bernoulli utility functions and second-order beliefs, and, nonetheless, are equally ambiguity-averse. It also provides sufficient conditions for the above two polar cases to be applicable.

Example 2 Let $T = \boldsymbol{R}_{++}$. Assume that v_i has a constant coefficient θ_i of relative risk aversion and w_i has a constant coefficient γ_i of relative risk aversion. Assume that $\theta_1 < \gamma_1$.

Suppose that μ_1 has a finite support $\{p_1^1, p_1^2, \ldots, p_1^N\}$ in Δ. For each n, write $q^n = \mu_1\left(\{p_1^n\}\right)$. Then $p^{I_1} = \sum_n q^n p_1^n$. We assume that it belongs to $\boldsymbol{R}^S_{++}$.

Let $\alpha \in (0, 1)$. Define μ_2 by letting, for each n, $p_2^n = \alpha p_1^n + (1 - \alpha) p^{I_1}$ and $\mu_2\left(\{p_2^n\}\right) = q^n$. That is, the distribution μ_1 of probabilities on the state space is contracted to the mean by factor α to obtain the distribution μ_2.

We claim that for every x and z, if $H_1^x(z) = 0$, then $H_2^x(z) = 0$, and, otherwise,

$$\frac{H_2^x(z)}{H_1^x(z)} = \frac{\gamma_2/\theta_2 - 1}{\gamma_1/\theta_1 - 1}\alpha^2. \tag{31}$$

Indeed,

$$\frac{-\dfrac{w_i''(x)}{w_i'(x)}}{-\dfrac{v_i''(x)}{v_i'(x)}} = \frac{\dfrac{\gamma_i}{x}}{\dfrac{\theta_i}{x}} = \frac{\gamma_i}{\theta_i}. \tag{32}$$

Moreover,

$$p^{I_2} = \sum_n q^n p_2^n = \sum_n q^n \left(\alpha p_1^n + (1-\alpha)p^{I_1}\right) = \alpha p^{I_1} + (1-\alpha)p^{I_1} = p^{I_1}.$$

Thus

$$\left(E^{p^{I_1}}[z]\right)^2 + \mathrm{Var}^{p^{I_1}}[z] = \left(E^{p^{I_2}}[z]\right)^2 + \mathrm{Var}^{p^{I_2}}[z]. \tag{33}$$

Since $\alpha\left(p_1^n - p^{I_1}\right) = p_2^n - p^{I_2}$ for every n,

$$\begin{aligned}
\mathrm{Var}^{\mu_2}[E^{\cdot}[z]] &= \sum_n q^n \left(\left(p_2^n - p^{I_2}\right)\cdot z\right)^2 \\
&= \alpha^2 \sum_n q^n \left(\left(p_1^n - p^{I_1}\right)\cdot z\right)^2 \\
&= \alpha^2 \mathrm{Var}^{\mu_1}[E^{\cdot}[z]].
\end{aligned} \tag{34}$$

By plugging (32), (33), and (34) into (29) for each of V_1 and V_2 and take the ratio of the two, then we obtain (31).

Now that we have established (31), it is easy to construct an numerical example of V_1 and V_2 that have different Bernoulli utility functions and second-order beliefs but are equally ambiguity-averse. In fact, let $\theta_1 = 1$, $\gamma_1 = 3$, $\theta_2 = 2$, $\gamma_2 = 18$, and $\alpha = 1/2$. Then

$$\frac{H_2^x(z)}{H_1^x(z)} = \frac{18/2 - 1}{3/1 - 1}\left(\frac{1}{2}\right)^2 = \frac{8}{2}\cdot\frac{1}{4} = 1.$$

That is, although w_2 is more risk-averse than v_2 to a fourfold extent that w_1 is more risk-averse than v_1, this difference is cancelled out by the difference in the second-order beliefs, where μ_1 is twice dispersed than μ_2.

Another case of interest is where $\theta_1 = 1$, $\gamma_1 = 2$, $\theta_2 = 1$, $\gamma_2 = 4$, and $\alpha = 1/2$. In this case, V_1 and V_2 have the same Bernoulli utility function v_1 and v_2 for risks without ambiguity, and the Bernoulli utility functions w_2 for certainty premiums under different probabilities on the state space is more risk-averse than w_1. Thus, I_2 is more concave than I_1, and, yet, V_1 is more ambiguity-averse than V_2 because μ_1 is twice as much dispersed as μ_2. That is, V_1 may make a more ambiguity-averse choice although I_2 is more concave.[6]

8.2 Function G in CVMMM (2011)

In this section, we express the measure of ambiguity aversion in the form of uncertainty-averse utility functions introduced by CVMMM (2011). We first review CVMMM's formulation when the state space S is finite. We then impose the differentiability of functions used in their formulation and provide additional assumptions that hinges on the differentiability.

Given the image $v(T)$ of a Bernoulli utility function $v : T \to \boldsymbol{R}$ (in fact, any open interval of $\boldsymbol{R}$ will do in place of $v(T)$), CVMMM considered a function $G : v(T) \times \Delta \to \boldsymbol{R} \cup \{+\infty\}$ that satisfy the following conditions.

Assumption 3 Lower semi-continuous G is lower semi-continuous, that is, its epigraph $\{(y, p, z) \in v(T) \times \Delta \times \boldsymbol{R} \mid G(y, p) \leq z\}$ is a closed subset of $\boldsymbol{R} \times \Delta \times \boldsymbol{R}$.

Quasi-convex G is quasi-convex.

Non-decreasing in the first coordinate $G(\,\cdot\,, p) : v(T) \to \boldsymbol{R} \cup \{+\infty\}$ is non-decreasing.

Grounded G is grounded, that is, $\inf_{p \in \Delta} G(y, p) = y$ for every $y \in v(T)$.

They then defined $I : v(T)^S \to \boldsymbol{R}$ by letting

$$I(g) = \inf_{p \in \Delta} G(p \cdot g, p) \tag{35}$$

for every $g \in v(T)^S$, and defined an uncertainty-averse utility function $V : T^S \to \boldsymbol{R}$ by letting $V(f) = I(v \circ f)$ for every $f \in T^S$. Conversely, when a normalized function I is given, they defined $G : v(T) \times \Delta \to \boldsymbol{R} \cup \{+\infty\}$ by letting $G(y, p) = \sup \{I(g) \mid g \in v(T)^S \text{ and } p \cdot g \leq y\}$ for every $(y, p) \in v(T) \times \Delta$. If V is quasi-concave, then $V(f) = \inf_{p \in \Delta} G(p \cdot (v \circ f), p)$ for every $f \in T^S$. This means that every quasi-concave V can be written via I in the form $I(g) = \inf_{p \in \Delta} G(p \cdot g, p)$ for some $G(\,\cdot\,, p) : v(T) \to \boldsymbol{R} \cup \{+\infty\}$, and letting $G(y, p) = \sup \{I(g) \mid p \in \Delta \text{ and } p \cdot g \leq y\}$ gives the minimal one among those which satisfy this equality.

[6] Wang (2019, Section 8.2.1) gave an example to make a similar point.

Before stating the conditions we will impose on the gradients and Hessians of G, we should clarify what they represent in terms of real-valued linear functions and quadratic forms, as the domain of G is $v(T) \times \Delta$, where Δ is not an open subset of $\boldsymbol{R}^S$. Denote by $e^\perp$ the hyperplane (in $\boldsymbol{R}^S$) with normal e, that is, $\{q \in \boldsymbol{R}^S \mid q \cdot e = 0\}$. Then, for every $p \in \Delta \cap \boldsymbol{R}^S_{++}$, the tangent space of Δ at p coincides with $e^\perp$. Since each real-valued linear function defined on $e^\perp$ can be represented as a dot product with the unique vector in $e^\perp$, the set of all real-valued linear functions defined on $e^\perp$ can be identified with $e^\perp$. The partial gradient $\nabla_p G(y,p)$, whenever it exists, is a real-valued linear function defined on $e^\perp$ and can thus be identified with a vector in $e^\perp$. The partial Hessian $\nabla_p^2 G(y,p)$, whenever it exists, is a quadratic form on $e^\perp$. According to Hirsch (1976, Section 6.1), this quadratic form is well defined (independent of the choice of a chart of $\Delta \cap \boldsymbol{R}^S_{++}$ around p) and can thus be identified with a quadratic form on $\boldsymbol{R}^S$ (and, hence, an $S \times S$ symmetric matrix) that takes value zero at e.

We impose the following differential conditions on G.

Assumption 4 Twice continuously differentiable $\{(y,p) \in v(T) \times \Delta \mid G(y,p) \in \boldsymbol{R}\}$ is an open subset of $v(T) \times \Delta$ that is included in $v(T) \times \boldsymbol{R}^S_{++}$ and, denoting this set by D, $G : D \to \boldsymbol{R}$ is twice continuously differentiable.

Differentiably strictly increasing in the first coordinate For every $(y,p) \in D$, $\nabla_y G(y,p) > 0$.

Differentiably strictly convex in p For every $(y,p) \in D$, if $\nabla_p G(y,p) = 0$, then the partial Hessian $\nabla_p^2 G(y,p)$ defines a positive-definite quadratic form.

First-order condition For every $g \in v(T)^S$, there is a $p \in \Delta$ such that $(p \cdot g, p) \in D$ and

$$\nabla_y G(p \cdot g, p) \left(g - (e \cdot g) S^{-1} e\right)^\top + \nabla_p G(p \cdot g, p) = 0. \tag{36}$$

The first two conditions are self explanatory. The third one is the differential version of strict convexity of $G(y, \cdot\,)$ around p where $\nabla_p G(y,p) = 0$. Under Assumption 5, to be imposed later, this is equivalent to requiring that the partial Hessian $\nabla_p^2 G(y,p)$ define a positive definite quadratic form when $p = p^I$. In the fourth condition, (36) is the first-order condition of the minimization problem $\inf_{p \in \Delta} G(p \cdot g, p)$, where its left-hand side represents a real-valued linear function on $e^\perp$. The condition guarantees the existence of a solution to the problem. By the strict differentiable quasi-convexity, the solution is unique.

Finally, we impose the following condition. It says that the solution to the minimization problem $\inf_{p \in \Delta} G(y,p)$ does not depend on the choice of $y \in v(T)$. If $g = ye$ for some

$y \in v(T)$, then it implies that the solution $\inf_{p \in \Delta} G(p \cdot g, p)$ coincides with the common solution.

Assumption 5 Common minimum There is a $p^I \in \Delta$ such that $G(y, p^I) \leq G(y, p)$ for every $y \in v(T)$ and every $p \in \Delta$.

When G is grounded, this condition is equivalent to requiring that $G(y, p^I) = y$ for every $y \in v(T)$. We will later see that the common minimum condition on G implies that the condition of common support along the diagonal for I.

Proposition 5 *If I is defined by a $G : v(T) \times \Delta \to \boldsymbol{R} \cup \{\infty\}$ that satisfies all the conditions of Assumptions 4 and 5 via* (35)*, then it satisfies all the conditions of Assumptions 1 and 2 and*

$$H^x(z) = \frac{\dfrac{\left(z - (e \cdot z)S^{-1}e\right)^\top \left((v'(x))^{-1} \nabla_p^2 G\left(v(x), p^I\right)\right)^{-1} \left(z - (e \cdot z)S^{-1}e\right)}{\left(E^{p^I}[z]\right)^2 + \mathrm{Var}^{p^I}[z]}}{-\dfrac{v''(x)}{v'(x)}} \tag{37}$$

for every $x \in T$ and $z \in \boldsymbol{R}^S \setminus \{0\}$.

To understand this proposition, recall that the partial Hessian, $\nabla_p^2 G\left(v(x), p^I\right)$, is a positive-definite quadratic form defined on $e^\perp$. In the numerator on the right-hand side of (37), the positive-definite quadratic form $\nabla_p^2 G\left(v(x), p^I\right)$ is divided by $v'(x)$ and, then, the inverse of the resulting positive-definite quadratic form is taken. Here, the inverse is understood as the positive-definite quadratic form that share the same eigenvectors as the original one and their corresponding eigenvalues are the reciprocals of those of the original one. For every $z \in \boldsymbol{R}^S$, $z - (e \cdot z)S^{-1}e$ is the orthogonal projection of z onto $e^\top$. Thus, the numerator is the value of the inverse of the quadratic form $(v'(x))^{-1} \nabla_p^2 G\left(v(x), p^I\right)$ evaluated at the orthogonal projection of z onto $e^\perp$.

In the numerator of the right-hand side of (37), the partial Hessian $\nabla_p^2 G\left(v(x), p^I\right)$ is divided by the marginal utility $v'(x)$. As shown by Part 2 of Corollary 7, this is to make the numerator invariant to any positive affine transformation of v and associated change in G to represent the same preference relation as V. The denominator of the right-hand side of (37) is equal to the Arrow-Pratt measure $-v''(x)/v'(x)$ of the Bernoulli utility function v. Thus, both the numerator and the denominator of our measure of ambiguity aversion are invariant to any positive affine transformation of the Bernoulli utility function v. The presence of this denominator distinguishes our measure from that of CVMMM (2006, Proposition 6). It allows us to compare ambiguity aversion of two decision makers with different risk attitudes, while their comparison is restricted to two decision makers sharing the same risk attitude. Indeed, they showed, given the same Bernoulli utility function v representing risk attitudes, if a function G_1 represents a more ambiguity-averse preference relation than another function G_2, then $G_1 \leq G_2$. This implies that they share the common minimum p^I, that is, $y = G_n(y, p^I) \leq G_n(y, p)$ for every n, every $y \in v(T)$, and every $p \in \Delta$. Since $G_1(y,p) - G_2(y,p) \leq 0 = G_1(y,p^I) - G_2(y,p^I)$ for every y and p, the second-order condition implies that $\nabla_p^2 G_1(y,p^I) - \nabla_p^2 G_2(y,p^I)$ is negative-semidefinite. This is equivalent to saying that $\left(\nabla_p^2 G_1(y,p^I)\right)^{-1} - \left(\nabla_p^2 G_2(y,p^I)\right)^{-1}$ is positive semidefinite. Thus, the measure $H^x(z)$ determined by G_1 is not smaller than that determined by G_2. This shows that our measure of ambiguity aversion does indeed extend that of CVMMM (2006, Proposition 6) under Assumptions 4 and 5.

