

Shadow Probability Theory and Ambiguity Measurement

Yehuda Izhakian^{*†}

January 9, 2012

Abstract

Ordering alternatives by their degree of ambiguity is a crucial element in decision processes in general and in asset pricing in particular. So far the literature has not provided an applicable measure of ambiguity allowing for such ordering. The current paper addresses this need by introducing a novel empirically applicable *ambiguity measure* derived from a new model of decision making under ambiguity, called *shadow probability theory*, in which probabilities of events are themselves random. In this model a complete distinction is attained between preferences and beliefs and between risk and ambiguity that enables the degree of ambiguity to be measured. The merits of the model are demonstrated by incorporating ambiguous probabilities into asset pricing and it is proved that the well-defined *ambiguity premium* that the paper proposes can be measured empirically.

Keywords: Ambiguity, Ambiguity Measure, Ambiguity Aversion, Knightian Uncertainty, Shadow Probability Theory, Choquet Expected Utility, Cumulative Prospect Theory, Ellsber Paradox, Ambiguity Premium.

JEL Classification Numbers: C44, C65, D81, D83, G11, G12.

^{*}Department of Economics, Stern School of Business, New York University, yud@stern.nyu.edu

[†]We benefited from discussions with Menachem Abudy, Yakov Amihud, John Asker, David Backus, Heski Bar-Isaac, Simon Benninga, Azi Ben-Rephael, Adam Brandenburger, Menachem Brenner, Luís Cabral, Allan Collard-Wexler, Ilan Cooper, Alex Cukierman, Itamar Drechsler, Ignacio Esponda, Xavier Gabaix, Itzhak Gilboa, Eitan Goldman, William Greene, Sergiu Hart, Saggi Katz, Ilan Kremer, Yacov Oded, Efe Ok, Thomas Sargent, David Schmeidler, Laura Veldkamp, Paul Wachtel, Avi Wohl and Stanley Zin. We would also like to thank the seminar participants at New York University, The Interdisciplinary Center (IDC) Herzliya and The Hebrew University of Jerusalem.

1 Introduction

How should uncertain alternatives be ranked by the criterion of ambiguity? Consider, for example, the following: a large urn contains 30 balls which are either black or yellow and a second smaller urn contains only 10 balls which are also either black or yellow. In both cases the proportions of black and yellow balls are unknown. Which of the following two bets is more ambiguous? "a ball drawn from the large urn is black" or "a ball drawn from the small urn is black."¹ Ordering different alternatives by their degree of ambiguity is an important element of any decision process, insofar as answering questions of this type is part of almost any real-life decision. However, so far the literature has not provided a useable measure of ambiguity that allows for such ordering. For example, the neoclassical finance literature, dealing with risk tolerance usually ignores the presence of ambiguity and unrealistically assumes that decision makers (DMs) are able to precisely evaluate the probability distribution of returns on assets. The main reason for this oversight is simply a lack of methods for ambiguity measurement. The goal of this paper is to provide a theoretical basis and applicable measure that addresses this need.

This paper makes three contributions to the existing literature. The first and the main contribution is that it introduces a novel, empirically applicable, *ambiguity measure*, underpinned by a new theoretical concept.² The second contribution is that it presents a decision-making model to derive this measure which achieves a complete distinction between preferences and beliefs and between risk and ambiguity, thereby enabling an exploration of the nature of ambiguity and the behavioral aspects of attitude toward ambiguity. The advantage of isolating ambiguity from risk and beliefs from preferences is twofold: first, the degree of ambiguity can be measured independently of preferences, especially in empirical studies; second, aspects of preferences concerning ambiguity can be easily monitored in behavioral studies. The third contribution is that it generalizes classical asset pricing theory to incorporate ambiguity, providing a well-defined mathematical formula for the ambiguity premium which is clearly distinguished from risk, and can be tested empirically.

Assuming that probabilities of events are themselves random, this paper introduces a novel

¹Formally, *risk* is defined as a situation in which the event to be realized is *a-priori* unknown, but the odds of all possible events are perfectly known. *Ambiguity*, or *Knightian uncertainty*, is the case where not only is the event to be realized *a-priori* unknown, but the odds of all possible events are also either not uniquely assigned or are unknown.

²Throughout this paper the term *ambiguity measure* is used literally and not formally.

model of decision making, called *shadow probability theory* (shadow theory for short), which aims to capture the multi-dimensional nature of uncertainty. In this model, there are two tiers of uncertainty, one with respect to consequences (outcomes) and the other with respect to the probabilities of these consequences. Each tier is modeled by a separate state space. This structure introduces a complete distinction of risk from ambiguity with regard to both beliefs and preferences. The degree of ambiguity and the DM's attitude toward it are then measured with respect to one space, while risk and risk attitude apply to the second space. As a consequence of random probabilities and probabilistic sensitivity (i.e., the nonlinear ways in which individuals may interpret probabilities) perceived probabilities are nonadditive. Ambiguity aversion results in a subadditive subjective probability measure, while ambiguity seeking results in a superadditive measure.

The main idea of shadow theory is that, like measuring the degree of risk by the variance of outcomes, the degree of ambiguity can be measured by the variance of probabilities.³ However, concerning the variance of probabilities, the question is: to the probability of which event is the variance applied? Given a classification of outcomes as a loss or as a gain, this paper proves that the degree of ambiguity can be measured by four times the variance of the probability of loss, which is equal to four times the variance of the probability of gain. Formally, our measure of ambiguity is given by

$$\mathcal{U}^2 = 4\text{Var}[P_L] = 4\text{Var}[P_G],$$

where P_L and P_G are the random probabilities of loss and gain, respectively, and the variance is taken with respect to second-order probabilities. The intuition behind this new measure is that ambiguity is caused by a perturbation of probabilities with respect to a meaningful reference point. Its main advantage is that it can be easily computed from the data and can be used in empirical tests (see, for example, Brenner and Izhakian (2011)).

Measuring the degree of ambiguity allows alternatives to be ranked by the criterion of ambiguity. It provides a way to address important questions that arise regarding the nature of ambiguity, in general, and the nature of the aggregate ambiguity of portfolios, in particular. The nature of ambiguity and the relationship between risk and ambiguity may shed some light on various puzzling financial phenomena. Notable examples are the fact that individuals tend to hold very small portfolios, 3-4 stocks (Goetzmann and Kumar (2008)), the equity premium

³Measuring risk by the variance of outcomes is admissible under some conditions; the same is true for measuring ambiguity by the variance of probabilities.

puzzle (Mehra and Prescott (1985)), the risk-free rate puzzle (Weil (1989)), the phenomenon of the observed equity volatility being too high to be justified by changes in the fundamentals (Shiller (1981)), and the home bias puzzle (Coval and Moskowitz (1999)).

To demonstrate the value of shadow theory and its measure of ambiguity, this paper generalizes asset pricing theory to incorporate ambiguity. Relaxing the assumption that probabilities are known, the price of an asset in our model is determined not only by its degree of risk and the DM's attitude toward this risk, but also by its degree of ambiguity and the DM's attitude toward it. The current paper constructs the *uncertainty premium* and proves that it can be separated into a *risk premium* and an *ambiguity premium*.⁴ It provides a well-defined ambiguity premium, completely distinguished from risk and attitude toward risk, which can be computed from the data. Using this model, in their empirical study Brenner and Izhakian (2011) show that ambiguity, measured by \mathcal{U}^2 , has a significant impact on the market portfolio's return. We are not aware of any other prior study that conducts direct empirical tests of models of decision making under ambiguity other than through parametric fitting and calibrations.⁵

Shadow theory relies on the *Choquet expected utility* (CEU) of Schmeidler (1989), whose axiomatic derivation paved the way for modeling decision making under ambiguity. Gilboa (1987) and Schmeidler (1989), in their pioneering studies, introduce the idea that, in the presence of ambiguity, the probabilities that reflect the DM's willingness to bet cannot be additive, i.e., the sum of the probabilities can be either smaller or greater than 1. Shadow theory combines the concept of nonadditive probabilities with the idea of *reference-dependent beliefs*. Reference dependency is applied to differentiate between the probability of gain and the probability of loss. It allows characterizing the DM's preferences concerning ambiguity, with regard to these obscure probabilities, with the different possible attitudes toward ambiguity when it concerns loss and when it concerns gain.

Tversky and Kahneman's (1992) *Cumulative prospect theory* (CPT) also applies a two-sided CEU to gains and to losses.⁶ However, CPT focuses on *reference-dependent preferences* and assumes a DM having different risk attitudes for losses and for gains, and asymmetric capacities

⁴This paper uses the term *uncertainty* to describe the aggregation of risk and ambiguity.

⁵Uppal and Wang (2003), Epstein and Schneider (2008), and Ju and Miao (2011), for example calibrate their model to the data. Several papers attribute different explanatory variables to ambiguity. For example, Anderson *et al.* (2009) attribute the degree of disagreement of professional forecasters to ambiguity, and Erbas and Mirakhor (2007) use the World Bank institutional quality indexes as a proxy for ambiguity.

⁶Using the perception of rank-dependent and cumulative functionals proposed by Weymark (1981), Quiggin (1982), Yaari (1987) and Schmeidler (1989), CPT generalizes the original prospect theory of Kahneman and Tversky (1979) from risk to uncertainty. It modifies the probability weighting functionals of the original prospect theory, such that it always satisfies stochastic dominance and supports an infinite state space.

with different arbitrary weighting schemes for losses and for gains.⁷ Shadow theory generalizes CPT and shows that capacities are not arbitrary and can be explained by the presence of ambiguity and the DM's preferences regarding it.⁸ It relies on the axiomatic foundation proposed by Wakker (2010) for both preferences concerning risk and preferences concerning ambiguity. Practically, shadow theory provides a bridge between the two main disciplines of modeling ambiguity: the multiple priors (Gilboa and Schmeidler (1989)) and the nonadditive probabilities (Gilboa (1987) and Schmeidler (1989)).

The rest of the paper is organized as follows. Section 2 introduces shadow theory. Section 3 applies this theory to a new decision-making model under uncertainty and characterizes DMs' attitudes toward ambiguity. Using this model, Section 4 suggests a novel measure of ambiguity. To demonstrate an application of shadow theory for asset pricing, Section 5 models the ambiguity premium and reviews some empirical results that test it. Section 6 discusses our theoretical results with respect to the related literature and Section 7 concludes.