8.3 Variational preferences and relative entropy

In this subsection, we consider variational preferences of Maccheroni, Marinacci, and Rustichini as a special case of uncertainty-averse utility functions of CVMMM. We use the same state space and the set T of consequences. They gave necessary and sufficient conditions for a preference relation to be represented in terms of $c : \Delta \to \boldsymbol{R}_+ \cup \{+\infty\}$ that satisfy the following conditions.

Assumption 6 Lower semi-continuous c is lower semi-continuous, that is, its epigraph is a closed subset of $\Delta \times \boldsymbol{R}$.

Convex c is convex.

Grounded c is grounded, that is, $\inf_{p \in \Delta} c(p) = 0$.

They then defined $I : v(T)^S \to \boldsymbol{R}$ by letting

$$I(g) = \inf_{p \in \Delta} (p \cdot g + c(p)) \tag{38}$$

for every $g \in v(T)^S$, and defined an uncertainty-averse utility function $V : T^S \to \boldsymbol{R}$ by letting $V(f) = I(v \circ f)$ for every $f \in T^S$.

We impose the following differential conditions on c.

Assumption 7 Twice continuously differentiable $\{p \in \Delta \mid c(p) \in \boldsymbol{R}\}$ is an open subset of Δ that is included in $\boldsymbol{R}^S_{++}$ and, denoting this set by D, $c : D \to \boldsymbol{R}_+$ is twice continuously differentiable.

Differentiably strictly convex For every $p \in D$, the Hessian $\nabla^2 c(p)$ is a positive definite quadratic form on $e^\perp$.

First-order condition For every $g \in v(T)^S$, there is a $p \in D$ such that $\left(g - (e \cdot g)S^{-1}g\right)^\top + \nabla c(p) = 0$.

Corollary 3 *Suppose that I is defined by a $c : D \to \boldsymbol{R}$ that satisfies all the conditions of Assumptions 6 and 7 via* (38)*. Then I satisfies all the conditions of Assumptions 1 and 2, there is a unique $p \in D$, which we denote by p^I, such that $c(p^I) = 0$, and*

$$H^x(z) = \frac{\dfrac{\left(z - (e \cdot z)S^{-1}e\right)^\top \left(\left(v'(x)\right)^{-1} \nabla^2 c\left(p^I\right)\right)^{-1} \left(z - (e \cdot z)S^{-1}e\right)}{\left(E^{p^I}[z]\right)^2 + \operatorname{Var}^{p^I}[z]}}{-\dfrac{v''(x)}{v'(x)}} \tag{39}$$

for every $x \in T$ and $z \in \boldsymbol{R}^S \setminus \{0\}$.

This result can be seen as a differential version of the more-ambiguity-averse than relation of Maccheroni, Marinacci, and Rustichini (2006, Proposition 8). The least-ambiguity-averse preference among their variational preferences is the ambiguity-neutral one (the expected utility function with the Bernoulli utility function v and the subjective probability p^I), which is represented by the function $c : \Delta \to \boldsymbol{R} \cup \{\infty\}$ such that $c(p) = 0$ if $p = p^I$ and $c(p) = \infty$ otherwise. This function takes the value ∞ and violates the conditions in Assumption 7, but it can be considered as the the most convex function as any deviation from p^I would infinitely increase the value of c. Since this function is better approximated by a more convex differentiable function $c : D \to \boldsymbol{R}$, the corresponding uncertainty-averse utility function V should be less ambiguity-averse, as indicated by the right-hand side of (39).

In the numerator of the right-hand side of (39), the Hessian $\nabla^2 c\left(p^I\right)$ is divided by the marginal utility $v'(x)$. This is to make the numerator invariant to any positive affine transformation of v and associated change in c to represent the same preference relation as V. This fact follows from Part 2 of Corollary 7 and (**??**) in the proof of this proposition. The denominator of the right-hand side of (39) is equal to the Arrow-Pratt measure $-v''(x)/v'(x)$ of the Bernoulli utility function v. Thus, both the numerator and the denominator of our measure of ambiguity aversion are invariant to any positive affine transformation of the Bernoulli utility function v. The presence of this denominator differentiates our measure from that of Maccheroni, Marinacci, and Rustichini (2006, Proposition 8), and it allows us to compare ambiguity aversion of two decision makers with different risk attitudes, while their comparison is restricted to two decision makers sharing the same risk attitude. Indeed, they showed, given the same Bernoulli utility function v representing risk attitudes, if a function c_1 represents a more ambiguity-averse preference relation than another function c_2, then $c_1 \leq c_2$. This implies that they share the common minimum p^I, that is, $0 = c_n(p^I) \leq c_n(p)$ for every n and every $p \in \Delta$. Since $c_1(p) - c_2(p) \leq 0 = c_1(p^I) - c_2(p^I)$ for every y and p, the second-order condition implies that $\nabla^2 c_1(p^I) - \nabla^2 c_2(y, p^I)$ is negative-semidefinite. This is equivalent to saying that $\left(\nabla^2 c_1(p^I)\right)^{-1} - \left(\nabla^2 G_2(p^I)\right)^{-1}$ is positive semidefinite. This shows that our measure of ambiguity aversion does indeed extend that of Maccheroni, Marinacci, and Rustichini (2006, Proposition 8) under Assumptions 7.

Consider two uncertainty-averse utility functions V_1 and V_2 representing variational preferences by (v_1, c_1) and (v_2, c_2). It will be illustrative to present two polar cases of comparison between V_1 and V_2. First, suppose that $v_1' = v_2'$, that is, the attitude towards risk is common between V_1 and V_2. Suppose also that $p^{I_1} = p^{I_2}$, that is, c_1 and c_2 attain the unique minimum 0 at the same probability. Then, $H_1^x(z) \geq H_2^x(z)$ for every $x \in T$ and $z \in \boldsymbol{R}^S \setminus \{0\}$ if and only if $\left(\nabla^2 c_1\left(p^I\right)\right)^{-1} - \left(\nabla^2 c_2\left(p^I\right)\right)^{-1}$ is positive semidefinite. Since this latter condition holds if and only if $\nabla^2 c_2\left(p^I\right) - \nabla^2 c_1\left(p^I\right)$ is positive semidefinite, we can see that V_1 is at least as ambiguity-averse as V_2 if and only if c_1 is at most as convex as c_2. As for the other polar case, we let $x \in T$ and $v_1'(x) = v_2'(x)$. This assumption is without loss of generality as it is imposed on a single consumption level x and can be satisfied by some positive affine transformation of either one of the two Bernoulli utility functions. Assume, in addition, that $c_1 = c_2$. Then $H_1^x(z) \geq H_2^x(z)$ for every $z \in \boldsymbol{R}^S \setminus \{0\}$ if and only if $-v_1''(x) \leq -v_2''(x)$. This holds if and only if $-v_1''(x)/v_1'(x) \leq -v_2''(x)/v_2'(x)$. That is, V_1 is at least as ambiguity-averse as V_2 at x if and only if v_2 is at least as risk-averse as v_1 at x. A similar point on the incomplete separability between risk aversion and ambiguity aversion was made by Guetlein (2016, Section II).

The prime example in the smooth case of variational preferences is the relative entropy, used by Hansen and Sargent. Let $p^I \in \Delta \cap \boldsymbol{R}^S_{++}$ and define the relative entropy of a probability $p \in \Delta \cap \boldsymbol{R}^S_{++}$ with respect to the reference probability p^I is defined by

$$R(p \mid p^I) = \sum_s p(s) \ln \frac{p(s)}{p^I(s)}.$$

Corollary 4 *Let $p^I \in \Delta \cap \boldsymbol{R}^S_{++}$ and $\theta > 0$. Define $c(p) = \theta R(p \mid p^I)$. Define I as in* (38)*. Then c satisfies all the conditions in Proposition 7. Let $v : T \to \boldsymbol{R}$ be the Bernoulli utility functions for risk. Then, for every $x \in T$ and $z \in \boldsymbol{R}^S \setminus \{0\}$,*

$$H^x(z) = -\frac{(v'(x))^2}{\theta v''(x)} \frac{\|z - (e \cdot z)S^{-1}e\|_{p^I}^2}{\|z\|_{p^I}^2}. \tag{40}$$

This corollary characterizes $H^x(z)$ in the case of relative entropy. If $z \in e^{\perp}$, then

$$H^x(z) = -\frac{(v'(x))^2}{\theta v''(x)}, \tag{41}$$

that is, $H^x(z)$ is independent of p^I.

Now compare two utility functions derived from the relative entropy. Let $p^{I_1} \in \Delta \cap \boldsymbol{R}^S_{++}$, $p^{I_2} \in \Delta \cap \boldsymbol{R}^S_{++}$, $\theta_1 > 0$, and $\theta_2 > 0$. Define $c_1(p) = \theta_1 R(p \mid p^{I_1})$ and $c_2(p) = \theta_2 R(p \mid p^{I_2})$. Define I_1 and I_2 by (38). Let $v_1 : T_1 \to \boldsymbol{R}$ and $v_2 : T_2 \to \boldsymbol{R}$ be the Bernoulli utility functions for risk. We assume that $T_1 = T_2$ and denote this common open interval by T. Then define $V_1 : T^S \to \boldsymbol{R}$ and $V_2 : T^S \to \boldsymbol{R}$ by letting $V_1(f) = I_1(v_1 \circ f)$ and $V_2(f) = I_2(v_2 \circ f)$ for every $f \in T^S$. The following corollary can be immediately derived from (40) and (41).

Corollary 5 *1. Suppose that $v_1 = v_2$, $p^{I_1} = p^{I_2}$. Then $\theta_1 < \theta_2$ if and only if $H^x_1(z) > H^x_2(z)$ for every $x \in T$ and every $z \in \boldsymbol{R}^S \setminus \{0\}$, that is, V_1 is more ambiguity-averse than V_2 at every consumption level in every direction.*

2. Suppose that $v_1 = v_2$. Then $\theta_1 < \theta_2$ if and only if $H^x_1(z) > H^x_2(z)$ for every $x \in T$ and every $z \in e^{\perp}$, that is, V_1 is more ambiguity-averse than V_2 at every consumption level in every direction whose coordinates add up to zero.

Part 1 of this corollary states that when two utility functions share the same Bernoulli utility function for risk and the same reference probability, one is more ambiguity-averse than the other at every consumption level in every direction if and only if the former has a higher penalty intensity for the relative entropy than the latter. Part 2 states that when two utility functions share the same Bernoulli utility function, one is more ambiguity-averse than the other at every consumption level in every direction whose coordinates add up to zero if and only if the former has a higher penalty intensity for the relative entropy than the latter. The two parts are different in whether they assume the common reference probability and in which direction they claim one utility function is more ambiguity-averse than the other.

Our measure of ambiguity aversion hinges of the twice continuous differentiability of the function I, but otherwise imposes no restriction on its functional form. This fact allows us to compare ambiguity aversion between two utility functions belonging to different classes of functional forms. The following proposition shows, indeed, that every relative-entropy utility function is as ambiguity-averse as some KMM utility function. A similar equivalence was established by CVMMM (2011, Corollary 22) via the function G in their formulation.

Proposition 6 *Let $p^I \in \Delta \cap \boldsymbol{R}^S_{++}$ and $\theta > 0$. Define $c : \Delta \to \boldsymbol{R}_+$ by letting $c(p) = \theta R(p \mid p^I)$ for every $p \in \Delta$. Define $I_1 : \boldsymbol{R}^S \to \boldsymbol{R}$ as in (38). Let $\varepsilon > 0$ and μ be a probability measure on Δ. Suppose that*

$$\int_\Delta p \, \mathrm{d}\mu(p) = p^I, \tag{42}$$

$$\int_\Delta \left(p - p^I\right) \left(p - p^I\right)^\top \mathrm{d}\mu(p) = \varepsilon^2 \left(I_{S\times S} - \frac{1}{S} ee^\top\right) \left[p^I\right] \left(I_{S\times S} - \frac{1}{S} ee^\top\right). \tag{43}$$

Define $\phi : \boldsymbol{R} \to \boldsymbol{R}$ by letting

$$\phi(y) = -\exp\left(-\frac{1}{\theta\varepsilon^2} y\right)$$

for every $y \in \boldsymbol{R}$. Define I_2 as in (27). Let $v : T \to \boldsymbol{R}$ and, for each $n = 1, 2$, define $V_n : T^S \to \boldsymbol{R}$ by letting $V_n(f) = I_n(v \circ f)$ for every $f \in T^S$. Then $H_1^x(z) = H_2^x(z)$ for every $x \in T$ and $z \in \boldsymbol{R}^S$.

Note that ϕ exhibits constant ambiguity aversion in the sense of Definition 6 of KMM. While the proposition does not explicitly claim this, there is indeed a second-order belief μ that satisfies (42) and (43) for a sufficiently small $\varepsilon > 0$. To see this, let $Z = (Z_1, \ldots, Z_S)$ be an S-variate random vector on some probability space such that for each s, Z_s takes value $(p^I(s))^{1/2}$ or $-(p^I(s))^{1/2}$ with probability 1/2 each, and Z_s and Z_t are independent whenever $s \neq t$. Then $E[Z] = 0$ and $\operatorname{Cov}[Z] = \left[p^I\right]$. Then, for each $\varepsilon > 0$, define an S-variate random vector Z^ε by

$$Z^\varepsilon = \varepsilon \left(I_{S\times S} - \frac{1}{S} ee^\top\right) Z + p^I,$$

then $e^\top Z^\varepsilon = 1$, that is, Z^ε always takes a value in $e^\perp + \{S^{-1}e\}$. Moreover, $E\left[Z^\varepsilon\right] = p^I$ and

$$\operatorname{Cov}\left[Z^\varepsilon\right] = \varepsilon^2 \left(I_{S\times S} - \frac{1}{S} ee^\top\right) \left[p^I\right] \left(I_{S\times S} - \frac{1}{S} ee^\top\right).$$

Since $p^I \in \boldsymbol{R}^S_{++}$, if $\varepsilon > 0$ is sufficiently small, then Z^ε always lies in $\in \boldsymbol{R}^S_{++}$. Let μ be the distribution of Z^ε for a sufficiently small $\varepsilon > 0$, then μ is a probability measure on $\Delta \cap \boldsymbol{R}^S_{++}$ that satisfies (42) and (43) for a sufficiently small $\varepsilon > 0$.

8.4 Confidence functions

In this subsection, we consider the class of ambiguity-averse utility functions introduced by Chateuneuf and Faro (2009) via confidence functions. Let the finite state space $\{1, 2, \ldots, S\}$, the nonempty open interval T of consumption levels, and a Bernoulli utility function $v : T \to \boldsymbol{R}$ for risks be as in Section 2. They assumed that there is a worst outcome in the set of consequences, but our model cannot really satisfy this assumption because our set of consequences, T, is an open interval of $\boldsymbol{R}$. We assume, instead, that the range $v(T)$ is bounded from below and, without loss of generality, that $\inf\, v(T) = 0$. This assumption is compatible with Inada condition. For example, it is satisfied if $T = \boldsymbol{R}_{++}$ and $v(x) = x^{1/2}$ for every $x \in \boldsymbol{R}_{++}$. Nonetheless, we should keep in mind that it restricts the risk attitude that v represents. For example, it excludes the case where $T = \boldsymbol{R}_{++}$ and v is the logarithmic Bernoulli utility function. More generally, when $\inf T = 0$ and the range $v(T)$ is bounded from below, then the relative risk aversion at sufficiently low consumption levels must be less than one.

Chateauneuf and Faro (2009) gave necessary and sufficient conditions for a preference relation to be represented in terms of $\phi : \Delta \to [0, 1]$ that satisfy the following conditions.

Assumption 8 Upper semi-continuous ϕ is upper semi-continuous, that is, its hypograph $\{(p, z) \in \Delta \times \boldsymbol{R} \mid \phi(p) \geq z\}$ is a closed subset of $\Delta \times \boldsymbol{R}$.

Concave ϕ is concave.

Normal ϕ is normal, that is, $\sup_{p \in \Delta} \phi(p) = 1$.

They then defined $I : v(T)^S \to \boldsymbol{R}$ by letting

$$I(g) = \inf_{p \in \Delta} \frac{p \cdot g}{\phi(p)} \tag{44}$$

for every $g \in v(T)^S$. Note that $p \cdot g/\phi(p) = +\infty$ whenever $\phi(p) = 0$, because $v(T)^S \subseteq \boldsymbol{R}^S_{++}$. They then defined an uncertainty-averse utility function $V : T^S \to \boldsymbol{R}$ by letting $V(f) = I(v \circ f)$ for every $f \in T^S$.

To be precise, they only required ϕ to be quasi-concave, not necessarily concave. However, it is easy to show that the concavification of ϕ, that is, the smallest concave function that dominates ϕ, gives rise to the same I and, hence, V. The convexitifcation coincides with what they called, in Corollary 5 of their paper, the maximal confidence function ϕ^*, which can be written, in our notation, as

$$\phi^*(g) = \inf_{g \in v(T)^S} \frac{p \cdot g}{I(g)}$$

for every $g \in v(T)^S$. Since their characterization of comparative ambiguity aversion (Proposition 8 of their paper) is given in terms of maximal confidence functions, it is appropriate in our analysis to assume that the confidence function ϕ is not only quasi-concave but also concave.

We impose the following differential conditions on ϕ.

Assumption 9 **Twice continuously differentiable** $\{p \in \Delta \mid \phi(p) > 0\}$ is an open subset of Δ that is included in $\boldsymbol{R}^S_{++}$ and, denoting this set by D, $\phi : D \to (0,1]$ is twice continuously differentiable.

Differentiably strictly concave at the maximum For every $p \in D$, if $\nabla\phi(p) = 0$, then the Hessian $\nabla^2\phi(p)$ defines a negative definite quadratic form on $e^{\perp}$.

First-order condition For every $g \in v(T)^S$, there is a $p \in D$ such that

$$\frac{1}{p \cdot g}\left(g - (e \cdot g)S^{-1}e\right)^{\top} - \frac{1}{\phi(p)}\nabla\phi(p) = 0. \tag{45}$$

Corollary 6 *Suppose that I is defined by a $\phi : D \to \boldsymbol{R}$ that satisfies all the conditions of Assumptions 8 and 9 via* (44)*. Then I satisfies all the conditions of Assumptions 1 and 2, there is a unique $p \in D$, which we denote by p^I, such that $\phi(p^I) = 0$, and*

$$H^x(z) = \frac{-\dfrac{(z - (e \cdot z)S^{-1}e)^{\top}\left((v'(x))^{-1}v(x)\nabla^2\phi\left(p^I\right)\right)^{-1}(z - (e \cdot z)S^{-1}e)}{\left(E^{p^I}[z]\right)^2 + \operatorname{Var}^{p^I}[z]}}{-\dfrac{v''(x)}{v'(x)}} \tag{46}$$

for every $x \in T$ and $z \in \boldsymbol{R}^S \setminus \{0\}$.

9 Conclusion

We have introduce a measure of ambiguity aversion that is applicable to twice continuously differentiable uncertainty-averse utility functions. Although it is defined in terms of the Hessians of utility functions, it has implications in binary choice behaviors (observable, in principle, in laboratory experiments) and also in asset prices and portfolio choices (observable, in principle, in asset markets). The measure have the form that can be easily interpreted when the utility function belongs to the class of the smooth ambiguity model, relative entropy, or confidence functions.

There are a couple of directions of future research. The first one is to extend our analysis to the case of infinitely many states. This is important because, as we did for the anatomy of the equity premium puzzle in Section 5, some numerical analysis uses continuous random variables such as log-normally distributed ones. Since the definition of our measure is based on quadratic forms, the space of acts should then be a subset of square-integrable random variables (random variables with finite variances) as in Maccheroni, Marinacci, and Ruffino (2013). The second one is to exend our analysis to a dynamic model. This is important as macroeconomic and financial applications are given in dynamic models. To ensure dynamic consistency, the measure of ambiguity aversion should then be defined for continuation utilities. The third one is to extend our analysis to the case where two uncertainty-averse utility functions are defined on different state spaces with different numbers of elements (states). It would then be meaningless to compare the measures of their ambiguity aversion in a common direction z, but it might still be meaningful to compare the largest and smallest measures, $\max_{z_1 \in \boldsymbol{R}^{S_1} \setminus \{0\}} H_1^x(z_1)$, $\max_{z_2 \in \boldsymbol{R}^{S_2} \setminus \{0\}} H_2^x(z_2)$, $\min_{z_1 \in e_1^{\perp} \setminus \{0\}} H_1^x(z_1)$, and $\min_{z_2 \in e_2^{\perp} \setminus \{0\}} H_2^x(z_2)$, at a common consumption level x, where e_n is the vector of 1's of dimension S_n. For example, we could say that V_1 is more ambiguity-averse than V_2 if $\min_{z_1 \in e_1^{\perp} \setminus \{0\}} H_1^x(z_1) > \max_{z_2 \in \boldsymbol{R}^{S_2} \setminus \{0\}} H_2^x(z_2)$, although this definition is much stronger than our original definition because the former requires the uniformity with respect to z_1 and z_2.

A Preliminary results on derivatives

For each $z \in \boldsymbol{R}^S$, denote by $[z]$ the $S \times S$ matrix of which the s-th diagonal entry is equal to z_s and the off-diagonal entries are all equal to zero. By the chain rule differentiation,

$$\frac{\partial V}{\partial f(s)}(f) = \frac{\partial I}{\partial y_s}(v \circ f) v'(f(s))$$

for every s and f. In the vector notation,

$$\nabla V(f) = \nabla I(v \circ f)[v' \circ f], \tag{47}$$

where the gradients are row vectors. Thus,

$$\frac{\partial^2 V}{\partial f(s') \partial f(s)}(f) = \begin{cases} \dfrac{\partial^2 I}{\partial y_{s'} \partial y_s}(v \circ f)\left(v'(f(s))\right)^2 + \dfrac{\partial I(v \circ f)}{\partial y_s} v''(f(s)) & \text{if } s' = s, \\ \dfrac{\partial^2 I}{\partial y_{s'} \partial y_s}(v \circ f) v'(f(s')) v'(f(s)) & \text{otherwise,} \end{cases}$$

for every s and s'. In the vector notation,

$$\nabla^2 V(f) = [v' \circ f] \nabla^2 I(v \circ f)[v' \circ f] + [\nabla I(v \circ f)][v'' \circ f]. \tag{48}$$

Thus, if $f = xe$ for some $x \in T$, then $v \circ f = v(x)e$, $v' \circ f = v'(x)e$, and $\nabla V(f) = p^I$. Since $[e]$ coincides with the $I \times I$ identity matrix,

$$\nabla V(xe) = v'(x)(p^I)^\top, \tag{49}$$

$$\nabla^2 V(xe) = (v'(x))^2 \nabla^2 I(v(x)e) + v''(x)[p^I]. \tag{50}$$

The following lemma extends the uniqueness part of Proposition 1 of Cerreia-Vioglio, Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2011) to the case where the consequences are not lotteries but real numbers.

Lemma 3 *Let $v_1 : T \to \boldsymbol{R}$ and $v_2 : T \to \boldsymbol{R}$ and suppose that they are strictly increasing. Let $I_1 : v_1(T)^S \to \boldsymbol{R}$ and $I_2 : v_2(T)^S \to \boldsymbol{R}$ and suppose that I_1 is normalized. Define $V_1 : T^S \to \boldsymbol{R}$ and $V_2 : T^S \to \boldsymbol{R}$ by letting $V_1(f) = I_1(v_1 \circ f)$ and $V_2(f) = I_2(v_2 \circ f)$ for every $f \in T^S$. Write $\psi = v_2 \circ v_1^{-1} : v_1(T) \to v_2(T)$. Let $\Psi : V_1\left(T^S\right) \to \boldsymbol{R}$.*

1. *$V_2 = \Psi \circ V_1$ if and only if $I_2 = \Psi \circ I_1 \circ \psi^{-1}$ on $v_2(T)^S$, that is, $I_2(z) = \Psi\left(I_1\left(\psi^{-1} \circ z\right)\right)$ for every $z \in v_2(T)^S$.*

2. *Suppose that $V_2 = \Psi \circ V_1$. Then I_2 is normalized if and only if $\psi = \Psi$ on $v_1(T)$.*

Proof of Lemma 3

1. For every $f \in T^S$, $V_2(f) = I_2(v_2 \circ f)$ and

$$(\Psi \circ V_1)(f) = \Psi(I_1(v_1 \circ f)) = \Psi\left(I_2\left(\left(\psi^{-1} \circ v_2\right) \circ f\right)\right) = \Psi\left(I_2\left(\psi^{-1} \circ (v_2 \circ f)\right)\right)$$

   Thus, $V_2(f) = (\Psi \circ V_1)(f)$ if and only if $I_2(v_2 \circ f) = \Psi\left(I_2\left(\psi^{-1} \circ (v_2 \circ f)\right)\right)$. Thus, $V_2 = \Psi \circ V_1$ if and only if $I_2(z) = \Psi\left(I_1\left(\psi^{-1} \circ z\right)\right)$ for every $z \in v_2(T)^S$.

2. Let $y \in v_2(T)$. Then there is an $x \in T$ such that $y = v_2(x)$. Thus

$$\begin{aligned} I_2(ye) &= I_2(v_2(x)e) = I_2(v_2 \circ (xe)) = V_2(xe) \\ &= \Psi(V_1(xe)) = \Psi(I_1(v_1 \circ (xe))) = \Psi(I_1(v_1(x)e)) = \Psi(v_1(x)) = \Psi(\psi^{-1}(y)). \end{aligned}$$

   Thus, I_2 is normalized if and only if $\Psi(\psi^{-1}(y)) = y$ for every $y \in v_2(T)$, which is, in turn, equivalent to saying that $\psi = \Psi$ on $v_1(T)$.

///

This lemma gives the condition under which two utility functions of the form $V_1(f) = I_1(v_1 \circ f)$ and $V_2(f) = I_2(v_2 \circ f)$, with I_1 and I_2 being normalized, represent the same preference relation on T^S. It is, in a nutshell, that the transformation ψ of v_1 to v_2 allows us to write I_2 in terms of I_1 as $I_2 = \psi \circ I_1 \circ \psi^{-1}$. In particular, we are interested in the case where ψ is affine, because v_1 and v_2 represent the same risk attitude if and only if ψ is affine. The lemma then implies that the transformation Ψ of V_1 to V_2 is also affine.

Implications of this equivalence on the gradients and Hessians of I_1 and I_2 are given in the following corollary.

Corollary 7 *Let $v_1 : T \to \boldsymbol{R}$ and $v_2 : T \to \boldsymbol{R}$ and suppose that they are continuously differentiable satisfying $v_1' > 0$ and $v_2' > 0$. Let $I_1 : v_1(T)^S \to \boldsymbol{R}$ and $I_2 : v_2(T)^S \to \boldsymbol{R}$ and suppose that they are normalized. Define $V_1 : T^S \to \boldsymbol{R}$ and $V_2 : T^S \to \boldsymbol{R}$ by letting $V_1(f) = I_1(v_1 \circ f)$ and $V_2(f) = I_2(v_2 \circ f)$ for every $f \in T^S$. Write $\psi = v_2 \circ v_1^{-1} : v_1(T) \to v_2(T)$. Suppose that there is a $\Psi : V_1\left(T^S\right) \to \boldsymbol{R}$ such that $V_2 = \Psi \circ V_1$.*

1. *If I_1, and I_2 are differentiable, then $\nabla I_1(v_1(x)e) = \nabla I_2(v_2(x)e)$ for every $x \in T$.*

2. *If I_1 and I_2 are twice continuously differentiable and ψ is affine, then $v_1'(x)\nabla^2 I_1(v_1(x)e) = v_2'(x)\nabla^2 I_2(v_2(x)e)$ for every $x \in T$.*

Part 2 shows that the Hessian of I multiplied by the marginal utility v' is invariant to the change in utility representation. This fact will help us understand the nature of the measure of ambiguity aversion to be introduced later. It would not hold without the assumption of the affinity of ψ.