2 Shadow probability theory

Shadow theory assumes two different tiers of uncertainty, one with respect to consequences and the other with respect to the probabilities of these consequences. Each tier is modeled by a separate state space. Uncertainty with respect to consequences (outcomes) is modeled by a *subordinated outcome space*, while uncertainty with respect to probabilities is modeled by a *directing probability space*. In this section, we model the subordinated space and then add the second tier of uncertainty, with respect to probabilities, by modeling the directing space. First, we demonstrate the concept of ambiguity that emerges from our model with an illustration.

2.1 Illustration

To illustrate the new insight gained by shadow theory about the nature of ambiguity and the way individuals perceive it, consider an Ellsberg urn with 90 colored balls, 30 of which are red and the other 60 either black or yellow. Ellsberg (1961) suggested the following two-part experiment. In each part of the experiment, before a ball is drawn at random from the urn,

⁷Capacities mean probabilities, possibly nonadditive. This paper usually uses the term probability in a broad sense, i.e., it can be nonadditive and either subjective or objective. The terms subjective probabilities and capacities are used interchangeably.

⁸To explain capacities, shadow theory does not assume asymmetric risk attitude, different ambiguity preferences for losses and for gains, or loss aversion.

a DM is offered two alternative bets; winning the bet entitles her to a sum of \$9. In the first part, the DM has to choose between two alternative bets: the ball drawn is red (R) or the ball drawn is black (B). Then, in the second part, the DM has to choose between betting on: the ball drawn is red or yellow (RY) or alternatively the ball drawn is black or yellow (BY). Behavioral experiments have demonstrated that individuals usually prefer R over B , but BY over RY .⁹

Shadow theory can help us to understand the choices that emerge from these experiment. The subordinated space is defined by the states of drawing different balls, i.e., $\{R, Y, B\}$, with random probabilities. The probability of B can be one of the possible values $\frac{0}{90}, \frac{1}{90}, \frac{2}{90}, \dots, \frac{60}{90}$. The precise probability of B is dominated by a second-order unobservable event in a directing space. Such an event can be, for example, "The experimenter put 30 red balls, 20 black balls and 40 yellow balls in the urn." The DM, however, does not have any additional information indicating which of the possible probabilities is more likely and thus she assigns an equal weight to each possibility.

The DM considers strictly positive outcomes (\$9) as a gain; and otherwise, (\$0) as a loss. Computing the variance of the probability of gain from B to obtain the measure of ambiguity indicates that its degree of ambiguity is $\mathcal{U}^2[B] = 0.1530$. Since the probability of R is known, $\frac{1}{3}$, its degree of ambiguity is $\mathcal{U}^2[R] = 0$. Clearly, since the expected outcomes of R and B are identical, an ambiguity-averse DM prefers R , with the lower degree of ambiguity, over B . The probability of gain from RY can take one of the possible values $\frac{30}{90}, \frac{31}{90}, \dots, \frac{90}{90}$, which in turn also implies an ambiguity degree of $\mathcal{U}^2[RY] = 0.1530$. The probability of gain from the alternative BY is exactly $\frac{2}{3}$, which implies $\mathcal{U}^2[BY] = 0$. Obviously, an ambiguity-averse DM prefers BY over RY . Table 1 is a stylized description of this example.

	R	Y	B	Prob	E	\mathcal{U}^2
(R)	9	0	0	$\frac{1}{3}$	3	0
(B)	0	0	9	$\frac{0}{90}, \frac{1}{90}, \frac{2}{90}, \dots, \frac{60}{90}$	3	0.1530
(RY)	9	9	0	$\frac{30}{90}, \frac{31}{90}, \frac{32}{90}, \dots, \frac{90}{90}$	6	0.1530
(BY)	0	9	9	$\frac{2}{3}$	6	0

Table 1: **The Ellsberg example**

⁹In expected utility theory, the DM's assessments of the likelihoods of R , B and Y can be described by some probability measure P . The DM is assumed to prefer a greater chance of winning \$9 to a smaller chance of winning \$9, such that the choices above imply that $P(R) > P(B)$ and $P(B \cup Y) > P(R \cup Y)$. However, since R , B and Y are mutually exclusive events, no such probability measure exists; hence, it is considered a paradox.

2.2 The subordinated space

Observable events and their consequences are defined by the subordinated space. The probabilities of these events are uncertain; as a consequence, perceived probabilities are nonadditive. The foundation of Schmeidler's (1989) CEU and Tversky and Kahneman's (1992) CPT of also assume that priors are nonadditive, but do not characterize the sources shaping these priors. The structure of the subordinated space relies on this foundation, while the reasoning for nonadditive priors is provided later by the directing space.

Let \mathbb{S} be a (finite or infinite) state space, called the *subordinated space*, endowed with an algebra of subsets of \mathbb{S} .^{10,11} It is assumed that exactly one state can be realized, which is unknown to the DM when she makes her choice. Subsets of \mathbb{S} are called *events* and are denoted by \mathcal{E} . The *complementary event*, \mathcal{E}^c , consists of all states $s \in \mathbb{S}$ not contained in \mathcal{E} . The set of events, Ξ , is a σ -algebra of subsets of \mathbb{S} . Define X to be the set of *consequences*, where, since this paper mostly deals with monetary outcomes, consequences are confined to real numbers, $X \subseteq \mathbb{R}$.¹²

An *act* is a function from states into consequences, $f : \mathbb{S} \rightarrow X$, describing the resulting consequence associated with each state $s \in \mathbb{S}$. The set of all possible acts is denoted \mathcal{F} . An act $f \in \mathcal{F}$ is represented as a sequence of $j = 1, \dots, n$ pairs

$$f = (\mathcal{E}_1 : x_1, \dots, \mathcal{E}_j : x_j, \dots, \mathcal{E}_n : x_n),$$

where x_j is the consequence if event \mathcal{E}_j occurs, i.e., x_j is the outcome under each state $s \in \mathcal{E}_j$, and $(\mathcal{E}_1, \dots, \mathcal{E}_n)$ is a partition of the state space \mathbb{S} . Acts, designating state-contingent consequences, are assumed to be equipped with *complete sign-ranking* with respect to outcomes,

$$x_1 \leq \dots \leq x_{k-1} \leq x_k \leq x_{k+1} \leq \dots \leq x_n,$$

for some $1 \leq k \leq n$. That is, all consequences are ranked not only with respect to each other, but also with respect to the neutral consequence x_k , called the *reference point*.¹³ Sometimes, when the context is clear, the act f with a vector of outcomes (x_1, \dots, x_n) is referred to as a random variable, possibly without specified probabilities, and designated $x_j = f(s_j)$ by f_j .

¹⁰To simplify our exposition, whenever possible our results are proved in static discrete settings; however all of the presented results can be applied to dynamic continuous settings.

¹¹Following Wakker and Tversky (1993), Wakker (2010, Appendix G) and Kothiyal *et al.* (2011), the state space, \mathbb{S} , can consist of an infinite number of states.

¹²Both sets \mathbb{S} and X are assumed to be nonempty.

¹³It is common to assume that the reference point is the status quo, exogenously given. In the financial world, when consequences are returns rather than quantitative outcomes, a natural objective reference point can possibly be 0 or the risk-free rate.

All consequences $x \in X$ are interpreted either as a gain or as a loss with respect to the designated reference point $x_k \in X$. Any consequence $x_j \in X$ is a *loss* if $x_j \leq x_k$ and a *gain* if $x_k < x_j$. The cumulative events of loss and gain are, thus, defined by $L = \mathcal{E}_1 \cup \dots \cup \mathcal{E}_k$ and $G = \mathcal{E}_{k+1} \cup \dots \cup \mathcal{E}_n$, respectively. To shorten our notations the conventions $\mathcal{E}_J = \mathcal{E}_1 \cup \dots \cup \mathcal{E}_j$ and $\mathcal{E}_{J\dots T} = \mathcal{E}_j \cup \dots \cup \mathcal{E}_t$ are used to denote cumulative events.¹⁴

Capacities are the DM's *subjective probabilities* of cumulative events. A *capacity* Q is a function on $2^{\mathbb{S}}$ assigning each event $A \subseteq \mathbb{S}$ with a number $Q(A)$, satisfying $Q(\emptyset) = 0$, $Q(\mathbb{S}) = 1$ and if $A \subset B \subset \mathbb{S}$ then $0 \leq Q(A) \leq Q(B)$; that is, capacities are monotonic. The probability of any loss event $1 \leq j \leq k$ is defined by

$$\pi_j^- = Q(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_j) - Q(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_{j-1})$$

and the probability of any gain event $k+1 \leq j \leq n$ is defined by

$$\pi_j^+ = Q(\mathcal{E}_j \cup \dots \cup \mathcal{E}_n) - Q(\mathcal{E}_{j+1} \cup \dots \cup \mathcal{E}_n),$$

where $\mathcal{E}_0 = \mathcal{E}_{n+1} = \emptyset$. It is important to note that the probabilities, π , do not necessarily add up to unity.¹⁵ Because $Q(\cdot)$ can be nonlinear, π need not be additive. That is, if $A \cap B = \emptyset$ usually $Q(A \cup B) \neq Q(A) + Q(B)$.¹⁶

The domain of preference relation, \succsim , is the nondegenerated set of acts \mathcal{F} and the relations $\prec, \succ, \sim, \precsim, \succsim$ are defined as usual. An act yielding the same consequence for any state $s \in \mathbb{S}$ is called a *constant act* and is designated by its constant consequence $x \in X$. The *certainty equivalent (CE)* of an act $f \in \mathcal{F}$ is a constant act, $x \in \mathcal{F}$, such that $f \sim x$.