Proof of Corollary 7

1. By Lemma 3, $I_2(z) = \psi\left(I_1\left(\psi^{-1}(z)\right)\right)$ for every $z \in T^S$. By differentiating both sides with respect to z, we obtain

$$\nabla I_2(z) = \psi'\left(I_1\left(\psi^{-1} \circ z\right)\right) \nabla I_1(\psi^{-1} \circ z) \left[\left(\psi^{-1}\right)' \circ z\right]. \tag{51}$$

   Let $x \in T$. By plugging $z = v_2(x)e$ into this equality, we obtain

$$\begin{aligned}\nabla I_2(v_2(x)e) &= \psi'\left(v_1(x)\right) \nabla I_1(v_1(x)e) \left[\left(\psi^{-1}\right)' \circ \left(v_2(x)\right)\right] \\ &= \psi'\left(v_1(x)\right) \left(\psi^{-1}\right)' \circ (v_2(x)e) \nabla I_1(v_1(x)e)\end{aligned}$$

   Since $\psi'\left(v_1(x)\right)\left(\psi^{-1}\right)' \circ (v_2(x)e) = 1$, this completes the proof.

2. Since ψ is affine, ψ' is constant. Denote it by κ. Then $\kappa = v_2'(x)/v_1'(x)$ for every $x \in T$ and, by (51),

$$\nabla I_2(z) = \kappa \nabla I_1(\psi^{-1} \circ z) \left[\kappa^{-1} e\right] = \nabla I_1(\psi^{-1} \circ z)$$

for every $z \in T^S$. By differentiating the right- and left-hand sides with respect to z, we obtain

$$\nabla^2 I_2(z) = \nabla^2 I_1(\psi^{-1} \circ z) \left[(\psi^{-1})' \circ z\right].$$

Let $x \in T$ and $z = v_2(x)e$, then $\psi^{-1} \circ z = v_1(x)e$ and $(\psi^{-1})' \circ z = \kappa^{-1} e = (v_1'(x)/v_2'(x))\, e$. Thus

$$\nabla^2 I_2(v(x)e) = \frac{v_1'(x)}{v_2'(x)} \nabla^2 I_1(v_(x)e).$$

The proof is completed by multiplying $v_2'(x)$ to both sides.

///

B Lemma on uncertainty and risk premiums

The following lemma gives an equivalent condition of Definition 3 in terms of the risk and uncertainty premiums.

Lemma 4 *Assume that V_1 and V_2 are as above. Assume that $T_1 = T_2$ and denote them by T. Then V_1 is at least as ambiguity-averse as V_2 in the ranking sense if and only if, for every $f \in T^S$ and every $y \in T$, $V_2(f) \geq V_2(ye)$ whenever $V_1(f) = V_1(ye)$, and $\bar{V}_2(f) \leq \bar{V}_2(ye)$ whenever $\bar{V}_1(f) = \bar{V}_1(ye)$.*

Proof of Lemma 4 The Only-If part follows from the definition of the at-least-as-ambiguity-as relation in the ranking sense and noting that $V_n(ye) = \bar{V}_n(ye)$.

To prove the If part, it suffices to show that for every $f \in T^S$ and every $g \in T^S$, $V_2(f) \geq \bar{V}_2(g)$ whenever $V_1(f) = \bar{V}_1(g)$. Indeed, if $V_1(f) > \bar{V}_1(g)$, then let $y \in T$ satisfy $\bar{V}_1(ye) = \bar{V}_1(g)$. Write as $f \wedge ye$ the vector of $\boldsymbol{R}^S$ whose s-th coordinate is equal to $\min\{f(s), y\}$. Since $f \wedge ye \in T^S$ and T is an open interval of $\boldsymbol{R}$, there is a $\delta > 0$ such that $f \wedge ye - \delta e \in T^S$. For each $\lambda \in [0, 1]$, define

$$h(\lambda) = (1 - \lambda)((f \wedge ye) - \delta e) + \lambda f \in T^S.$$

Then $h(0) = (f \wedge ye) - \delta e$, $h(1) = f$, $ye - h(0) \in \boldsymbol{R}^S_{++}$, and $f - h(\lambda) \in \boldsymbol{R}^S_{++}$ for every $\lambda \in [0, 1)$. Thus,

$$V_1(h(0)) < V_1(ye) = \bar{V}_1(ye) = \bar{V}_1(g) < V_1(h(1)).$$

Hence, there is a $\lambda \in (0,1)$ such that $V_1(h(\lambda)) = \bar{V}_1(g)$. Thus, by the hypothesis, $V_1(h(\lambda)) = \bar{V}_1(g)$. Since $f - h(\lambda) \in \boldsymbol{R}^S_{++}$, $V_1(f) > \bar{V}_1(g)$.

Now, suppose that $V_1(f) = \bar{V}_1(g)$. Let $x \in T$ satisfy $V_1(xe) = V_1(f)$. Then, $V_1(f) = \bar{V}_1(xe)$ and $V_1(xe) = \bar{V}_1(g)$. By the hypothesis, $V_2(f) \geq \bar{V}_2(xe)$ and $V_2(xe) \geq \bar{V}_2(g)$. Since $V_1(xe) = \bar{V}(xe)$, $V_2(f) \geq \bar{V}_2(g)$. ///

C The at-least-as-ambiguity-averse-as relation in the ranking sense

As we have repeatedly mentioned, according to the standard definition of comparative ambiguity aversion by Ghirardato and Marinacci (2002) and others, if V_1 is at least as ambiguity-averse as V_2, then the Bernoulli utility function v_1 is just as risk-averse as v_2. In the next lemma, we show that according to the definition of comparative ambiguity aversion in the ranking sense (Definition 3), v_1 is at most as risk-averse as v_2 if the probability implicit at the constant acts is not degenerate; and the former is just as risk-averse as the latter if there is an unambiguous and, yet, non-constant act. This substantiates our claim, stated just before Definition 3, that using the fictitious utility function $\bar{V}$ in the definition serves essentially for the same purpose as expanding the domain of the original utility function V to a set of Anscombe-Aumann acts.

Lemma 5 *Assume that V_1 and V_2 are as above. Assume that $T_1 = T_2$ and $p^{I_1} = p^{I_2}$. Denote the probabilities by p. Assume that there is a subset R of S such that $0 < p(R) < 1$. Suppose that V_1 is at least as ambiguity-averse as V_2 in the ranking sense (Definition 3). Then:*

1. *v_1 is at most as risk-averse as v_2.*
2. *Assume, in addition, that there is a subset R of S such that $0 < p(R) < 1$ and for each n and for every act f, if f is measurable with respect to the partition $(R, S \setminus R)$, then $V_n(f) = E^p[v_n \circ f]$. Then v_1 is just as risk-averse as v_n.*

An act f is measurable with respect to the partition $(R, S \setminus R)$ if and only if it is constant on R and $S \setminus R$. Thus, if we denote the consequences on R and $S \setminus R$ by x and y, then $E^p[v_n \circ f] = p(R)v_n(x) + (1 - p(R))\, v_n(y)$. This assumption means that both decision makers sees event R and its complement $S \setminus R$ as unambiguous and, thus, have expected utility functions for acts that are constant on R and $S \setminus R$.

Proof of Lemma 5 Write $\varphi = v_2 \circ v_1^{-1}$.

1. Since v_1 is at most as risk-averse as v_2 if and only if φ is concave, it suffices to prove that for all x, y, and z, if

$$p(R)v_1(x) + (1 - p(R))v_1(y) = v_1(z), \tag{52}$$

then

$$p(R)v_2(x) + (1 - p(R))v_2(y) \leq v_2(z). \tag{53}$$

Indeed, then,

$$\varphi\left(p(R)v_1(x) + (1 - p(R))v_1(y)\right) = v_2(z) \geq p(R)\varphi\left(v_1(x)\right) + (1 - p(R))\varphi\left(v_1(y)\right),$$

and Theorem 88 of Hardy, Littlewood, and Polya (1952) is applicable because $0 < p(R) < 1$. Define two acts f and g by letting $f = ze$ and

$$g(s) = \begin{cases} x & \text{if } s \in R \\ y & \text{otherwise.} \end{cases}$$

Then (52) holds if and only if $V_1(f) = \bar{V}_1(g)$, and (53) holds if and only if $V_2(f) \geq \bar{V}_2(g)$. Since V_1 is at least as ambiguity-averse as V_2 in the ranking sense, (53) holds whenever (52) holds.

2. It suffices to prove that v_1 is at least as risk-averse as v_2. Since this is true if and only if φ is convex, it suffices to prove that for all x, y, and z, if

$$p(R)v_1(x) + (1 - p(R))v_1(y) = v_1(z), \tag{54}$$

then

$$p(R)v_2(x) + (1 - p(R))v_2(y) \geq v_2(z). \tag{55}$$

Indeed, then,

$$\varphi\left(p(R)v_1(x) + (1 - p(R))v_1(y)\right) = v_2(z) \leq p(R)\varphi\left(v_1(x)\right) + (1 - p(R))\varphi\left(v_1(y)\right),$$

and Theorem 88 of Hardy, Littlewood, and Polya (1952) is applicable because $0 < p(R) <$

1. Define two acts f and g by letting

$$f(s) = \begin{cases} x & \text{if } s \in R \\ y & \text{otherwise,} \end{cases}$$

and $g = ze$. By the assumption on the partition $(R, S \setminus R)$, (54) holds if and only if $V_1(f) = \bar{V}_1(g)$, and (55) holds if and only if $V_2(f) \geq \bar{V}_2(g)$. Since V_1 is at least as ambiguity-averse as V_2 in the ranking sense, (55) holds whenever (54) holds. ///

D Other Proofs

Proof of Theorem 1 By (7),

$$\kappa^x(0, z) = 0. \tag{56}$$

Since v and I are twice continuously differentiable, so is V. The gradient of the function $(k, \varepsilon) \mapsto V((x-k)e) - V(xe + \varepsilon z)$ is equal to $(-\nabla V((x-k)e)e, -\nabla V(xe+\varepsilon z)z)$. Since $\nabla V((x-k)e)e = v'(x-k) > 0$, the implicit function theorem implies that $\kappa^x(\varepsilon, z)$ is a twice continuously differentiable function of ε. Moreover, by differentiating the left-hand side of (7) with respect to ε, we obtain

$$\begin{aligned} 0 &= -\frac{\partial \kappa^x}{\partial \varepsilon}(\varepsilon, z) \nabla V\left(\left(x - \kappa^x(\varepsilon, z)\right) e\right) e - \nabla V(xe + \varepsilon z) z \\ &= -\frac{\partial \kappa^x}{\partial \varepsilon}(\varepsilon, z) v'\left(x - \kappa^x(\varepsilon, z)\right) - \nabla V(xe + \varepsilon z) z. \end{aligned} \tag{57}$$

If $\varepsilon = 0$, then

$$-\frac{\partial \kappa^x}{\partial \varepsilon}(\varepsilon, z) v'\left(x - \kappa^x(\varepsilon, z)\right) - \nabla V(xe + \varepsilon z) z = -\frac{\partial \kappa^x}{\partial \varepsilon}(0, z) v'\left(x\right). \tag{58}$$

Since this is equal to zero by (7),

$$-\frac{\partial \kappa^x}{\partial \varepsilon}(0, z) = 0.$$

By differentiating the left-hand side of (57), we obtain

$$-\frac{\partial^2 \kappa^x}{\partial \varepsilon^2}(\varepsilon, z) v'\left(x - \kappa^x(\varepsilon, z)\right) + \left(\frac{\partial \kappa^x}{\partial \varepsilon}(\varepsilon, z)\right)^2 v''\left(x - \kappa^x(\varepsilon, z)\right) - z^\top \nabla V^2(xe + \varepsilon z) z.$$

If $\varepsilon = 0$, then by (56) and (58),

$$
\begin{aligned}
&-\frac{\partial^2 \kappa^x}{\partial \varepsilon^2}(\varepsilon, z) v'(x - \kappa^x(\varepsilon, z)) + \left(\frac{\partial \kappa^x}{\partial \varepsilon}(\varepsilon, z)\right)^2 v''(x - \kappa^x(\varepsilon, z)) - z^\top \nabla^2 V(xe + \varepsilon z) z \\
=& -\frac{\partial^2 \kappa^x}{\partial \varepsilon^2}(0, z) v'(x) - z^\top \nabla^2 V(xe) z.
\end{aligned}
$$

Since this is equal to zero by (7),

$$\frac{\partial^2 \kappa^x}{\partial \varepsilon^2}(0, z) = -\frac{1}{v'(x)} z^\top \nabla^2 V(xe) z.$$

Analogously,

$$\frac{\partial^2 \bar{\kappa}^x}{\partial \varepsilon^2}(0, z) = -\frac{1}{v'(x)} z^\top \nabla^2 \bar{V}(xe) z.$$

Thus,

$$\frac{\dfrac{\partial^2 \kappa^x}{\partial \varepsilon^2}(0, z) - \dfrac{\partial^2 \bar{\kappa}^x}{\partial \varepsilon^2}(0, z)}{\dfrac{\partial^2 \bar{\kappa}^x}{\partial \varepsilon^2}(0, z)} = H^x(z).$$

Since $\dfrac{\partial \kappa^x}{\partial \varepsilon}(\varepsilon, z) \to \dfrac{\partial \kappa^x}{\partial \varepsilon}(0, z) = 0$ and $\dfrac{\partial \bar{\kappa}^x}{\partial \varepsilon}(\varepsilon, z) \to \dfrac{\partial \bar{\kappa}^x}{\partial \varepsilon}(0, z) = 0$, by L'Hôpital's rule,

$$\frac{\dfrac{\partial \kappa^x}{\partial \varepsilon}(\varepsilon, z) - \dfrac{\partial \bar{\kappa}^x}{\partial \varepsilon}(\varepsilon, z)}{\dfrac{\partial \bar{\kappa}^x}{\partial \varepsilon}(\varepsilon, z)} \to H^x(z)$$

as $\varepsilon \to 0$. Since $\kappa^x(\varepsilon, z) \to \kappa^x(0, z) = 0$ and $\bar{\kappa}^x(\varepsilon, z) \to \bar{\kappa}^x(0, z) = 0$, again by L'Hôpital's rule,

$$\frac{\kappa^x(\varepsilon, z) - \bar{\kappa}^x(\varepsilon, z)}{\bar{\kappa}^x(\varepsilon, z)} \to H^x(z)$$

as $\varepsilon \to 0$. ///

Proof of Corollary 1 1. Suppose that V_1 is at least as ambiguity-averse as V_2 at x in the direction of z. Then (12) holds. As $\varepsilon \to 0$, by Theorem 1, its left- and right-hand sides converge to $H_1^x(z)$ and $H_2^x(z)$. Thus, $H_1^x(z) \geq H_2^x(z)$.

2. The Only-If part can be proved just like part 1, only by using $h_1 \geq h_2 + \delta$ in place of $h_1 \geq h_2$. To prove the If part, suppose that $H_1^x(z) > H_2^x(z)$. Let h_1 and h_2 satisfy

$H_1^x(z) > h_1 > h_2 > H_2^x(z)$ and $\delta = h_1 - h_2$. Theorem 1 implies that for every sufficiently small $\varepsilon > 0$, $\bar{\kappa}_1(\varepsilon, z)$ and $\bar{\kappa}_2(\varepsilon, z)$ are well defined and

$$\frac{\kappa_1^x(\varepsilon, z) - \bar{\kappa}_1^x(\varepsilon, z)}{\bar{\kappa}_1^x(\varepsilon, z)} > h_1,$$
$$\frac{\kappa_2^x(\varepsilon, z) - \bar{\kappa}_2^x(\varepsilon, z)}{\bar{\kappa}_2^x(\varepsilon, z)} < h_2.$$

Write $\bar{k}_1 = \bar{\kappa}_1^x(\varepsilon, z)$ and $\bar{k}_2 = \bar{\kappa}_1^x(\varepsilon, z)$. Then

$$\bar{V}_1\left(\left(x - \bar{k}_1\right)e\right) = \bar{V}_1\left(xe + \varepsilon z\right),$$
$$\bar{V}_2\left(\left(x - \bar{k}_2\right)e\right) = \bar{V}_2\left(xe + \varepsilon z\right).$$

Moreover, $\kappa_1(\varepsilon, z) > \bar{k}_1 + h_1\bar{k}_1$ and $\kappa_2(\varepsilon, z) < \bar{k}_2 + h_2\bar{k}_2$. Thus,

$$V_1\left(\left(x - \bar{k}_1 - h_1\bar{k}_1\right)e\right) > V_1\left(xe + \varepsilon z\right),$$
$$V_2\left(\left(x - \bar{k}_2 - h_2\bar{k}_2\right)e\right) < V_2\left(xe + \varepsilon z\right).$$

Thus, V_1 is more ambiguity-averse than V_2 at x in the direction of z. ///

Proof of Proposition 2 1. First we prove by contraposition that $p^{I_1} = p^{I_2}$. Suppose that $p^{I_1} \neq p^{I_2}$. Since $p^{I_1} \cdot e = 1 = p^{I_2} \cdot e$, neither p^{I_1} nor p^{I_2} is a nonnegative multiple of the other. Thus, there is a $z \in \boldsymbol{R}^S$ such that $p^{I_1} \cdot z > 0 > p^{I_2} \cdot z$. Let $x \in T$. Since $\nabla V_n(xe) = \bar{\nabla} V_n(xe) = v_n'(x)\left(p^{I_n}\right)^\top$, define

$$\xi_n(\varepsilon) = V_n(xe + \varepsilon z) - V_n(xe) - v_n'(x)\left(p^{I_n} \cdot z\right)\varepsilon$$

and similarly for $\bar{\xi}_n(\varepsilon)$ using $\bar{V}_n$ in place of V_n. By twice differentiability, $\xi_n(\varepsilon)/\varepsilon \to 0$ and $\bar{\xi}_n(\varepsilon)/\varepsilon \to 0$ as $\varepsilon \to 0$. Moreover, for every $\varepsilon > 0$,

$$\frac{V_n(xe + \varepsilon z) - V_n(xe)}{\varepsilon} = v_n'(x)\left(p^{I_n} \cdot z\right) + \frac{\xi_n(\varepsilon)}{\varepsilon}.$$

Since $p^{I_1} \cdot z > 0$, $V_1(xe + \varepsilon z) - V_1(xe) > 0$ for every sufficiently small $\varepsilon > 0$. We can similarly show that $\bar{V}_1(xe - \varepsilon z) - \bar{V}_1(xe) < 0$ for every sufficiently small $\varepsilon > 0$. Since $V_1(xe) = v(x) = \bar{V}_1(xe)$, $V_1(xe + \varepsilon z) > \bar{V}_1(xe - \varepsilon z)$ for every sufficiently small $\varepsilon > 0$. Since $p^{I_2} \cdot z < 0$, we can similarly show that $V_2(xe + \varepsilon z) < \bar{V}_2(xe - \varepsilon z)$ for every sufficiently small $\varepsilon > 0$. By taking $f = xe + \varepsilon z$ and $g = xe - \varepsilon z$, we see that V_1 is not at least as ambiguity-averse as V_2.

To prove (14) and (15), let $x \in T$ and $z \in T$ and suppose that $E^{p^{I_1}}[z] = E^{p^{I_2}}[z] = 0$. Let $\varepsilon > 0$ and assume that it is sufficiently small. By applying Lemma 4 to $f = xe + \varepsilon z$

and using $E^{p^{I_1}}[z] = E^{p^{I_2}}[z] = 0$, we obtain $x - \kappa_2^x(\varepsilon, z) \geq x - \kappa_1^x(\varepsilon, z)$ and $x - \bar{\kappa}_1^x(\varepsilon, z) \geq x - \bar{\kappa}_2^x(\varepsilon, z)$, that is, (14) and (15) hold.

2. Let $f \in T^S$ and write $f = xe + z$ with $x \in T$ and $E^{p^{I_1}}[z] = E^{p^{I_2}}[z] = 0$. Then $V_n(f) = V_n\left((x - \kappa_n^x(1, z))\, e\right)$ and $\bar{V}_n(f) = \bar{V}_n\left((x - \bar{\kappa}_n^x(1, z))\, e\right)$. By applying (14) and (15) to $\varepsilon = 1$, $V_2\left((x - \kappa_2^x(1, z))\, e\right) \geq V_2\left((x - \kappa_1^x(1, z))\, e\right)$ and $\bar{V}_2\left((x - \bar{\kappa}_2^x(1, z))\, e\right) \leq \bar{V}_2\left((x - \bar{\kappa}_1^x(1, z))\, e\right)$. Hence, $V_2(f) \geq V_2\left((x - \kappa_1^x(1, z))\, e\right)$ and $\bar{V}_2(f) \leq \bar{V}_2\left((x - \bar{\kappa}_1^x(1, z))\, e\right)$. Thus, the conditions in Lemma 4 are satisfied when $y = x - \kappa_1^x(1, z)$ and $y = x - \bar{\kappa}_1^x(1, z)$. Hence, V_1 is at least as ambiguity-averse as V_2 in the ranking sense. ///

Proof of Proposition 1 We will prove Part 1 only, as Part 2 can similarly proved. Suppose that V_1 is at least as ambiguity-averse in choice as V_2 at x in the direction of z. By (8), (9), (10), and (11), $\bar{k}_1 + h_1\bar{k}_1 \leq \kappa_1^x(\varepsilon, z)$, $\bar{k}_1 \geq \bar{\kappa}_1^x(\varepsilon, z)$, $\bar{k}_2 + h_2\bar{k}_2 \geq \kappa_2^x(\varepsilon, z)$, and $\bar{k}_2 \leq \bar{\kappa}_2^x(\varepsilon, z)$. Thus,

$$h_1 = \frac{\left(\bar{k}_1 + h_1\bar{k}_1\right) - \bar{k}_1}{\bar{k}_1} \leq \frac{\kappa_1^x(\varepsilon, z) - \bar{\kappa}_1^x(\varepsilon, z)}{\bar{\kappa}_1^x(\varepsilon, z)},$$
$$h_2 = \frac{\left(\bar{k}_2 + h_1\bar{k}_2\right) - \bar{k}_2}{\bar{k}_2} \geq \frac{\kappa_2^x(\varepsilon, z) - \bar{\kappa}_2^x(\varepsilon, z)}{\bar{\kappa}_2^x(\varepsilon, z)}.$$

Since $h_1 \geq h_2$, we obtain (12). Conversely, if (12) holds, then let $\bar{k}_n = \bar{\kappa}_n^x(\varepsilon, z)$ for each n, h_1 be the left-hand side of (12), and h_2 be the right-hand side of (12). Then $h_1 \geq h_2$, and (8), (9), (10), and (11) holds with equalities. ///

Proof of Corollary 2 If V_1 is at least as ambiguity-averse as V_2 in the ranking sense, then, by Part 1 of Proposition 2, $p^{I_1} = p^{I_2}$, and (14) and (15) hold. Thus,

$$\kappa_1^x(\varepsilon, z) - \bar{\kappa}_1^x(\varepsilon, z) \geq \kappa_2^x(\varepsilon, z) - \bar{\kappa}_2^x(\varepsilon, z). \tag{59}$$

By (15) and (59),

$$\frac{\kappa_1^x(\varepsilon, z) - \bar{\kappa}_1^x(\varepsilon, z)}{\bar{\kappa}_1^x(\varepsilon, z)} \geq \frac{\kappa_2^x(\varepsilon, z) - \bar{\kappa}_2^x(\varepsilon, z)}{\bar{\kappa}_2^x(\varepsilon, z)}. \tag{60}$$

By Part 1 of Proposition 1, V_1 is at least as ambiguity-averse in choice as V_2 at x in the direction of z. ///

Proof of Proposition 3 It is routine to show that for every y,

$$
\begin{aligned}
b_1'(z) &= 2z, \\
b_1''(z) &= 2, \\
b_2'(z) &= 4z^3 + z, \\
b_2''(z) &= 12z + 1.
\end{aligned}
$$

By the definition of I_n,

$$
\begin{aligned}
\nabla I_n(g) &= \left(\frac{1}{2} - b_n'(g(1) - g(2)), \frac{1}{2} + b_n'(g(1) - g(2))\right), \\
\nabla^2 I_n(g) &= \begin{pmatrix} -b_n''(g(1) - g(2)) & b_n''(g(1) - g(2)) \\ b_n''(g(1) - g(2)) & -b_n''(g(1) - g(2)) \end{pmatrix}.
\end{aligned}
$$

Letting $g(1) = g(2)$, we obtain $p^{I_n} = \nabla I_n(g) = (1/2, 1/2)$ for each n and

$$
\begin{aligned}
\nabla^2 I_1(g) &= \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}, \\
\nabla^2 I_2(g) &= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.
\end{aligned}
$$

Thus, $\nabla^2 I_1(g) - \nabla^2 I_2(g)$ is negative semidefinite, that is, $z^\top \nabla^2 I_1(g) z \leq z^\top \nabla^2 I_2(g) z$ for every $z \in \boldsymbol{R}^2$. Moreover, the weak inequality holds as a strict inequality if $z = (1, -1)$. By (6), this proves Part 2. As for Part 3, note that

$$
V_n(\overline{x}, \underline{x}) = I_n(3, 2) = \frac{3+2}{2} - b_n(3-2) = \frac{5}{2} - b_n(1) = \begin{cases} \frac{3}{2} & \text{if } n = 1, \\ 1 & \text{if } n = 1. \end{cases}
$$

Let $x \in (v^{-1}(1), v^{-1}(3/2))$. Then

$$
\begin{aligned}
V_1(\overline{x}, \underline{x}) &> v(x) = \bar{V}_1(x, x), \\
V_2(\overline{x}, \underline{x}) &< v(x) = \bar{V}_2(x, x).
\end{aligned}
$$

This completes the proof. ///

Proof of Theorem 2 By the chain rule differentiation,

$$
\frac{\partial \pi^x}{\partial \varepsilon}(\varepsilon, z) = \frac{1}{\nabla V(xe + \varepsilon z)e} z^\top \nabla^2 V(xe + \varepsilon z) z - \frac{\nabla V(xe + \varepsilon z) z}{(\nabla V(xe + \varepsilon z)e)^2} e^\top \nabla^2 V(xe + \varepsilon z) z.
$$

Since $\nabla V(ex) = v'(x)p^I$, $p^I \cdot e = 1$, and $p^I \cdot z = 0$,

$$\frac{\partial \pi^x}{\partial \varepsilon}(0, z) = \frac{z^\top \nabla^2 V(xe)z}{v'(x)}. \tag{61}$$

Similarly,

$$\frac{\partial \bar{\pi}^x}{\partial \varepsilon}(0, z) = \frac{z^\top \nabla^2 \bar{V}(xe)z}{v'(x)}. \tag{62}$$