Let V be a real function $V : \mathcal{F} \rightarrow \mathbb{R}$ assigning to each act $f \in \mathcal{F}$ a value $V(f)$ such that

$$V(f) = \sum_{j=1}^k [Q(\mathcal{E}_{1\dots j}) - Q(\mathcal{E}_{1\dots j-1})] U(x_j) + \sum_{j=k+1}^n [Q(\mathcal{E}_{j\dots n}) - Q(\mathcal{E}_{j+1\dots n})] U(x_j), \quad (1)$$

where $U : X \rightarrow \mathbb{R}$ is a strictly increasing continuous *utility function* satisfying $U(x_k) = 0$. Similarly, the value of an act with an infinite support is defined by¹⁷

$$V(f) = - \int_{-\infty}^k Q(\{s \in \mathbb{S} \mid U(f_s) < t\}) dt + \int_k^{\infty} Q(\{s \in \mathbb{S} \mid U(f_s) > t\}) dt. \quad (2)$$

Assume that the preference relation, \succsim , on the set of acts \mathcal{F} is *truly mixed* and satisfies: *weak ordering, monotonicity, continuity, gain-loss consistency, sign-tradeoff consistency*. Wakker

¹⁴Similarly, these notational conventions are applied to other variables, such as $P_J = P(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_j)$ and $P_{J\dots T} = P(\mathcal{E}_j \cup \dots \cup \mathcal{E}_t)$.

¹⁵To save on notations, π refers to both π^- and π^+ .

¹⁶Similarly, capacities can be applied directly to consequences, without specifying an underlying state space, i.e., $Q(\underline{c} \leq x \leq \bar{c})$, where \underline{c} and \bar{c} are constants; see Jaffray (1989).

¹⁷Considering an infinite support, notations are abused and k stands for x_k .

(2010, Theorem 12.3.5) proves that \succsim can be represented by the real function V such that $V(f) \geq V(g)$ iff $f \succsim g$, for any $f, g \in \mathcal{F}$.¹⁸

Caution 2.1. To apply the Choquet integration, $\int_{\mathbb{S}} f dQ$, over negative values it is assumed that there exists a constant value $C \geq |x_1|$ such that $g_j = f_j + C \geq 0$ for any event $\mathcal{E}_j \in \Xi$ and the integration, thus, takes the form $\int_{\mathbb{S}} g dQ - C$.

The utility function, $U(\cdot)$, characterizes the DM's preferences toward risk. A concave function implies risk aversion and a convex function implies risk loving. The utility function $U(\cdot)$ is unique up to an affine transformation and takes the form $-U(-x)$ for negative outcomes.

The concept of ambiguity aversion asserts that individuals are sensitive to probabilities. Moreover, as subjective probabilities, Q , are shaped by the nonlinear ways in which individuals process them they are nonadditive. The next section models the way individuals perceive probabilities and the impact of ambiguity and attitude toward it on perceived probabilities.

2.3 The directing space

Probabilities over the subordinated space are assumed to be random and dominated by unobservable events in a separate latent space, called a *directing space*. The meaning of *shadow probability* arises from the randomness of probabilities themselves. While making her choice, the DM does not know which event will be realized, either in the subordinated space or in the directing space, which means that she knows neither the realized outcome nor the realized probabilities of outcomes.

Objective probabilities P of events $\mathcal{E} \in \Xi$ occurring in the subordinated space are random and governed by directing events ε in a (finite or infinite) state space Ω , called the *directing space*. A *directing event* $\varepsilon = \{\omega_1, \dots, \omega_t\}$ is a subset of Ω , where $\omega_1, \dots, \omega_t \in \Omega$ are *directing states*. The set of directing events Σ is a σ -algebra of subsets of Ω and $(\varepsilon_1, \dots, \varepsilon_m)$ is a partition of the directing space Ω , i.e., $\varepsilon_i \cap \varepsilon_l = \emptyset, \forall i \neq l$. A consequence of event ε_i is an additive probability measure P_i over the subordinated space \mathbb{S} , where P_i stands for $P(\cdot \mid \varepsilon_i)$. \mathbb{P} denotes the set of consequential probability measures. In a finite subordinated space the probability measure P_i takes the form of a probability vector $P_i = (P_{i,1}, \dots, P_{i,j}, \dots, P_{i,n})$, assigning to each subordinated event its probability, where $P_{i,j}$ stands for $P(\mathcal{E}_j \mid \varepsilon_i)$.

A *directing act*, \hat{f} , is a function from the directing space into the set of probability measures,

¹⁸Alternative axiomatizations that could be adopted are Tversky and Kahneman (1992), Wakker and Tversky (1993), Chew and Wakker (1996) and Kothiyal *et al.* (2011).

$\hat{f} : \Omega \rightarrow \mathbb{P}$, describing the resulting probability measure associated with each directing state $\omega \in \Omega$. The *set of directing acts* is denoted $\widehat{\mathcal{F}}$. For simplicity's sake, when possible, directing acts are assumed to take a finite number of values. A directing act $\hat{f} \in \widehat{\mathcal{F}}$ is defined by a sequence $i = 1, \dots, m$ of pairs

$$\hat{f} = (\varepsilon_1 : P_1, \dots, \varepsilon_i : P_i, \dots, \varepsilon_m : P_m),$$

where P_i is the probability measure of event ε_i occurring. A directing act induces an *event-wise directing act* which assigns each subordinated event \mathcal{E}_j with its possible probabilities. That is,

$$\hat{f}_j = (\varepsilon_1 : P_{1,j}, \dots, \varepsilon_i : P_{i,j}, \dots, \varepsilon_m : P_{m,j}),$$

where \hat{f}_j stands for $\hat{f}(\mathcal{E}_j)$. The *set of event j 's directing acts* is denoted $\widehat{\mathcal{F}}_j$. When the context is clear, \hat{f}_j is referred to as directing act and the index j designates the event. In this context, \hat{f}_j is simply considered as a random variable formulating the random probability of event \mathcal{E}_j . A *constant directing act* associates any directing event with the same probability measure P . A subordinated act f associated with a constant directing act can be viewed as a *roulette lottery*, i.e., all probabilities are perfectly known.

A second-order capacity (second-order probability) $\chi(\cdot)$ on 2^Ω assigns to each directing event $\varepsilon \subseteq \Omega$ a number $\chi(\varepsilon)$, satisfying $\chi(\emptyset) = 0$, $\chi(\Omega) = 1$, and if $A \subset B \subset \Omega$ then $0 \leq \chi(A) \leq \chi(B)$. It is important to note that $\chi(\cdot)$ need not be additive; However, to save on notations and without loss of generality, it is considered to be additive such that $\sum_{i=1}^m \chi_i = 1$, where χ_i denotes $\chi(\varepsilon_i)$. Figure 1 provides a diagrammatic representation of the two spaces: the subordinated space and the directing space, and the relation between them.

Any subordinated act $f \in \mathcal{F}$ is associated with a directing act $\hat{f} \in \widehat{\mathcal{F}}$. The DM's preferences can be considered as if they are applied to pairs (f, \hat{f}) , i.e., preferences over a subset of $\mathcal{F} \times \widehat{\mathcal{F}}$. However, as explained in the next section, decisions in our model are made in two separate phases, such that these preference can be broken down into two independent preferences: \succsim over \mathcal{F} and \succsim^2 over $\widehat{\mathcal{F}}$. We conclude this subsection by relating our model to a more Savage-style structure. The objects of choice in Savage-type theory would be all measurable functions from the product space $\mathbb{S} \times \Omega$ to the outcome space X .¹⁹ Preferences over these objects of choice (Savage-acts) are then taken as primitive in this framework.

¹⁹The product space can alternatively be defined by $\mathbb{S} \times \mathbb{P}$.

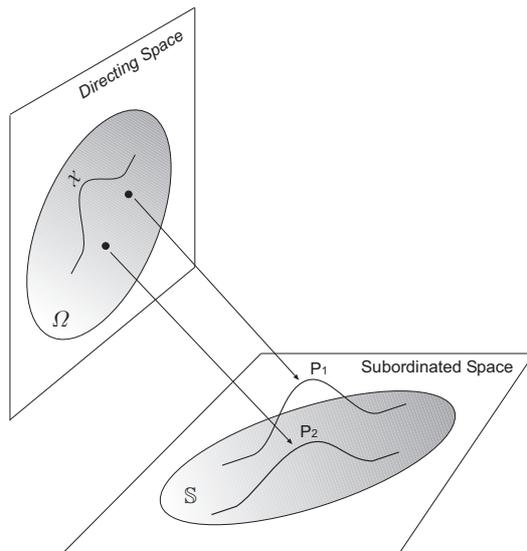


Figure 1: **The two spaces**

3 Ambiguity

The main idea at the basis of our model is that risk and risk preferences apply to the subordinated space, whereas ambiguity and preferences concerning it apply to the directing space. The model assumes two differentiated phases in the decision-making process: *probability framing* and *valuation*. In the framing phase, based on the information she has and her preferences concerning ambiguity, the DM forms a representation of her subjective probabilities for all the events which are relevant to her decision. Then, in the valuation phase, based on her preferences concerning risk, the DM assesses the value of each act and chooses accordingly. Separating the decision process to two sequential phases, where each phase is related to a different space, allows us to distinguish between risk and ambiguity and between the preferences concerning them. This section concentrates on the implications of the directing space for ambiguity and preferences concerning it along the process of framing probabilities.

3.1 Attitudes toward ambiguity

The DM is assumed to have a (second-order) preference relation, \succsim^2 , over the set of directing acts, $\widehat{\mathcal{F}}$. Although the DM does not have a direct choice over directing acts, her subjective perception of likelihoods, resulting from an aversion to or love of ambiguity, derives from the nature of the directing acts. Directing acts are latent and cannot be chosen independently of subordinated acts. As a result second-order preferences concerning the directing acts are

unobservable. In some decision problems, however, second-order preferences can be inferred from observable choices over acts in the subordinated space, for example, when the DM chooses between two alternatives with identical subordinated acts but different directing acts.

It is important to emphasize that even though preferences with respect to the directing acts are not observable they are not as unfamiliar as they might first appear. For example, in the Ellsberg urn experiment, directing acts may be considered as bets on the composition of the urn. In portfolio investment in mean-variance space, directing acts can be viewed as bets on the means, variances and covariances of the investment opportunities (see Izhakian (2011)). Similarly, directing acts in model uncertainty applications can be considered as bets about the true values of the parameters of the model.

Attitude toward ambiguity can be characterized by the DM's preferences over random probabilities of events and their expected probabilities, i.e., preferences over directing acts and constant directing acts. An ambiguity-averse DM prefers the expectations of the random probabilities over the random probabilities themselves. An ambiguity-loving DM prefers the random probabilities over their expectations and an ambiguity-neutral DM is indifferent between them. The next definition settles this idea formally.