Since $\pi^x(\varepsilon, z) - \bar{\pi}^x(\varepsilon, z) \to 0$ and $\bar{\pi}^x(\varepsilon, z) \to 0$ as $\varepsilon \to 0$, by L'Hôpital's rule,

$$\frac{\pi^x(\varepsilon, z) - \bar{\pi}^x(\varepsilon, z)}{\bar{\pi}^x(\varepsilon, z)} \to \frac{\dfrac{\partial \pi^x}{\partial \varepsilon}(0, z) - \dfrac{\partial \bar{\pi}^x}{\partial \varepsilon}(0, z)}{\dfrac{\partial \bar{\pi}^x}{\partial \varepsilon}(0, z)}$$

as $\varepsilon \to 0$. The theorem then follows from (61) and (62). ///

Proof of Theorem 3 By definition,

$$\begin{aligned}
\rho^x(\varepsilon, z) - \bar{\rho}^x(\varepsilon, z) &= \frac{1}{(x + \pi^x(\varepsilon, z))\,(x + \bar{\pi}^x(\varepsilon, z))} \left(-\,(x + \bar{\pi}^x(\varepsilon, z))\,\pi^x(\varepsilon, z) + (x + \pi^x(\varepsilon, z))\,\bar{\pi}^x(\varepsilon, z)\right) \\
&= \frac{x}{(x + \pi^x(\varepsilon, z))\,(x + \bar{\pi}^x(\varepsilon, z))} \left(\bar{\pi}^x(\varepsilon, z) - \pi^x(\varepsilon, z)\right).
\end{aligned}$$

Thus,

$$\frac{\rho^x(\varepsilon, z) - \bar{\rho}^x(\varepsilon, z)}{\bar{\rho}^x(\varepsilon, z)} = \frac{x}{x + \pi^x(\varepsilon, z)} \frac{\pi^x(\varepsilon, z) - \bar{\pi}^x(\varepsilon, z)}{\bar{\pi}^x(\varepsilon, z)}.$$

Since $\pi^x(\varepsilon, z) \to 0$ as $\varepsilon \to 0$, the proof is completed by Theorem 2. ///

Proof of Theorem 4 By definition,

$$\rho^x(\varepsilon, z) - \tilde{\rho}^x(\varepsilon, z) = -\frac{\dfrac{\nabla I(v \circ (xe + \varepsilon z))z}{\nabla I(v \circ (xe + \varepsilon z))e}}{x + \pi^x(\varepsilon, z)}.$$

Thus,

$$\frac{\rho^x(\varepsilon, z) - \tilde{\rho}^x(\varepsilon, z)}{\tilde{\rho}^x(\varepsilon, z)} = -\frac{\dfrac{\nabla I(v \circ (xe + \varepsilon z))z}{\nabla I(v \circ (xe + \varepsilon z))e}}{\dfrac{\nabla I(v \circ (xe + \varepsilon z))z}{\nabla I(v \circ (xe + \varepsilon z))e} - \pi^x(\varepsilon, z)}.$$

Note here that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\frac{\nabla I(v\circ(xe+\varepsilon z))z}{\nabla I(v\circ(xe+\varepsilon z))e}\right) = \frac{1}{\nabla I(v\circ(xe+\varepsilon z))e} z^\top \nabla^2 I(v'\circ(xe+\varepsilon z))\left((v'\circ(xe+\varepsilon z))\circ z\right)$$
$$- \frac{\nabla I(v\circ(xe+\varepsilon z))z}{\left(\nabla I(v\circ(xe+\varepsilon z))e\right)^2} e^\top \nabla^2 I(v'\circ(xe+\varepsilon z))\left(v'\circ(xe+\varepsilon z)\right).$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\frac{\nabla I(v\circ(xe+\varepsilon z))z}{\nabla I(v\circ(xe+\varepsilon z))e}\right)\Bigg|_{\varepsilon=0} = v'(x) z^\top \nabla^2 I(v(x)e)z$$

By (5), (61), and L'Hôpital's rule, as $\varepsilon \to 0$,

$$-\frac{\dfrac{\nabla I(v\circ(xe+\varepsilon z))z}{\nabla I(v\circ(xe+\varepsilon z))e}}{\dfrac{\nabla I(v\circ(xe+\varepsilon z))z}{\nabla I(v\circ(xe+\varepsilon z))e} - \pi^x(\varepsilon,z)} \to -\frac{v'(x)z^\top\nabla^2 I(v(x)e)z}{v'(x)z^\top\nabla^2 I(v(x)e)z - \dfrac{z^\top\nabla^2 V(xe)z}{v'(x)}}$$
$$= \frac{\left(v'(x)\right)^2 z^\top\nabla^2 I(v(x)e)z}{z^\top\nabla^2 V(xe)z - \left(v'(x)\right)^2 z^\top\nabla^2 I(v(x)e)z}$$
$$= \frac{z^\top\left(\nabla^2 V(xe) - \nabla^2\bar{V}(xe)\right)z}{z^\top\nabla^2\bar{V}(xe)z} = H^x(z).$$

///

Proof of Lemma 1 Since V is concave, $\gamma^x(\varepsilon,z)$ is a concave function of ε with $xe+\varepsilon z \in T^S$ and $\partial\gamma^x(0,z)/\partial_+\varepsilon$ is well defined. Define

$$A = \{(f,y)\in T^S\times\boldsymbol{R} \mid V(f)\geq y\},$$
$$B = \left\{(xe+\delta z, y)\in\boldsymbol{R}^S\times\boldsymbol{R} \mid v(x)+\delta\frac{\partial\gamma^x}{\partial_+\varepsilon}(0,z) < y\right\}.$$

Then A is convex because V is concave, and B is convex by construction. Moreover, $A\cap B=\varnothing$. Indeed, let $\delta\in\boldsymbol{R}$ and $y\in\boldsymbol{R}$, and assume that $xe+\delta z\in T^S$ and $(xe+\delta z,y)\in A$. Since $\gamma^x(\cdot,z)$ is concave,

$$\gamma^x(\delta,z)\leq\gamma^x(0,z)+\delta\frac{\partial\gamma^x}{\partial_+\varepsilon}(0,z),$$

that is,

$$V(xe+\delta z)\leq v(x)+\delta\frac{\partial\gamma^x}{\partial_+\varepsilon}(0,z).$$

Since $V(xe+\delta z) \geq y$,

$$y \leq v(x) + \delta \frac{\partial \gamma^x}{\partial_+ \varepsilon}(0,z).$$

Thus, $xe+\delta z \notin B$. Thus $A \cap B = \varnothing$.

By the separating hyperplane theorem, there are a $\hat{p} = (p, p^0) \in (\boldsymbol{R}^S \times \boldsymbol{R}) \setminus \{0\}$ and a $c \in \boldsymbol{R}$ such that $\hat{p} \cdot a \geq c \geq \hat{p} \cdot b$ for every $a \in A$ and every $b \in B$. Since $(xe, v(x))$ belongs to A and the closure of B, $c = p \cdot (xe) + p^0 v(x)$. If $p^0 > 0$, then $\{\hat{p} \cdot a \mid a \in A\}$ is not bounded from below and $\{\hat{p} \cdot b \mid b \in B\}$ is not bounded from above. Thus $p^0 \leq 0$.

We now show that $p \in \boldsymbol{R}^S_+$. Indeed, let $f \in T^S$ and assume that $f - xe \in \boldsymbol{R}^S_+$. Since V is non-decreasing, $V(f) \geq v(x)$. Hence $(f, (V(f)+v(x))/2) \in A$. Thus, $p \cdot f + p^0(V(f) + v(x))/2 \geq c = p \cdot (xe) + p^0 v(x)$. Thus, $p \cdot (f - xe) \geq p^0(v(x) - V(f)) \geq 0$ because $p^0 \leq 0$ and $v(x) - V(f) \leq 0$. Thus, $p \in \boldsymbol{R}^S_+$.

Let $x' \in T$ and assume that $x' > x$. Then $(x'e, v(x')) \in A$. If $p = 0$, then $p^0 < 0$. Hence, $p \cdot (x'e) + p^0 v(x') = p^0 v(x') < p^0 v(x) = p \cdot (xe) + p^0 v(x) = c$, which is a contradiction. Thus $p \in \boldsymbol{R}^S_+ \setminus \{0\}$. Thus $p \cdot e > 0$. By dividing $\hat{p}$ by $p \cdot e$ whenever necessary, we can assume that $p \cdot e = 1$, that is, $p \in \Delta$.

Since $(x'e, v(x')) \in A$ for every $x' \in T$,

$$p \cdot (x'e) + p^0 v(x') \geq c = p \cdot (xe) + p^0 v(x).$$

Thus the first-order condition of minimizing the left-hand side of the above equality over $x' \in T$ is that

$$p \cdot e + p^0 v'(x) = 1 + p^0 v'(x) = 0,$$

that is,

$$p^0 = -\frac{1}{v'(x)} < 0. \tag{63}$$

Since

$$\left(xe + \delta z, v(x) + \delta \frac{\partial \gamma^x}{\partial_+ \varepsilon}(0,z) \right) = (xe, v(x)) + \delta \left(z, \frac{\partial \gamma^x}{\partial_+ \varepsilon}(0,z) \right)$$

belongs to the closure of B for every $\delta \in \boldsymbol{R}$ and since $\{\hat{p} \cdot b \mid b \text{ belongs to the closure of } B\}$ is bounded from above,

$$p \cdot z + p^0 \frac{\partial \gamma^x}{\partial_+ \varepsilon}(0,z) = 0. \tag{64}$$

By (63) and (64),

$$v'(x)(p \cdot z) = \frac{\partial \gamma^x}{\partial_+ \varepsilon}(0, z).$$

It remains to prove that $v'(x)p$ is a supergradient of V at xe. Since $(f, V(f)) \in A$ for every $f \in T^S$,

$$p \cdot f + p^0 V(f) \geq c = p \cdot (xe) + p^0 v(x).$$

By (63),

$$V(f) \leq c = (v'(x)p) \cdot (f - xe) + V(xe).$$

That is, $v'(x)p$ is a supergradient of V at xe. ///

Proof of Lemma 2 For each ε close to 0, write

$$\xi^x(\varepsilon) = v(x + \varepsilon) - \left(v(x) + v'(w)\varepsilon + \frac{1}{2} v''(x)\varepsilon^2 \right).$$

Then $\xi^x(\varepsilon)/\varepsilon^2 \to 0$ as $\varepsilon \to 0$. Moreover, for every n,

$$\begin{aligned} \bar{V}_{p_n}(xe + \varepsilon_n z) - \bar{V}_{p_n}(xe) &= \sum_s p_n(s) \left(v'(x)\varepsilon_n z(s) + \frac{1}{2} v''(x)(\varepsilon_n z(s))^2 + \xi^x(\varepsilon_n z(s)) \right) \\ &= v'(x)(p_n \cdot z)\varepsilon_n + \frac{1}{2} v''(x)\|z\|_{p_n}^2 \varepsilon^2 + \sum_s p_n(s)\xi^x(\varepsilon_n z(s)), \end{aligned}$$

and a similar expression is obtained for $\bar{V}_{q_n}(xe + \varepsilon_n z) - \bar{V}_{q_n}(xe)$. Since $\bar{V}_{p_n}(xe + \varepsilon_n z) \leq \bar{V}_{q_n}(xe + \varepsilon_n z)$ and $\bar{V}_{p_n}(xe) = v(x) = \bar{V}_{q_n}(xe)$,

$$\bar{V}_{p_n}(xe + \varepsilon_n z) - \bar{V}_{p_n}(xe) \leq \bar{V}_{q_n}(xe + \varepsilon_n z) - \bar{V}_{q_n}(xe),$$

that is,

$$\begin{aligned} & v'(x)(p_n \cdot z)\varepsilon_n + \frac{1}{2} v''(x)\|z\|_{p_n}^2 \varepsilon_n^2 + \sum_s p_n(s)\xi^x(\varepsilon_n z(s)) \\ \leq & v'(x)(q_n \cdot z)\varepsilon_n + \frac{1}{2} v''(x)\|z\|_{q_n}^2 \varepsilon_n^2 + \sum_s q_n(s)\xi^x(\varepsilon_n z(s)). \end{aligned}$$

Since $p_n \cdot z = q_n \cdot z$ and $v''(x) < 0$, this inequality can be rewritten as

$$\|z\|_{p_n}^2 - \|z\|_{q_n}^2 \geq \frac{2}{v''(x)} \sum_s (q_n(s) - p_n(s)) \frac{\xi^x(\varepsilon_n z(s))}{(\varepsilon_n z(s))^2} (z(s))^2.$$

As $n \to \infty$, the left-hand side converges to $\|z\|_p^2 - \|z\|_q^2$, while the right-hand side converges to 0. Thus $\|z\|_p^2 - \|z\|_q^2 \geq 0$. ///

Proof of Proposition 4 First, we show that I satisfies all four conditions of Assumptions 1 and 2. Indeed, the twice differentiability follows from that of v and ϕ. If $z = ye$ for some $y \in v(T)$, then $p \cdot z = y = I(z)$ for every $p \in \Delta$. Thus, I is normalized. By differentiating both sides of $\phi(I(z)) = \int_\Delta \phi(p \cdot z)\, \mathrm{d}\mu(p)$ with respect to z, we obtain

$$\phi'(I(z))\nabla I(z) = \int_\Delta \phi'(p \cdot z)\, p^\top \mathrm{d}\mu(p) \tag{65}$$

Since $\phi' > 0$ and $\int_\Delta p\, \mathrm{d}\mu(p) \in \boldsymbol{R}_{++}^S$, $\nabla I(z) \in \boldsymbol{R}_{++}^S$. If $z = ye$ for some $y \in v(T)$, the equality is reduced to $\nabla I(ye) = \int_\Delta p^\top \mathrm{d}\mu(p)$. We can thus let $p^I = \int_\Delta p\, \mathrm{d}\mu(p)$ to show that the last condition of Assumptions 1 and 2 is met.