Definition 3.1. *Let a directing act be $\hat{f}_j = (\varepsilon_1 : P_{1,j}, \dots, \varepsilon_m : P_{m,j})$ and its related constant directing act be $\bar{f}_j = (\varepsilon_1 : E[P_j], \dots, \varepsilon_m : E[P_j])$, where $E[P_j] = \sum_{i=1}^m \chi_i P_{i,j}$ is the expected probability of event \mathcal{E}_j with respect to the second-order capacities χ . Ambiguity aversion (loving) as regards event \mathcal{E}_j is defined by $\bar{f}_j \succsim^2 \hat{f}_j$ ($\hat{f}_j \succsim^2 \bar{f}_j$) and ambiguity neutrality is defined by $\bar{f}_j \sim^2 \hat{f}_j$.²⁰*

This definition focuses on preferences concerning ambiguity that related to single events. Section 4 aggregates these preferences to preferences over entire acts. A DM is said to be ambiguity averse if $\bar{f}_j \succsim^2 \hat{f}_j$ for any $\mathcal{E}_j \in \Xi$. Ambiguity-loving DMs and ambiguity-neutral DMs are defined similarly. Definition 3.1 can be applied to formulate different attitudes toward ambiguity for different subsets of events; for example, ambiguity loving for losses and ambiguity aversion for gains. Although it allows the flexibility of DMs having different attitudes for different events, it is assumed that the DM is consistent with respect to her attitudes toward ambiguity and, thus, her attitude might vary only across events of gain and loss.

It is important to note that ambiguity and attitude toward it consider only the probabilities and not consequences. The type and the magnitude of a consequence resulting from an event are

²⁰Strict preferences toward ambiguity can be defined similarly.

related neither to subjective preferences toward ambiguity nor to objective degree of ambiguity. To illustrate this notion consider an event with unknown probability of losing \$100. Changing the magnitude of this loss to \$10 or to \$1000 does not have any effect on the degree of ambiguity.

The next theorem ties the DM's preferences concerning ambiguity to a functional representation.

Theorem 3.2. *Let $v : \widehat{\mathcal{F}}_j \rightarrow \mathbb{R}$ be a real function assigning a value $v(\hat{f}_j)$ to each directing act $\hat{f}_j \in \widehat{\mathcal{F}}_j$, such that*

$$v(\hat{f}_j) = \sum_{i=1}^m [\chi(\varepsilon_1 \cup \dots \cup \varepsilon_i) - \chi(\varepsilon_1 \cup \dots \cup \varepsilon_{i-1})] \psi(P_{i,j}),$$

where $\varepsilon_0 = \emptyset$ and $\psi : [0, 1] \rightarrow \mathbb{R}$ is a strictly increasing continuous function.²¹ Assume that the preference relation, \succsim^2 , on the set of acts $\widehat{\mathcal{F}}_j$ satisfies: weak ordering, monotonicity, continuity and sign-tradeoff consistency. The preference \succsim^2 can then be represented by the function v , such that $v(\hat{f}_j) \geq v(\hat{g}_j)$ iff $\hat{f}_j \succsim^2 \hat{g}_j$, for any $\hat{f}_j, \hat{g}_j \in \widehat{\mathcal{F}}_j$.

The proof of Theorem 3.2 is directly derived by Wakker (2010, Theorem 12.3.5).²² The function $\psi(\cdot)$, referred to as a *probability-sake* (*sake* for short) function, forms the DM's attitude toward ambiguity. Similarly to preferences concerning risk, there are three types of preferences concerning ambiguity: ambiguity aversion, ambiguity loving and ambiguity neutrality. Ambiguity neutrality takes the form of a linear sake function $\psi(\cdot)$, ambiguity aversion the form of a concave sake function and ambiguity loving the form of a convex sake function.

Two special types of preferences concerning ambiguity can be defined. *Constant relative ambiguity aversion* (CRAA), which takes the functional form $\psi(P) = \frac{P^{1-\eta}}{1-\eta}$, and *constant absolute ambiguity aversion* (CAAA), which takes the functional form $\psi(P) = -\frac{e^{-\eta P}}{\eta}$, where η is the coefficient of ambiguity aversion. With this notion of ambiguity and preferences concerning it, we suggest the following definition of subjective probabilities.

Model 3.3. *The **subjective probability** $Q(\mathcal{E}_j)$ of any event $\mathcal{E}_j \in \Xi$ is formed by²³*

$$Q(\mathcal{E}_j) = \psi^{-1} \left(\sum_{i=1}^m \chi(\varepsilon_i) \psi(P(\mathcal{E}_j | \varepsilon_i)) \right). \quad (3)$$

It can be easily observed that, in fact, Equation (3) models the *certainty equivalent*, in terms of probabilities, of the random probability of a subordinated event \mathcal{E}_j . The subjective proba-

²¹This theorem provides the general form where capacities, χ , are nonadditive.

²²Preferences \succsim^2 and their utility representation over the directing space are degenerated in the sense that no distinction of losses from gains is made, such that true mixing and gain-loss consistency are not enforced.

²³Considering a directing space with an infinite support, subjective probabilities are formed in a manner similar to Equation (2).

bilities of an ambiguity-neutral DM are equal to the expected probabilities of each event, taken with respect to the second-order probabilities, i.e., $Q(\mathcal{E}_j) = E[P_j]$. Probabilities, in this case, are additive and the certainty-equivalent probabilities are equal to the expected probabilities. The subjective probabilities of an ambiguity-averse DM are lower than the expected probabilities, i.e., $Q(\mathcal{E}_j) < E[P_j]$, which results in subadditive probabilities. The subjective probabilities of an ambiguity-loving DM are greater than the expected probabilities, i.e., $Q(\mathcal{E}_j) > E[P_j]$, which results in superadditive probabilities.

The subjective probabilities as formed by Model 3.3 provide a reason for the nonadditivity, which is arbitrary in CEU and CPT. This result coincides with the nonadditive priors of Gilboa (1987) and Schmeidler (1989). Recall that in shadow theory, before evaluating any subordinated act $f \in \mathcal{F}$ associated with the subordinated space, \mathbb{S} , the DM's preferences over directing acts, $\widehat{\mathcal{F}}$, associated with the directing space, Ω , serve to shape her subjective probabilities, Q , assigned to each subordinated event $\mathcal{E}_j \in \Xi$. Formally, given the subjective probabilities of Equation (3), they are integrated into Equation (1) to obtain the value of an act f .

Model 3.4. *In shadow theory, the value of any act $f \in \mathcal{F}$ with a finite support is*

$$V(f) = \sum_{j=1}^k \left[\psi^{-1} \left(\sum_{i=1}^m \chi_i \psi(P_{i,1\dots J}) \right) - \psi^{-1} \left(\sum_{i=1}^m \chi_i \psi(P_{i,1\dots J-1}) \right) \right] U(x_j) + \sum_{j=k+1}^n \left[\psi^{-1} \left(\sum_{i=1}^m \chi_i \psi(P_{i,J\dots N}) \right) - \psi^{-1} \left(\sum_{i=1}^m \chi_i \psi(P_{i,J+1\dots N}) \right) \right] U(x_j),$$

and the value of any act with an infinite support is

$$V(f) = - \int_{-\infty}^k \psi^{-1} \left(\sum_{i=1}^m \chi_i \psi(P_i(\{s \in \mathbb{S} \mid U(f) < t\})) \right) dt + \int_k^{\infty} \psi^{-1} \left(\sum_{i=1}^m \chi_i \psi(P_i(\{s \in \mathbb{S} \mid U(f) > t\})) \right) dt.$$

The functional representation of the DM's aggregate preferences, proposed by Model 3.4, makes a complete distinction between beliefs and preferences and between risk and ambiguity. First-order beliefs are formed by the random probability measures $P_{i=1,\dots,m}$; second-order beliefs are formed by the probability measure χ ; preferences concerning risk are formed by the utility function $U(\cdot)$; and preferences concerning ambiguity are formed by the sake function $\psi(\cdot)$. The separation attained by Model 3.4 allows for isolating and studying the distinct impact of each component on values of acts. However, even more important, it enables the simplification of this model to an applicable form such that the degree of ambiguity can be measured, as proposed later. This simplification paves the way for empirical studies to test the impact of

ambiguity using the data.

Risk and ambiguity preferences, in Model 3.4, can be different for losses and for gains. Thus, the functional representation for losses can take the form $U(\cdot) = U^-(\cdot)$ and $\psi(\cdot) = \psi^-(\cdot)$, and for gains $U(\cdot) = U^+(\cdot)$ and $\psi(\cdot) = \psi^+(\cdot)$. Different utility functions, $U^-(\cdot)$ and $U^+(\cdot)$, can capture, for example, *loss aversion*.²⁴ Different sake functions, $\psi^-(\cdot)$ and $\psi^+(\cdot)$, can capture, for example, ambiguity loving for losses and ambiguity aversion for gains.²⁵ Pessimism and optimism can also be incorporated into Model 3.4 by the distortions $w^-(Q)$ and $w^+(Q)$ of the subjective probabilities Q of losses and gains, respectively. *Pessimism* holds if worsening the rank increases the weighting assigned by $w(\cdot)$, i.e., bigger weights are assigned for worse ranks.²⁶ *Optimism* holds if improving the rank increases the weighting assigned by $w(\cdot)$, i.e., bigger weights are assigned for better ranks. In principle, a DM in this extension can exhibit ambiguity aversion while still having an optimistic perception of likelihoods.

3.2 Ambiguous probabilities

The goal of this section is to simplify Model 3.4 to a friendlier applicative form. Using this simplified form, it is then proved that the ambiguity associated with an event can be measured by the variance of its probability. For these purposes the following notations are introduced.

Let

$$p_j = E[P_j] = \sum_{i=1}^m \chi_i P_{i,j}$$

be the expected probability of event \mathcal{E}_j and

$$\zeta_j^2 = \text{Var}[P_j] = \sum_{i=1}^m \chi_i (P_{i,j} - p_j)^2$$

be its variance. The covariance of the probability of two events \mathcal{E}_j and \mathcal{E}_l is defined by

$$\zeta_{j,l} = \text{Cov}[P_j, P_l] = \sum_{i=1}^m \chi_i (P_{i,j} - p_j)(P_{i,l} - p_l).$$

Now, to simplify the exposition of Model 3.4 and make it more applicable, subjective probabilities, $Q(\cdot)$, are approximated by taking a second-order Taylor approximation around their expectations.²⁷ Since this approximation deals with probabilities, a condition on the sake func-

²⁴Loss aversion is modeled by a steeper utility function for losses than for gains (see, for example, Barberis and Huang (2001)). Some behavioral studies report risk aversion for gains and risk seeking for losses (see, for example, Wang and Fischbeck (2004) and Abdellaoui *et al.* (2008)).

²⁵Some behavioral study document ambiguity loving for losses and ambiguity aversion for gains (see, for example, Dobbs (1991) and Bier and Connell (1994)).

²⁶To simplify notation, the weighting function w is used in reference to both w^- and w^+ .

²⁷The same method is applied by Arrow (1965) and Pratt (1964) to outcomes within the expected utility

tion $\psi(\cdot)$ is enforced to assure that the approximated subjective probabilities are nonnegative.

Theorem 3.5. *Assume a continuous twice-differentiable sake function $\psi(\cdot)$, satisfying*

$$\frac{1}{2} \left(\frac{\psi''(p_l)}{\psi'(p_l)} \zeta_l^2 - \frac{\psi''(p_{j \cup l})}{\psi'(p_{j \cup l})} \zeta_{j \cup l}^2 \right) \leq p_j, \quad (4)$$

for any events $\mathcal{E}_j, \mathcal{E}_l \in \Xi$, where $j \cup l$ stands for $\mathcal{E}_j \cup \mathcal{E}_l$. The subjective probability of event \mathcal{E}_j is then

$$Q(\mathcal{E}_j) \approx p_j + \frac{1}{2} \frac{\psi''(p_j)}{\psi'(p_j)} \zeta_j^2.$$

Similarly, the subjective probability of a cumulative event $\mathcal{E}_{J \dots T} = \mathcal{E}_j \cup \dots \cup \mathcal{E}_t$ is $Q(\mathcal{E}_{J \dots T}) \approx p_{J \dots T} + \frac{1}{2} \frac{\psi''(p_{J \dots T})}{\psi'(p_{J \dots T})} \zeta_{J \dots T}^2$. Theorem 3.5 characterizes the DM's capacities, which satisfy $Q(\emptyset) = 0$ and $Q(\mathbb{S}) = 1$. Lemma A.1 proves that if $A \subset B \subset \mathbb{S}$ then $Q(A) \leq Q(B)$ and, thus, $Q(\cdot)$ is a capacity. Condition (4) implies that the coefficient of ambiguity aversion, $-\frac{\psi''(\cdot)}{\psi'(\cdot)}$, is bounded, such that $-2\frac{p_j}{\zeta_j^2} \leq -\frac{\psi''(p_j)}{\psi'(p_j)} \leq 2\frac{p_j}{\zeta_j^2}$ and $-2\frac{p_j}{\zeta_j^2 + 2\zeta_{j,l}} \leq -\frac{\psi''(p_j)}{\psi'(p_j)} \leq 2\frac{p_j}{\zeta_j^2 + 2\zeta_{j,l}}$. It restricts the level of ambiguity aversion (the concavity of ψ) such that the marginal probability premium (defined below) of an event must be lower than the marginal expected probability of that event. This condition assures that the approximated probability, $Q(\cdot)$, is nonnegative and that the probability of an event is not lower than the probability of any of its sub-events. This condition, however, is required only for the purpose of approximation and it is not enforced over the precise subjective probabilities defined in Equation (3). Henceforth it is assumed that the sake function $\psi(\cdot)$ falls under Condition (4).

The following definition establishes a new terminology.

Definition 3.6. *The expression*

$$\varphi_j = -\frac{1}{2} \frac{\psi''(p_j)}{\psi'(p_j)} \zeta_j^2$$

is referred to as the **probability premium** of event \mathcal{E}_j and ζ_j^2 is referred to as the **ambiguity of event j** (**e-ambiguity** for short). The expression $-\frac{\psi''(p_j)}{\psi'(p_j)}$ is referred to as the coefficient of absolute ambiguity aversion and $-p_j \frac{\psi''(p_j)}{\psi'(p_j)}$ is referred to as the coefficient of relative ambiguity aversion.²⁸

The probability premium is composed of two components: the DM's preferences concerning ambiguity, framed by $-\frac{\psi''(\cdot)}{\psi'(\cdot)}$, and the degree of e-ambiguity, measured by ζ^2 . Preferences

framework, whereas in this case it is applied to consequential probabilities. The method is restricted to the case of small consequences, and is applicable in the current model since consequences are probabilities ranging between 0 and 1.

²⁸These definitions are equivalent to the Arrow-Pratt coefficient of absolute risk aversion and coefficient of relative risk aversion, respectively.

concerning ambiguity can be aversion ($-\frac{\psi''(\cdot)}{\psi'(\cdot)} > 0$) loving ($-\frac{\psi''(\cdot)}{\psi'(\cdot)} < 0$) or neutrality ($-\frac{\psi''(\cdot)}{\psi'(\cdot)} = 0$). Clearly a higher ambiguity aversion or a higher degree of e-ambiguity implies a greater probability premium and a lower subjective probability. When the probability of an event is perfectly known, its e-ambiguity is zero, i.e., $\zeta^2 = 0$, and thus is the probability premium. When the DM is ambiguity neutral then $-\frac{\psi''(\cdot)}{\psi'(\cdot)} = 0$, which also implies a zero probability premium. Subadditive probabilities are obtained when the DM exhibits ambiguity aversion, i.e., $\frac{\psi''(\cdot)}{\psi'(\cdot)} < 0$, and superadditive probabilities when she exhibits ambiguity loving, i.e., $\frac{\psi''(\cdot)}{\psi'(\cdot)} > 0$. It is important to note that by Condition (4) the coefficient of absolute ambiguity aversion is bounded, such that for a high enough level of ambiguity aversion the subjective probability of each event tends to zero.

In general, the subjective probability measure, $Q(\cdot)$, is nonadditive, that is

$$Q(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_t \cup \mathcal{E}_{t+1} \cup \dots \cup \mathcal{E}_j) \neq Q(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_t) + Q(\mathcal{E}_{t+1} \cup \dots \cup \mathcal{E}_j).$$

This measure has an additive component, the expected probability

$$p(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_t \cup \mathcal{E}_{t+1} \cup \dots \cup \mathcal{E}_j) = p(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_t) + p(\mathcal{E}_{t+1} \cup \dots \cup \mathcal{E}_j),$$

with an additional nonadditive component, the probability premium. The source of nonadditivity is the e-ambiguity, measured by $\zeta_{1\dots j}^2$, which in most cases satisfies

$$\zeta_{1\dots J}^2 \neq \zeta_{1\dots T-1}^2 + \zeta_{T\dots J}^2.$$

The ambiguities of two events (e-ambiguities) are not independent.²⁹ Nonadditivity is caused by the correlation between the probabilities of events. Since e-ambiguity is the variance of the possible probabilities of the event, the e-ambiguity of a union of events comprises the covariances between the probabilities of its sub-events. For example, the e-ambiguity of a union of two events is the sum of the e-ambiguity of each event, plus twice the covariance between the probabilities of these events, as the next proposition proves.

Proposition 3.7. *For any $1 < t < j \leq n$, e-ambiguity satisfies*

$$\zeta_{1\dots J}^2 = \zeta_{1\dots T}^2 + \zeta_{T+1\dots J}^2 + 2\zeta_{1\dots T, T+1\dots J}.$$

A special case of e-ambiguity arises when considering an event \mathcal{E} and its complement event \mathcal{E}^c , since the e-ambiguity of their union equals zero, i.e., $\zeta_{\mathcal{S}}^2 = 0$. The next lemma applies to this case.

²⁹One can write $\phi(\mathcal{E}_j, \mathcal{E}_l) = \frac{1}{2} \left(\frac{\psi''(p_{j \cup l})}{\psi'(p_{j \cup l})} \zeta_{j \cup l}^2 - \frac{\psi''(p_l)}{\psi'(p_l)} \zeta_l^2 - \frac{\psi''(p_j)}{\psi'(p_j)} \zeta_j^2 \right)$ to obtain the *Möbius transform* of two events (see, for example, Chateauneuf and Jaffray (1989) and Grabisch *et al.* (2000)).

Lemma 3.8. *The covariance between the probability of event $\mathcal{E} \in \Xi$ and the probability of its complement event $\mathcal{E}^c \in \Xi$ satisfies*

$$\text{Cov} [P (\mathcal{E}), P (\mathcal{E}^c)] = -\text{Var} [P (\mathcal{E})] = -\text{Var} [P (\mathcal{E}^c)],$$

and thus $\zeta_{\mathcal{S}}^2 = 0$.

The results of Proposition 3.7 and Lemma 3.8 coincide with the findings of *support theory* of Tversky and Koehler (1994) and Rottenstreich and Tversky (1997). Support theory demonstrates that the judged probability of an event generally increases when its description is unpacked into disjoint components and decreases by unpacking the alternative description. One can easily conclude from Proposition 3.7 that, when $\zeta_{1\dots T, T+1\dots J} < 0$ and DMs are ambiguity averse, unpacking an event into disjoint components increases its probability. The conclusion arising from the behavioral findings of support theory is that the probabilities of events in the same state space are negatively correlated.

Now that we have studied some essential properties of e-ambiguity, Model 3.4 can be simplified such that the value of an act f takes the form

$$V(f) = \sum_{j=1}^k [p_j - \varphi_{1\dots J} + \varphi_{1\dots J-1}] U(x_j) + \sum_{j=k+1}^n [p_j - \varphi_{J\dots N} + \varphi_{J+1\dots N}] U(x_j), \quad (5)$$

where $\varphi_{J\dots T} = -\frac{1}{2} \frac{\psi''(p_{J\dots T})}{\psi'(p_{J\dots T})} \zeta_{J\dots T}^2$. Recall that attitude toward ambiguity can be different for gains and for losses. Thus, the coefficient of ambiguity attitude can take the form $-\frac{\psi''_L(p_{1\dots J})}{\psi'_L(p_{1\dots J})}$ for any loss event $1 \leq j \leq k$ and $-\frac{\psi''_G(p_{1\dots J})}{\psi'_G(p_{1\dots J})}$ for any gain event $k+1 \leq j \leq n$.