By differentiating both sides of (65) with respect to z, we obtain

$$\phi''(I(z))\nabla I(z)^\top \nabla I(z) + \phi'(I(z))\nabla^2 I(z)^\top = \int_\Delta \phi''(p \cdot z)\, pp^\top\, \mathrm{d}\mu(p).$$

Thus,

$$\nabla^2 I(z) = \frac{1}{\phi'(I(z))}\left(\int_\Delta \phi''(p \cdot z)\, pp^\top\, \mathrm{d}\mu(p) - \phi''(I(z))\nabla I(z)^\top \nabla I(z)\right). \tag{66}$$

In particular, if $z = ye$ for some $y \in T^S$, then $p \cdot z = y = I(z)$ for every $p \in \Delta$ and $\nabla I(z) = (p^I)^\top$. Thus

$$\nabla^2 I(ye) = \frac{\phi''(y)}{\phi'(y)}\left(\int_\Delta pp^\top\, \mathrm{d}\mu(p) - p^I (p^I)^\top\right) = \frac{\phi''(y)}{\phi'(y)} \int_\Delta (p - p^I)(p - p^I)^\top\, \mathrm{d}\mu(p). \tag{67}$$

Thus, for every $x \in T$ and every $z \in \boldsymbol{R}^S$,

$$z^\top \nabla^2 I(v(x)e) z = \frac{\phi''(v(x))}{\phi'(v(x))} \int_\Delta \left((p - p^I) \cdot z\right)^2\, \mathrm{d}\mu(p). \tag{68}$$

Therefore,

$$H^x(z) = \frac{\dfrac{\phi''(v(x))}{\phi'(v(x))}}{\dfrac{v''(x)}{v'(x)}} v'(x) \frac{\int_\Delta \left((p - p^I) \cdot z\right)^2\, \mathrm{d}\mu(p)}{\|z\|_{p^I}^2}. \tag{69}$$

Note here that $p \cdot z = E^p[z]$ and $p^I \cdot z = E^{p^I}[z]$. Since $E^\mu[E^\cdot[z]] = E^{p^I}[z]$ by the law of iterated expectation,

$$\int_{\Delta} \left((p - p^I) \cdot z\right)^2 \mathrm{d}\mu(p) = \int_{\Delta} \left(E^p[z] - E^{p^I}[z]\right)^2 \mathrm{d}\mu(p) = \mathrm{Var}^{\mu}[E^{\cdot}[z]]. \tag{70}$$

By definition,

$$\|z\|_{p^I}^2 = \left(E^{p^I}[z]\right)^2 + \mathrm{Var}^{p^I}[z]. \tag{71}$$

Recall also that $w(x) = \phi(v(x))$ for every $x \in T$. By differentiating both sides with respect to x, we obtain

$$w'(x) = \phi'(v(x))v'(x). \tag{72}$$

By differentiating both sides with respect to x, we obtain

$$w''(x) = \phi''(v(x))(v'(x))^2 + \phi'(v(x))v''(x). \tag{73}$$

By dividing each side of (73) by the same side of (72), we obtain

$$\frac{w''(x)}{w'(x)} = \frac{\phi''(v(x))}{\phi'(v(x))}v'(x) + \frac{v''(x)}{v'(x)},$$

that is,

$$\frac{\dfrac{\phi''(v(x))}{\phi'(v(x))}v'(x)}{\dfrac{v''(x)}{v'(x)}} = \frac{-\dfrac{w''(x)}{w'(x)}}{-\dfrac{v''(x)}{v'(x)}} - 1. \tag{74}$$

By plugging (74), (70), and (71) into (69), we obtain (29). ///

Proof of Proposition 5 In the four conditions of Assumptions 1 and 2, it follows from the groundedness of G that I is normalized. Indeed, $I(ye) = \min_p G(y, p) = y$ for every $y \in v(T)$. The other three conditions will be proved later.

The left-hand side of (36) is a real-valued linear function on $e^{\perp}$ that is, in turn, a function of (g, p). Its derivative with respect to p is the quadratic form on $e^{\perp}$ represented by

$$\left(g - (e \cdot g)S^{-1}e\right)\left(\nabla_y^2 G(p \cdot g, p)g^{\top} + \nabla_p \nabla_y G(p \cdot g, p)\right) + \nabla_y(\nabla_p G(p \cdot g, p))^{\top} g^{\top} + \nabla_p^2 G(p \cdot g, p).$$

If $g = ye$ for some $y \in v(T)$ and the first-order condition (36) is met, then $p = p^I$ and this can be rewritten as

$$y\nabla_y(\nabla_p G(y,p^I))^\top e^\top + \nabla_p^2 G(y,p^I).$$

For every $z \in e^\perp \setminus \{0\}$,

$$z^\top \left(y\nabla_y(\nabla_p G(y,p^I))^\top e^\top + \nabla_p^2 G(y,p^I)\right) z = z^\top \nabla_p^2 G(y,p^I) z > 0.$$

Thus, the quadratic form is positive definite.

The derivative of the left-hand side of (36) with respect to g is the bilinear form on $e^\perp$ represented by

$$\nabla_y G(p\cdot g,p)\left(I_{S\times S} - \frac{1}{S}ee^\top\right) + \nabla_y^2 G(p\cdot g,p)\left(I_{S\times S} - \frac{1}{S}ee^\top\right) gp^\top + \nabla_y\left(\nabla_p G(p\cdot g,p)\right)^\top p^\top$$

If $g = ye$ for some $y \in v(T)$ and the first-order condition (36) is met, then $p = p^I$ and this can be rewritten as

$$\nabla_y G(y,p^I)\left(I_{S\times S} - \frac{1}{S}ee^\top\right) + \nabla_y\left(\nabla_p G(y,p^I)\right)^\top (p^I)^\top$$

Since $G(y,p^I) = y$ and $\nabla_p G(y,p^I) = 0$ for every $y \in v(T)$, $\nabla_y G(y,p^I) = 1$ and $\nabla_y\left(\nabla_p G(y,p^I)\right) = 0$ for every $y \in v(T)$. Thus, the above expression is equal to

$$I_{S\times S} - \frac{1}{S}ee^\top.$$

For every $z \in e^\perp \setminus \{0\}$,

$$\left(I_{S\times S} - \frac{1}{S}ee^\top\right) z = z.$$

That is, the partial derivative of the left-hand side of (36) with respect to p can be identified with the identity mapping on $e^\perp$.

As we stated after presenting Assumption 4, for every $g \in v(T)^S$, the minimization problem $\inf_{p\in\Delta} G(p\cdot g,p)$ has a unique solution. Define $\chi : v(T)^S \to \Delta$ by letting $\chi(g)$ be the solution for $g \in v(T)^S$. By the invertibility of the quadratic form representing the partial derivative of the left-hand side of (36) with respect to p, for every $y \in v(T)$, χ is continuously differentiable at ye and $D\chi(ye)$ coincides with $-\left(\nabla_p^2 G(y,p^I)\right)^{-1}$ on $e^\perp$.

Since $z-(e\cdot z)S^{-1}e \in e^{\perp}$ for every $z \in \boldsymbol{R}^S$, this means that

$$\left(z-(e\cdot z)S^{-1}e\right)^{\top} D\chi(ye)\left(z-(e\cdot z)S^{-1}e\right) = -\left(z-(e\cdot z)S^{-1}e\right)^{\top}\left(\nabla_p^2 G(y,p^I)\right)^{-1}\left(z-(e\cdot z)S^{-1}e\right). \tag{75}$$

By the definition of G and χ, $I(g) = G(\chi(g)\cdot g, \chi(g)$ for every $g \in v(T)^S$. By the first-order condition (36),

$$\nabla I(g) = \nabla_y G(\chi(g)\cdot g, \chi(g))\chi(g)^{\top}$$

for every $g \in v(T)^S$. Since G is twice continuously differentiable and χ is continuously differentiable, this shows that I is twice continuously differentiable. Since $\nabla_y G(\chi(g)\cdot g, \chi(g)) > 0$ and $\chi(g) \in \Delta \cap \boldsymbol{R}^S_{++}$, this shows that I is differentiably strictly increasing. It also implies that for every $y \in v(T)$, $\nabla I(ye) = p^I$, because $\nabla_y G(y, p^I) = 1$ and $\chi(ye) = p^I$. We have thus shown that I satisfies all the conditions of Assumptions 1 and 2.

By differentiating both sides of the above equality with respect to g, we obtain

$$\begin{aligned}\nabla^2 I(g) =& \nabla_y G(\chi(g)\cdot g, \chi(q))D\chi(g)\\ &+ \chi(g)\left(\nabla_y^2 G(\chi(g)\cdot g, \chi(q))(\chi(g)^{\top} + g^{\top}D\chi(g)) + \nabla_p\nabla_y G(\chi(g)\cdot g, \chi(q))D\chi(g)\right)\end{aligned}$$

When $g = ye$ for some $y \in v(T)$, $\chi(g) = p^I$. Since $\nabla_y G(y, p^I) = 1$, $\nabla_y^2 G(y, p^I) = 0$, and $\nabla_p \nabla_y G(\chi(g)\cdot g, \chi(g)) = \nabla_y\left(\nabla_p G(y, p^I)\right) = 0$ for every $y \in v(T)$. Thus, this can be rewritten as

$$\nabla^2 I(ye) = D\chi(ye). \tag{76}$$

Since $\nabla I(ye) = (p^I)^{\top}$ for every $y \in v(T)$, $\nabla^2 I(ye)y = 0$ for every $y \in v(T)$. Thus, for every $z \in \boldsymbol{R}^S$,

$$z^{\top}\nabla^2 I(ye)z = \left(z-(e\cdot z)S^{-1}e\right)^{\top}\nabla^2 I(ye)\left(z-(e\cdot z)S^{-1}e\right).$$

By (75) and (76),

$$\begin{aligned}z^{\top}\nabla^2 I(ye)z =& \quad \left(z-(e\cdot z)S^{-1}e\right)^{\top} D\chi(ye)\left(z-(e\cdot z)S^{-1}e\right)\\ =& \quad \left(z-(e\cdot z)S^{-1}e\right)^{\top}\nabla^2 I(ye)\left(z-(e\cdot z)S^{-1}e\right)\\ =& -\left(z-(e\cdot z)S^{-1}e\right)^{\top}\left(\nabla_p^2 G(y,p^I)\right)^{-1}\left(z-(e\cdot z)S^{-1}e\right).\end{aligned}$$

By letting $y = v(x)$ and plugging this equality into (6), we complete the proof. ///

Proof of Corollary 3 Define $G : v(T) \times \Delta \to \boldsymbol{R} \cup \{+\infty\}$ by letting $G(y,p) = y + c(p)$ for every $(y,p) \in v(T) \times \Delta$. Since $\nabla_p G(y,p) = \nabla c(p)$ and $\nabla_p^2 G(y,p) = \nabla^2 c(p)$ for every $(y,p) \in D$, if c satisfies the conditions of Assumptions 6 and 7, then G satisfies the conditions of Assumptions 3 and 4. By applying the First-order condition of c to $g = ye$ for any $y \in v(T)$, we see that there is a $p \in D$ such that $\nabla c(p) = 0$. Since c is strictly convex and grounded, $c(p) = 0$ and p is the only vector in Δ that satisfies this condition. We can thus let p^I be this p. Then Assumption 5 is satisfied. Since $\nabla_p^2 G(y,p) = \nabla^2 c(p)$ for every $(y,p) \in D$, by plugging this into (37), we complete the proof. ///

Proof of Corollary 4 To show that c satisfies all the conditions of Assumption 7, it suffices to show that $R(\,\cdot \mid p^I)$ satisfies all the conditions of Assumption 7. Indeed, $\nabla_p R(p \mid p^I)$ is the restriction, onto $e^\perp$, of the linear function defined by the dot product with

$$\left(\ln \frac{p(1)}{p^I(1)} + 1, \ldots, \ln \frac{p(S)}{p^I(S)} + 1\right),$$

and, thus, can be identified with its orthogonal projection onto $e^\perp$, which is equal to

$$\left(\ln \frac{p(1)}{p^I(1)} - \frac{1}{S}\sum_t \ln \frac{p(t)}{p^I(t)}, \ldots, \ln \frac{p(S)}{p^I(S)} - \frac{1}{S}\sum_t \ln \frac{p(t)}{p^I(t)}\right).$$

The Hessian $\nabla_p^2 R(p \mid p^I)$ is the restriction, onto $e^\perp$, of the quadratic form defined by the $S \times S$ diagonal matrix

$$\begin{pmatrix} \frac{1}{p(1)} & & \\ & \ddots & \\ & & \frac{1}{p(S)} \end{pmatrix}. \tag{77}$$

Since this $S \times S$ is positive definite, so is the Hessian $\nabla_p^2 R(p \mid p^I)$. Moreover, $R(p^I \mid p^I) = 0$ and $\nabla_p R(p^I \mid p^I) = 0$. Thus, the minimum is attained at p^I and its value is equal to zero. To prove the Onto property of c, it suffices to show that for every $z \in e^\perp$, there is a $p \in \Delta \cap \boldsymbol{R}^S_{++}$ such that $\nabla c(p) = z$. Indeed, define $p \in \Delta \cap \boldsymbol{R}^S_{++}$ by letting, for each s,

$$p(s) = \frac{p^I(s)\mathrm{e}^{z(s)}}{\sum_t p^I(t)\mathrm{e}^{z(t)}}.$$

Then

$$\ln \frac{p(s)}{p^I(s)} = z(s) - \ln \left(\sum_t p^I(t) \mathrm{e}^{z(t)} \right).$$

Hence $\nabla c(p) = z$.