Example 3.9. *Assume an ambiguity-averse DM who exhibits CAAA. Since p is additive and $-\frac{\psi''(\cdot)}{\psi'(\cdot)} = \eta$, one can verify that the value of an act f in Equation (5) is simplified to*

$$V(f) = \sum_{j=1}^n p_j U(x_j) - \frac{1}{2} \eta \sum_{j=1}^k [\zeta_{1\dots J}^2 - \zeta_{1\dots J-1}^2] U(x_j) - \frac{1}{2} \eta \sum_{j=k+1}^n [\zeta_{J\dots N}^2 - \zeta_{J+1\dots N}^2] U(x_j).$$

The first component of the value function is the conventional expected utility. The last two components are the disutility caused to an ambiguity-averse DM by the presence of ambiguity.

If the DM's preferences toward ambiguity are CRAA, then the value of an act f is

$$V(f) = \sum_{j=1}^n p_j U(x_j) - \frac{1}{2} \eta \sum_{j=1}^k \left[\frac{\zeta_{1\dots J}^2}{p_{1\dots J}} - \frac{\zeta_{1\dots J-1}^2}{p_{1\dots J-1}} \right] U(x_j) - \frac{1}{2} \eta \sum_{j=k+1}^n \left[\frac{\zeta_{J\dots N}^2}{p_{J\dots N}} - \frac{\zeta_{J+1\dots N}^2}{p_{J+1\dots N}} \right] U(x_j).$$

In both cases, if the DM is ambiguity neutral, no disutility occurs and $V(f) = \sum_{j=1}^n p_j U(x_j)$.

4 Ambiguity measurement

Almost any real-life decision is concerned with ambiguity. For example, what are the chances that a date will be successful? Which job opportunity is more likely to yield the best result for career growth? Or which investment approach is more promising for long term gain? An important part of any decision process is first to rank the different alternatives by their degree of ambiguity. The key point for addressing this need is to determine a well-defined measure of ambiguity. The main goal of this section is to provide such a measure.

4.1 Ordering ambiguous events

A preliminary step in ordering acts by their degree of ambiguity is to define an order over primitive events. This order, induced by the DM's preferences, is defined as follows.

Definition 4.1. *Let the random probabilities of events $\mathcal{E}_j, \mathcal{E}_l \subseteq \mathbb{S}$, having the same expected probability, i.e., $p_j = p_l$, be \hat{f}_j and \hat{f}_l , respectively. Event \mathcal{E}_j is more ambiguous than event \mathcal{E}_l if $\hat{f}_l \succsim^2 \hat{f}_j$ by any ambiguity-averse DM.*

Definition 4.1 provides the subjective ordering that emerges from the DM's preferences. An objective ordering can be defined as follows.³⁰

Definition 4.2. *Event $\mathcal{E}_j \subseteq \mathbb{S}$ is more ambiguous than event $\mathcal{E}_l \subseteq \mathbb{S}$ if there exists a random variable ϵ such that*

$$P_j - p_j =_d P_l - p_l + \epsilon \geq 0,$$

where $=^d$ means equal in distribution and $E[\epsilon \mid P_l] = E[\epsilon] = 0$.³¹

The subjective ordering of Definition 4.1 coincides with the objective ordering of Definition 4.2, as the following proposition proves.

Proposition 4.3. *Given two events $\mathcal{E}_j, \mathcal{E}_l \subseteq \mathbb{S}$ having the same expected probability, Definitions 4.1 and 4.2 of more ambiguous events are equivalent.*

Events can also be ordered by their degree of ambiguity, measured by ζ^2 .

Definition 4.4. *Let $\mathcal{E}_j, \mathcal{E}_l \subseteq \mathbb{S}$. Event \mathcal{E}_j is more ambiguous than event \mathcal{E}_l if $\zeta_j^2 \geq \zeta_l^2$.*

Definition 4.4 states that the higher the fluctuation of probability of an event is the greater its e-ambiguity. Ordering events by ζ^2 coincides with the ordering of Definitions 4.1 and 4.2 if probabilities are equably symmetrically distributed or if the DM's preferences toward ambiguity

³⁰Rothschild and Stiglitz (1970) apply a similar idea for risk, with respect to outcomes.

³¹The condition $E[\epsilon \mid P_l] = E[\epsilon]$ means that ϵ is mean-independent of P_l , i.e., a mean-preserving spread.

are quadratic or of the CAAA type. Formally, the probability of event \mathcal{E}_j is said to be **equably symmetrically** distributed if it satisfies $P_{s+i+1,j} - P_{s+i,j} = \Delta$ and $\chi_{s-i} = \chi_{s+i}$, $\forall i = -s, \dots, s$, where s is the point of symmetry.

Proposition 4.5. *Given two events $\mathcal{E}_j, \mathcal{E}_l \subseteq \mathbb{S}$ having the same expected probability, Definition 4.4 is equivalent to Definitions 4.1 and 4.2 if one of the following conditions holds:*

- (i) *The probabilities of \mathcal{E}_j and \mathcal{E}_l are equably symmetrically distributed;*
- (ii) *The DM's preferences toward ambiguity are of the CAAA type;*
- (iii) *The DM's preferences toward ambiguity are quadratic.*

Henceforth, it is assumed that the probabilities of all events are equably symmetrically distributed. If needed, at any point this assumption can be replaced by assuming CAAA or a quadratic sake function. At this point the order of events by e-ambiguity is well defined. This order induces only a *partial order* on the set of acts \mathcal{F} . E-ambiguity induces a *total order* on subsets of \mathcal{F} satisfying the following condition.

Definition 4.6. *Let the acts $f, g \in \mathcal{F}$, whose probabilities are equably symmetrically distributed, have the same expected probabilities. Act f is more ambiguous than act g if for any event $\mathcal{E} \subseteq \mathbb{S}$, any ambiguity-averse DM $\hat{g}(\mathcal{E}) \succsim^2 \hat{f}(\mathcal{E})$.*

Assuming aversion to ambiguity, Definition 4.6 states that if for any event the random probability associated with act g is preferred to the random probability associated with act f , then f is more ambiguous than g . This definition together with Proposition 4.5 implies that if for any subset of states $\mathcal{E} \subseteq \mathbb{S}$ the e-ambiguity of g is not higher than the e-ambiguity of f , i.e., $\zeta_{f,\mathcal{E}}^2 \geq \zeta_{g,\mathcal{E}}^2$, then f is more ambiguous than g .³² In other words, the idea of Definition 4.6 is that act f is more ambiguous than act g if the probabilities associated with it are consistently more volatile than the probabilities associated with act g .

Practically, Definition 4.6 formulates *first-order stochastic dominance with respect to ambiguity*. This definition can be rephrased as follows: act f is *stochastically dominated* by act g with respect to ambiguity. Figure 2 illustrates two acts, where f is stochastically dominated by g . The values on the y-axis are the degrees of ambiguity of the cumulative events lying on the x-axis. Recall that the degree of ambiguity of the entire state space is always zero. Thus, both acts have a zero degree of ambiguity for the empty event, \emptyset , and the state space, \mathbb{S} .

³²The notations $p_{f,j}$ and $\zeta_{f,j}^2$ stand for the expected probability and the e-ambiguity of event \mathcal{E}_j under act f .

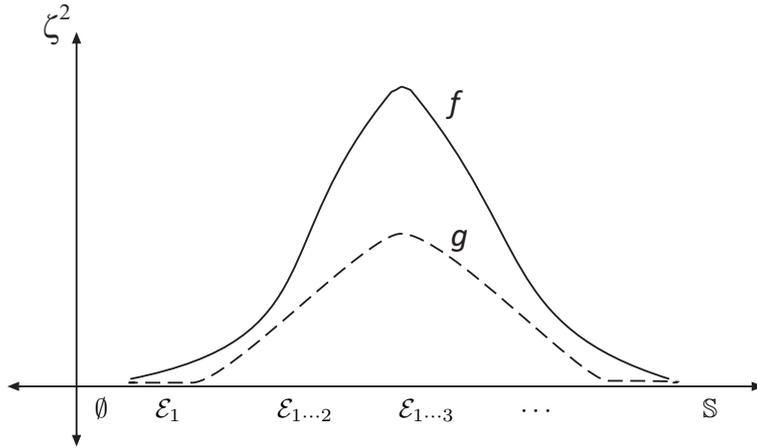


Figure 2: **First order stochastic dominance with respect to ambiguity**

4.2 The ambiguity measure

The ambiguity measure introduced in this section considers symmetric acts. Formally, an act f is said to be a *symmetric act* around the point of symmetry x_s if its probabilities are equably symmetrically distributed and it satisfies $x_s - x_{s-j} = x_{s+j} - x_s$ and $P_{i,s-j} = P_{i,s+j}$, $\forall j = -s, \dots, s$ and $\forall i = 1, \dots, m$. For measuring ambiguity, the more restrictive assumption of normally distributed outcomes, which allows measuring risk by variance, can be relaxed to symmetric distribution. The measure of ambiguity proposed in this section utilizes the cumulative probability of loss and the cumulative probability of gain. It relies on Lemma 3.8, which implies that the variance of the cumulative probability of loss is equal to the variance of the cumulative probability of gain, that is, $\text{Var}[P_L] = \text{Var}[P_G]$. The next model proposes one of the main results of this paper: a new measure of ambiguity. It asserts that the degree of ambiguity embedded in an act can be measured by twice the sum of the variance of its cumulative probability of loss and the variance of its cumulative probability of gain.

Model 4.7. *The **degree of ambiguity** of act $f \in \mathcal{F}$, denoted \mathcal{U}^2 , can be measured by*

$$\mathcal{U}^2[f] = 2\text{Var}[P_L] + 2\text{Var}[P_G] = 4\text{Var}[P_L]. \quad (6)$$

The normalized, to the units of probability, measure is defined by

$$\mathcal{U}[f] = 2\sqrt{\text{Var}[P_L]}. \quad (7)$$

The measure of ambiguity, \mathcal{U}^2 , allows for ordering acts that satisfy first-order stochastic dominance with respect to ambiguity. Theorem 4.8 below proves that, indeed, ordering according to \mathcal{U}^2 coincides with the ordering provided by an ambiguity-averse DM. The ambiguity measure, \mathcal{U}^2 , is inspired by the insight that a probability's fluctuation should be measured

relative to a meaningful reference point. Since decisions are concerned with potential loss, with a non-zero probability (truly mixed), the natural reference point is the consequence which distinguishes losses from gains. While making choices, this consequence is significant to the DM, who is concerned about the perturbation of probabilities of loss and gain.

The minimal possible degree of ambiguity, $\mathcal{U}^2 = 0$, is obtained when all probabilities are perfectly known. The maximal possible degree of ambiguity, $\mathcal{U}^2 = 1$, is obtained when the probability of loss (or gain) is either 0 or 1 with equal odds. In this most extreme case, the variance of the probability of loss attains its maximal possible value, $\frac{1}{4}$. Variances of probabilities are therefore normalized by 4 to provide an ambiguity measure ranging between 0 and 1. Notice that \mathcal{U}^2 depends on a reference point, x_k , which determines the set of gain consequences and the set of loss consequences. If $x_k = x_1$ or $x_k = x_n$, i.e., outcomes are considered either all as gain or all as loss, then the degree of ambiguity equals zero.³³

If there is a reference point agreed upon by all DMs, which makes a clear distinction between losses and gains, the ambiguity measure \mathcal{U}^2 can be considered an *objective* measure of ambiguity; otherwise it is considered a *subjective* measure. Concerning financial assets, for example, the risk-free rate of return can possibly be an objective reference point agreed upon by all financial DMs.

The measure of ambiguity, \mathcal{U}^2 , also takes into account the impact of the correlations between probabilities, across-events. One can define an absolute measure of ambiguity as follow. The *absolute degree of ambiguity* of act f can be measured by³⁴

$$\widehat{\mathcal{U}}^2 [f] = 4 \sum_{j=1}^n \text{Var} [P_j]. \quad (8)$$

Great caution should be exercised when using $\widehat{\mathcal{U}}^2$; by definition probabilities are almost always correlated, such that $\widehat{\mathcal{U}}^2$ is biased in the sense that it ignores these correlations. In other words, this measure disregards an important item of information concerning the nature of probabilities and as a result also the nature of ambiguity.

The point to emphasize is that the measure of ambiguity \mathcal{U}^2 is not influenced by the magnitude of outcomes in general and the magnitude of loss or gain in particular. Increasing or decreasing the outcomes of an act does not change its degree of ambiguity, but it does change its degree of risk. A decision-making process considers not only the degree of ambiguity but

³³These are the only two cases in which acts are not truly mixed.

³⁴One may consider $\sum_{j=1}^k [p_{1\dots J} \zeta_{1\dots J}^2 - p_{1\dots J-1} \zeta_{1\dots J-1}^2] + \sum_{j=k+1}^n [p_{J\dots N} \zeta_{J\dots N}^2 - p_{j+1\dots N} \zeta_{j+1\dots N}^2]$ as relative measure of ambiguity (with respect to expected probabilities).

also the degree of risk. Hence, when making choices these two factors together play a role. A consolidated *uncertainty measure*, which aggregates risk and ambiguity, can be defined by

$$\mathcal{U}(f) = \frac{\text{Var}[f]}{1 - \mathcal{U}^2[f]},$$

see Izhakian (2012).

Taking the first part of the Ellsberg experiment as an example, the probability of drawing a black (B) ball can be one of the possible values $\frac{0}{90}, \frac{1}{90}, \frac{2}{90}, \dots, \frac{60}{90}$. The DM, who doesn't have any additional information regarding which probability is more likely, assigns an equal weight to each alternative. Considering only strictly positive outcomes as a gain (\$9 in this case), the normalized degree of ambiguity (to units of probability) of this bet is $\mathcal{U}[B] = 0.3912$. In the second part of the experiment, when betting on red or yellow (RY), the probability of gain can be one of the possible values $\frac{30}{90}, \frac{31}{90}, \dots, \frac{90}{90}$, which implies that the degree of ambiguity is also $\mathcal{U}[RY] = 0.3912$. Table 2 is a stylized description of this example, where $E[x]$ and $\text{Var}[x]$ are computed using expected probabilities.

	R	Y	B	P_G	$E[P_G]$	$E[x]$	$\text{Var}[x]$	\mathcal{U}
(R)	9	0	0	$\frac{1}{3}$	$\frac{1}{3}$	3	18	0
(B)	0	0	9	$\frac{0}{90}, \frac{1}{90}, \frac{2}{90}, \dots, \frac{60}{90}$	$\frac{1}{3}$	3	18	0.3912
(RY)	9	9	0	$\frac{30}{90}, \frac{31}{90}, \frac{32}{90}, \dots, \frac{90}{90}$	$\frac{2}{3}$	6	18	0.3912
(BY)	0	9	9	$\frac{2}{3}$	$\frac{2}{3}$	6	18	0

Table 2: **The Ellsberg example**

Assume now that instead of 60 black and yellow balls in an unknown proportion, the urn contains only 30 black and yellow balls in an unknown proportion in addition to 30 red balls. The degree of ambiguity of the bet on B decreases to $\mathcal{U}[B] = 0.2981$. If the amount of balls which may be black or yellow in an unknown proportion is 90, then the degree of ambiguity of the bet on B increases to $\mathcal{U}[B] = 0.4377$. If the urn contains only 60 balls, all of them black or yellow in an unknown proportion, then the degree of ambiguity, which in this case is identical for black and yellow, is $\mathcal{U} = 0.5868$. If there are only 10 balls in the urn (and again the proportion of black and yellow is unknown), then $\mathcal{U} = 0.6324$. Finally, if there is only one ball in the urn, of unknown color, then $\mathcal{U} = 1$, and in the other extreme case if there is an infinite number of balls in the urn, then $\mathcal{U} = \frac{1}{\sqrt{12}}$. Table 3 is a stylized description of these variations.

To prove that \mathcal{U}^2 measures ambiguity, it has to be shown that indeed, given two acts with an identical expected outcome and an identical degree of risk, an ambiguity-averse DM prefers the

#balls	R	Y	B	P_G	$\mathcal{U}[B]$
90	30	$0, \dots, 60,$	$0, \dots, 60$	$\frac{0}{90}, \frac{1}{90}, \dots, \frac{60}{90}$	0.3912
60	30	$0, \dots, 30$	$0, \dots, 30$	$\frac{0}{60}, \frac{1}{60}, \dots, \frac{30}{60}$	0.2981
120	30	$0, \dots, 90$	$0, \dots, 90$	$\frac{0}{120}, \frac{1}{120}, \dots, \frac{90}{120}$	0.4377
∞	0	$0 \dots \infty$	$0 \dots \infty$	$0 \dots 1$	0.2886
60	0	$0, \dots, 60$	$0, \dots, 60$	$\frac{0}{60}, \frac{1}{60}, \dots, \frac{60}{60}$	0.5868
10	0	$0, \dots, 10$	$0, \dots, 10$	$\frac{0}{10}, \frac{1}{10}, \dots, \frac{10}{10}$	0.6324
1	0	$0, \dots, 1$	$0, 1$	$0, 1$	1

Table 3: **Degrees of ambiguity**

act with the lower \mathcal{U}^2 . The next theorem ties \mathcal{U}^2 to the DM's preferences concerning ambiguity. To differentiate the impact of ambiguity, the ordering is made over acts with identical properties except for the degree of ambiguity. That is, the acts have the same set of outcomes and the same expected probability for each outcome, such that the only difference between them is the dispersion of probabilities around their expected probabilities, for each event. To eliminate the impact of preferences concerning risk, the DM is assumed to be risk neutral. Since acts are assumed to be symmetric, with a point of symmetry x_s equal to the expected outcome, it is assumed that the reference point satisfies $x_k \leq x_s$; otherwise, the DM will not consider these acts.

Theorem 4.8. *Assume symmetric acts $f, g \in \mathcal{F}$, satisfying first-order stochastic dominance with respect to ambiguity, sharing the same set of outcomes, X , and having the same expected probability for each outcome, i.e., $E[P(f_j)] = E[P(g_j)], \forall j = 1, \dots, n$.³⁵ Act f is more ambiguous than act g , i.e., $\mathcal{U}^2[g] \leq \mathcal{U}^2[f]$, iff any ambiguity-averse DM, with a reference point $x_k \leq x_s$, prefers g to f .*

Theorem 4.8 proves that if two acts are identical except in their degree of ambiguity any ambiguity-averse DM prefers the act with the lower \mathcal{U}^2 over the act with the higher \mathcal{U}^2 , which implies that ordering acts by \mathcal{U}^2 is identical to the ordering provided by a DM who exhibits aversion to ambiguity.

³⁵Here, f and g are referred to as random variables; thus, the index j designates outcome rather than event.

5 Implementation for asset pricing

The main advantage of shadow theory, implemented by Model 3.4, is that it achieves a complete separation of ambiguity from risk and beliefs from preferences. To demonstrate this merit, this section presents an application of the theory to asset pricing. The prices that investors are willing to pay for assets might be affected by the fact that they do not know the precise probabilities of returns. They might require an additional premium, in terms of return, for bearing ambiguity in addition to the premium they require for bearing risk. This section models these two premiums, the *risk premium* and the *ambiguity premium*, in separated closed forms. The empirical tests of this model conducted by Brenner and Izhakian (2011) are then described.

5.1 The ambiguity premium

The *risk premium* is the premium that a DM is willing to pay for replacing a risky bet with its expected outcome. The *ambiguity premium* is the premium that a DM is willing to pay for replacing an ambiguous bet with a risky but non-ambiguous bet having an identical expected outcome. The *uncertainty premium* is the total premium that a DM is willing to pay for replacing an ambiguous bet with its expected outcome, i.e., the accumulation of risk premium and ambiguity premium. Formally, the uncertainty premium, denoted \mathcal{K} , is defined by

$$U(E[x] - \mathcal{K}) = \sum_{j=1}^k [p_j - (\varphi_{1\dots J} - \varphi_{1\dots J-1})] U(x_j) + \sum_{j=k+1}^n [p_j - (\varphi_{j\dots N} - \varphi_{j+1\dots N})] U(x_j). \quad (9)$$

The *certainty equivalent*, $CE = E[x] - \mathcal{K}$, of an act $f \in \mathcal{F}$ is a constant act $CE \in \mathcal{F}$ satisfying $V(CE) = V(f)$. In other words, the certainty equivalent is the constant act for which the DM is willing to exchange a risky ambiguous (uncertain) act. The next theorem approximates this premium and separates it into risk premium and ambiguity premium.