To prove (40), note that by (77), $\left((v'(x))^{-1} \nabla^2 c\left(p^I\right)\right)^{-1}$ is equal to the restriction onto $e^\top$ of

$$\frac{v'(x)}{\theta}[p].$$

Thus,

$$H^x(z) = \frac{\dfrac{\dfrac{v'(x)}{\theta} \|z - (e \cdot z)S^{-1}e\|_{p^I}^2}{\|z\|_p^2}}{-\dfrac{v''(x)}{v'(x)}} = -\frac{(v'(x))^2}{\theta v''(x)} \frac{\|z - (e \cdot z)S^{-1}e\|_{p^I}^2}{\|z\|_{p^I}^2}.$$

///

Proof of Proposition 6 By (74) and the definition of ϕ,

$$\frac{-\dfrac{w''(x)}{w'(x)}}{-\dfrac{v''(x)}{v'(x)}} - 1 = -\frac{1}{\theta\varepsilon^2} \frac{(v'(x))^2}{v''(x)}. \tag{78}$$

By (42) and (43),

$$\begin{aligned}
E^\mu[E^{\cdot}[z]] &= \int_\Delta p \cdot z \,\mathrm{d}\mu(p) = z^\top \int_\Delta p \,\mathrm{d}\mu(p) = z^\top p^I = p^I \cdot z, \\
\mathrm{Var}^\mu[E^{\cdot}[z]] &= \int_\Delta (p \cdot z - p^I \cdot z)^2 \,\mathrm{d}\mu(p) \\
&= \int_\Delta ((p - p^I)^\top z)^2 \,\mathrm{d}\mu(p) \\
&= z^\top \left(\int_\Delta \left(p - p^I\right) \left(p - p^I\right)^\top \mathrm{d}\mu(p) \right) z \\
&= z^\top \left(\varepsilon^2 \left(I_{S\times S} - \frac{1}{S} ee^\top \right) [p^I] \left(I_{S\times S} - \frac{1}{S} ee^\top \right) \right) z \\
&= \varepsilon^2 (z - (e \cdot z)S^{-1}e)^\top [p^I] (z - (e \cdot z)S^{-1}e) \\
&= \varepsilon^2 \|z - (e \cdot z)S^{-1}e\|_{p^I}^2.
\end{aligned}$$

By (29),

$$H_2^x(z) = \left(-\frac{1}{\theta\varepsilon^2}\frac{(v'(x))^2}{v''(x)}\right)\frac{\varepsilon^2\|z-(e\cdot z)S^{-1}e\|_{p^I}^2}{\left(E^{p^I}[z]\right)^2+\mathrm{Var}^{p^I}[z]} = -\frac{(v'(x))^2}{\theta v''(x)}\frac{\|z-(e\cdot z)S^{-1}e\|_{p^I}^2}{\|z\|_{p^I}^2}$$

for every x and z. By (40), this is equal to $H_1^x(z)$. ///

Proof of Corollary 6 Define $G : v(T)\times\Delta\to\boldsymbol{R}\cup\{+\infty\}$ by letting $G(y,p)=y/\phi(p)$ for every $(y,p)\in v(T)\times\Delta$. Then all the conditions in Assumption 3 are met. Indeed, the lower semi-continuity of G follows from the upper semi-continuity of ϕ; the quasi-convexity of G follows from the concavity of ϕ; the groundedness of G follows from the normality of ϕ; and the non-decreasingness of $G(\,\cdot\,p)$ follows from the non-negativity of ϕ. Let $D=\{(y,p)\in v(T)^S\times\Delta \mid \phi(p)>0\}$, then G is twice continuously differentiable on D because ϕ is twice continuously differentiable on $\{p\in\Delta\mid\phi(p)>0\}$. Moreover, for every $(y,p)\in D$,

$$\begin{aligned}
\nabla_y G(y,p) &= \frac{1}{\phi(p)},\\
\nabla_y^2 G(y,p) &= 0,\\
\nabla_p G(y,p) &= -\frac{y}{(\phi(p))^2}\nabla\phi(p),\\
\nabla_p^2 G(y,p) &= -\frac{y}{(\phi(p))^2}\left(\nabla^2\phi(p)-\frac{2}{\phi(p)}\nabla\phi(p)^\top\nabla\phi(p)\right).
\end{aligned}$$

Thus $\nabla_y G(y,p)>0$ and, if $\nabla_p G(y,p)=0$, then $\nabla\phi(p)=0$ and

$$\nabla_p^2 G(y,p) = -\frac{y}{(\phi(p))^2}\nabla^2\phi(p).$$

Since $\nabla^2\phi(p)$ defines a negative-define quadratic form, $\nabla_p^2 G(y,p)$ defines a positive-definite quadratic form.

It is straightforward to show that since $\nabla^2\phi(p)$ defines a negative semi-definite quadratic form on the hyperplane (in $e^\perp$) with normal $\nabla\phi(p)$. Moreover, (36) holds if and only if (45) holds. Thus, all the conditions in Assumption 4 follow from the corresponding conditions in Assumption 9. Furthermore, if $\nabla_p G(y,p)=0$, then $\nabla\phi(p)=0$ and $\phi(p)=1$ because ϕ is concave. Moreover, since $\nabla^2\phi(p)$ determines a negative-definite quadratic form, such a p is unique. Hence we can let this p be p^I in Assumption 5.

Since $\nabla_p^2 G(v(x),p^I)=-v(x)\nabla^2\phi(p^I)$ for every $x\in T$, by plugging this into (37), we complete the proof. ///

References

[1] Ahn, D., Choi, S., Gale, D. and Kariv, S. 2014, Estimating ambiguity aversion in a portfolio choice experiment. *Quantitative Economics* 5:195-223. doi: 10.3982/QE243

[2] Altug, S., F. Collard, C. Çakmakli, S. Mukerji, and H. Özsöylev, 2018, Ambiguous business cycles: A quantitative assessment, working paper.

[3] Ang, A., 2014, *Portfolio Management: A Systematic Approach to Factor Investing*, Oxford University Press.

[4] Attanasi, G., C. Gollier, A. Montesano, and N. Pace, 2014, Eliciting ambiguity aversion in unknown and in compound lotteries: A smooth ambiguity model experimental study, *Theory and Decision* 77, 485-530.

[5] Backus, D., A. Ferriere, and S. Zin, 2015, Risk and ambiguity in models of business cycles, *Journal of Monetary Economics* 69: 42-63.

[6] Banz, Rolf W., 1981, The relationship between return and market values of common stocks, *Journal of Financial Economics* 9, 3–18.

[7] Black, F., 1972, Capital market equilibrium with restricted borrowing, *Journal of Business* 45, 444–455.

[8] Bossaerts, P., P. Ghirardato, S. Guarnaschelli, and W.R. Zame, 2010, Ambiguity in asset markets: Theory and experiment, *Review of Financial Studies* 23: 1325–1359.

[9] Canner, N., N. G. Mankiw, and D. N. Weil, 1997, An asset allocation puzzle, *American Economic Review* 87: 181–191.

[10] Campbell, J. Y., A. Lo, and C. MacKinlay. 1997, *The Econometrics of Financial Markets*. Princeton University Press.

[11] Cerreia-Vioglio, S., F. Maccheroni, and M. Marinacci, forthcoming, Ambiguity aversion and wealth effects, Journal of Economic Theory.

[12] Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio, 2011, Uncertainty averse preferences, Journal of Economic Theory 146, 1275–1330.

[13] Chen, H., Ju, N., and J. Miao, 2014, Dynamic asset allocation with ambiguous return predictability. *Review of Economic Dynamics* 17, 799–823.

[14] Chateauneuf, A., and J. H. Faro, 2009, Ambiguity through confidence functions, Journal of Mathematical Economics 45, 535–558.

[15] Collard, F., S. Mukerji, K. Sheppard, and J. Tallon, 2018, Ambiguity and the historical equity premium, *Quantitative Economics* 945–993.

[16] Ellsberg, D., 1961, Risk, ambiguity, and the Savage axiom, *Quarterly Journal of Economics* 75, 643–669.

[17] Epstein, L. G. 2010, A paradox for the "smooth ambiguity" model of preference, Econometrica 78: 2085–2099.

[18] Epstein, L. G., and J. Miao, 2003, A two-person dynamic equilibrium under ambiguity, *Journal of Economic Dynamics and Control* 27, 1253–1288.

[19] Epstein, L. G., and Zin, S. E. (1989). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. Econometrica, 57(4), 937-969.

[20] Gallant, A. R., M. Jahan-Parvar, and H. Liu, forthcoming, Does smooth ambiguity matter for asset pricing? *Review of Financial Studies.*

[21] Garlappi, L., R. Uppal, and T. Wang. 2007. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *Review of Financial Studies* 20:41–81.

[22] Ghirardato, P., and M. Marinacci, 2002, Ambiguity made precise: A comparative foundation, Journal of Economic Theory 102, 251–289.

[23] Ghirardato, P., F. Maccheroni, and M. Marinacci, 2004, Differentiating ambiguity and ambiguity aversion, *Journal of Economic Theory* 118, 133–173.

[24] Cerreia-Vioglio, S., P. Ghirardato, F. Maccheroni, M. Marinacci, and M. Siniscalchi, 2011, Rational preferences under uncertainty, Economic Theory 48, 341–375.

[25] Ghirardato P., F. Maccheroni, and M. Marinacci, 2004, Differentiating ambiguity and ambiguity attitude, Journal of Economic Theory 118, 133–173.

[26] Ghirardato, P., and M. Siniscalchi, 2012, Ambiguity in the small and in the large, Econometrica 80, 2827–2847.

[27] Gibbons, M. R., S. A. Ross, and J. Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

[28] Gilboa, I., and D. Schmeidler. 1989. Maxmin Expected Utility with Non-Unique Prior, *Journal of Mathematical Economics* 18, 141–153.

[29] Guetlein, M.-C., 2016, Comparative risk aversion in the presence of ambiguity, American Economic Journal: Microeconomics 8, 51–63.

[30] Halevy, Y. (2007). Ellsberg revisited: An experimental study. Econometrica, 75(2), 503-536.

[31] Hansen, L. P., 2007, Beliefs, doubts and learning: Valuing macroeconomic risk, *American Economic Review* 97, 1–30.

[32] Hansen, L. P., and T. J. Sargent, 2001, Acknowledging misspecification in macroeconomic theory, Review of Economic Dynamics 4, 519–535.

[33] Hara, C., and T. Honda, forthcoming, Implied ambiguity: mean-variance efficiency and pricing errors, *Management Science*.

[34] Hara, C., 2020, A ranking over the "more risk aversion" relations and its applications to the smooth ambiguity model, *KIER Discussion Paper Series* No. 1019, Institute of Economic Research, Kyoto University.

[35] Hardy, G. H., J. E. Littlewood, and G. Polya, 1952, *Inequalities*, 2nd edition, Cambridge University Press.

[36] Hirsch, M. W., 1976, Differential Topology, Springer-Verlag.

[37] Jahan-Parvar, M. R., and H. Liu, 2014, Ambiguity aversion and asset prices in production economies, *Review of Financial Studies* 27, 3060–3097.

[38] Ju, N., and J. Miao, 2012, Ambiguity, learning, and asset returns. *Econometrica* 80, 559–591.

[39] Klibanoff, P., M. Marinacci, and S. Mukerji. 2005. A smooth model of decision making under ambiguity, *Econometrica* 73, 1849–1892.

[40] Kocherlakota, N. R., 1996, The Equity Premium: It's Still a Puzzle, *Journal of Economic Literature* 34, 42–71.

[41] Maccheroni, F., M. Marinacci, and D. Ruffino. 2013. Alpha as Ambiguity: Robust Mean-Variance Portfolio Analysis. *Econometrica* 81:1075–1113.

[42] Maccheroni, F., M. Marinacci, and A. Rustichini, 2006, Ambiguity aversion, robustness, and the variational representation of preferences, *Econometrica* 74, 1447–1498.

[43] Mas-Colell, A., M. Whinston, and J. R. Green, 1995, Microeconomic Theory, Oxford University Press.

[44] Mehra, R. and E. C. Prescott, 1985, The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145–161.

[45] Pratt, J. W., 1964, Risk aversion in the small and in the large, *Econometrica* 32, 122–136.

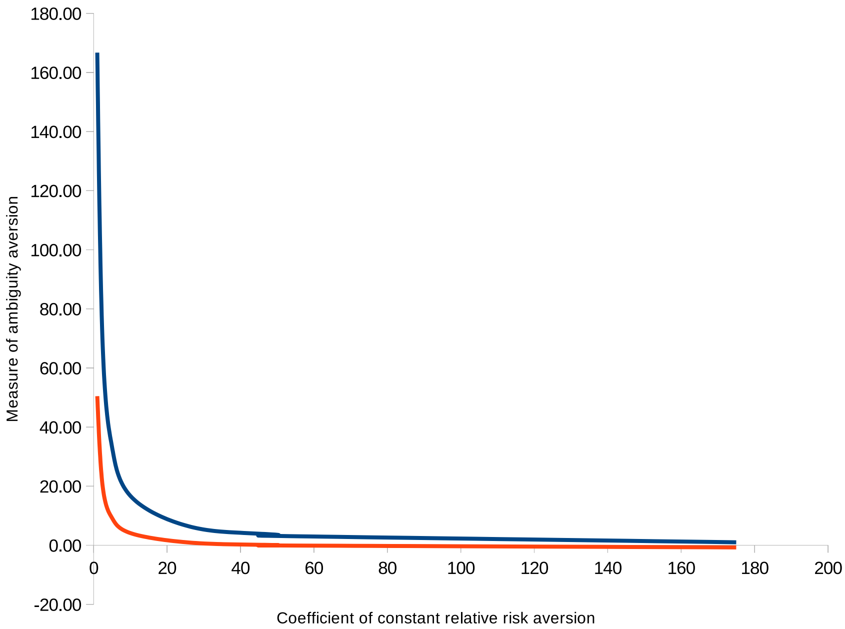
[46] Rigotti, L., C. Shannon, and T. Strzalecki, Subjective beliefs and ex ante trade, Econometrica 76, 1167–1190.

[47] Schmeidler, D. 1989, Subjective probability and expected utility without additivity, Econometrica 57, 571–587.

[48] Thimme, J., and C. Völkert, 2015, Ambiguity in the cross-section of expected returns: An empirical assessment, *Journal of Business and Economic Statistics* 33, 418–429.

[49] Wang, F., 2019, Comparative ambiguity aversion, manuscript, HEC Paris.

Figure 1: Measures of ambiguity aversion implied by the ambiguity premium and the equity premium



The blue curve is the graph of the measure of ambiguity aversion as a function the coefficient of relative risk aversion when the ambiguity premium is equal to ten percents of the expected consumption level. The red curve is the graph of the measure of ambiguity aversion as a function the coefficient of relative risk aversion when the equity premium is equal to 6.18%.