Theorem 5.1. *Assume a DM whose reference point x_k is relatively close to zero, satisfying $0 \leq x_k \leq E[x]$, and her preferences are characterized by a twice-differentiable utility function, $U(\cdot)$, and a twice-differentiable sake function, $\psi(\cdot)$. The uncertainty premium is then*

$$\mathcal{K} \approx \underbrace{-\frac{1}{2} \frac{U''(E[x])}{U'(E[x])} \text{Var}[x]}_{\mathcal{R}} - \underbrace{\frac{1}{8} \left[\frac{\psi''(E[P_L])}{\psi'(E[P_L])} + \frac{\psi''(E[P_G])}{\psi'(E[P_G])} \right]}_{\mathcal{A}} E[x] \mathcal{U}^2[x], \quad (10)$$

where \mathcal{R} is the risk premium and \mathcal{A} is the ambiguity premium.³⁶

³⁶The proof of Theorem 5.1 applies the same method as that of Arrow (1965) and Pratt (1964), while outcomes

Concerning financial decisions, outcomes can possibly be rates of return, r . In this case, the uncertainty premium, in terms of return, takes the form

$$\mathcal{K} \approx -\frac{1}{2} \frac{U''(E[r])}{U'(E[r])} \text{Var}[r] - \frac{1}{8} \left[\frac{\psi''(E[P_L])}{\psi'(E[P_L])} + \frac{\psi''(E[P_G])}{\psi'(E[P_G])} \right] \mathcal{U}^2[r]. \quad (11)$$

Theorem 5.1 provides a complete distinction in two aspects. First, it distinguishes between risk premium and ambiguity premium, such that these two premiums are orthogonal. Second, within each premium, it distinguishes between sources of premiums, i.e., preference and beliefs.

The risk premium, $\mathcal{R} \approx -\frac{1}{2} \frac{U''(E[r])}{U'(E[r])} \text{Var}[r]$, is the Arrow-Pratt risk premium. Independently, a higher risk, measured by $\text{Var}[r]$, or higher risk aversion, measured by the coefficient of absolute risk aversion, $-\frac{U''(\cdot)}{U'(\cdot)}$, both result in a higher risk premium. The degree of risk, $\text{Var}[r]$, is a matter of the DM's beliefs, while risk aversion is a matter of her preferences concerning it.

The ambiguity premium, $\mathcal{A} \approx -\frac{1}{8} \left[\frac{\psi''(E[P_L])}{\psi'(E[P_L])} + \frac{\psi''(E[P_G])}{\psi'(E[P_G])} \right] \mathcal{U}^2[r]$, possesses attributes resembling those of the risk premium, but with respect probabilities rather than to consequences. A complete separation between beliefs about random probabilities, measured by \mathcal{U}^2 , and preferences concerning it, measured by the coefficient of absolute ambiguity aversion, $-\frac{\psi''(\cdot)}{\psi'(\cdot)}$, is achieved. Ambiguity aversion implies a positive ambiguity premium, since $-\frac{\psi''(\cdot)}{\psi'(\cdot)} > 0$. Ambiguity loving implies a negative ambiguity premium, since $-\frac{\psi''(\cdot)}{\psi'(\cdot)} < 0$. Ambiguity neutrality implies a zero ambiguity, since $-\frac{\psi''(\cdot)}{\psi'(\cdot)} = 0$. A zero ambiguity premium is also obtained when probabilities are perfectly known, i.e., $\mathcal{U}^2 = 0$. Independently, a higher degree of ambiguity or a higher aversion to it, both increase the ambiguity premium.

As an example, the next corollary shows the different premiums in the case of a DM typified by constant relative risk aversion (CRRA) and constant absolute ambiguity aversion (CAAA).

Corollary 5.2. *Assuming a DM who is characterized by CRRA,³⁷*

$$U(x_j) = \begin{cases} \frac{x_j^{1-\gamma} - x_k^{1-\gamma}}{1-\gamma}, & \gamma \neq 0 \\ \ln(x_j) - \ln(x_k), & \gamma = 0 \end{cases},$$

and CAAA,

$$\psi(P_i) = -\frac{e^{-\eta P_i}}{\eta},$$

then the uncertainty premium is

$$\mathcal{K} \approx \underbrace{\gamma \frac{1}{2} \text{Var}[r]}_{\mathcal{R}} + \underbrace{\eta \frac{1}{4} E[r] \mathcal{U}^2[r]}_{\mathcal{A}}.$$

are also restricted to be relatively small.

³⁷A more standard formulation of CRRA, $U(x) = \frac{x^{1-\gamma}}{1-\gamma}$ for $\gamma \neq 1$ and $U(x) = \ln(x)$ for $\gamma = 1$ otherwise, is not always normalized to $U(x_k) = 0$.

When the expected probabilities of loss and gain are relatively close to $\frac{1}{2}$, the ambiguity premium can be simplified to

$$\mathcal{A} \approx -\frac{1}{4} \frac{\psi''(\mathbb{E}[P_L])}{\psi'(\mathbb{E}[P_L])} \mathcal{U}^2[r],$$

and the *expected return* (equity premium) is thus

$$\mathbb{E}[r] \approx r_f - \frac{1}{2} \frac{U''(\mathbb{E}[r])}{U'(\mathbb{E}[r])} \text{Var}[r] - \frac{1}{4} \frac{\psi''(\mathbb{E}[P_L])}{\psi'(\mathbb{E}[P_L])} \mathcal{U}^2[r], \quad (12)$$

where r_f stands for the risk-free rate of return. Brenner and Izhakian (2011) test this equation empirically as described next.

5.2 Empirical results

To demonstrate the empirical aspects of shadow theory and its explanatory power, this section reviews the results of Brenner and Izhakian (2011). Their study employs the ambiguity measure, \mathcal{U}^2 , as an explanatory factor of the aggregate return on the stock market. They assume a pricing representative investor, whose reference point is the risk-free rate. The return on the stock market is assumed to be normally distributed, but the parameters, mean and variance, governing its distribution are assumed to be random.

Using the S&P 500 intraday data, Brenner and Izhakian (2011) extract the monthly degree of ambiguity by utilizing the following four-step methodology. The first step is sampling 20 to 22 groups, each comprising 27 observations (15-minutes returns) from the monthly data. A group can be selected randomly or can simply be the observation of a specific day in the month. The second step is computing the mean and variance of each group. Assuming that returns are normally distributed, the third step is computing the probability of a negative return (loss) for each group, using its mean and variance. At this point, for each month there are 20-22 probabilities of loss. The last step is computing the variance of these probabilities to obtain the monthly degree of ambiguity \mathcal{U}^2 .

Employing the degree of ambiguity, Brenner and Izhakian (2011) conduct a series of tests to study its explanatory power on the stock market's return and its relationships with other risk factors. They show that ambiguity has a significantly negative impact on returns in both contemporaneous and prediction testing, which means that monthly return on the stock market is not only negatively affected by the ambiguity in the same month but also by the ambiguity in the previous month. These results indicate that typically investors in the stock market are ambiguity loving.

6 Related Literature

Since the seminal works of Knight (1921) and Ellsberg (1961), utility theory research has been making a concerted effort to treat decision processes under uncertainty and explain the violation of expected utility theory. This effort has generated the ideas that, in the presence of ambiguity, the DM's belief takes the form either of multiple priors or of a single but non-additive prior. In their *max-min expected utility with multiple priors* (MEU) model, Gilboa and Schmeidler (1989) assert that an ambiguity-averse DM possesses a set of priors and evaluates her ex-ante welfare conditional upon the worst prior. The *subjective nonadditive* probabilities of Gilboa (1987) and the *Choquet expected utility* (CEU) of Schmeidler (1989) state that uncertainty and aversion to it can be represented by a single subadditive prior (capacity). Tversky and Kahneman's (1992) CPT implements a two-sided variant of CEU with different capacities for gains and for losses.

Shadow theory's contribution to this literature is threefold. First, it achieves a complete separation between ambiguity and attitude toward it. Second, it provides a bridge between the MEU discipline and the CEU and CPT disciplines. Third, it proves that capacities are not arbitrary and can be explained by the presence of ambiguity and the DM's preferences concerning it. Capacities, in shadow theory, emerge from the randomness of probability distribution and the nonlinear ways in which individuals may perceive these probabilities. This randomness can be viewed as a random selection, dominated by a second-order prior, of one prior from a set of priors.

The concept of modeling attitudes toward ambiguity by relaxing the reduction between first-order and second-order probabilities, suggested by Segal (1987), inspires other models: Klibanoff *et al.*'s (2005) smooth model of ambiguity, its recursive form, also proposed by Klibanoff *et al.* (2009), its generalization to include intertemporal substitution, proposed by Ju and Miao (2011) and Hayashi and Miao (2011), and the second-order probability sophistication of Nau (2006), Chew and Sagi (2008) and Ergin and Gul (2009). Unlike our model, in these model preferences toward ambiguity are taken with respect to expected utilities or certainty equivalents, so that a complete distinction between the impact of risk preferences and the impact of ambiguity preferences on decisions is not trivial. Applying preferences concerning ambiguity to probabilities, our model achieves a complete distinction between the effects of risk preferences and the effects of ambiguity preferences.

Shadow theory can be interpreted as a model of robustness in the presence of *model uncer-*

tainty. This class of models assumes an uncertainty about the true probability law governing the realization of states, and a DM, with her concerns about misclassification, looks for robust decision making; see, for example, Hansen and Sargent (2001), Maccheroni *et al.* (2006) and Hansen and Sargent (2007). Generally speaking, ambiguity in this line of models is formulated by the deviation of probability from a reference probability (reference model), measured by *relative entropy*.³⁸ Shadow theory relaxes the requirement of having a reference model and required only a reference point, which is easier to identify.

Shadow theory is also related to Siniscalchi's (2009) *vector expected utility*, which assumes a baseline probability and different sources of ambiguity with respect to expected utility. Other models that consider reference expected utility include those of Roberts (1980), Quiggin *et al.* (2004) and Grant and Polak (2007), for example, or consider a reference prior (Einhorn and Hogarth (1986), and Gajdos *et al.* (2008), for example). Kopylov's (2006) ϵ -contamination suggests the addition of an element of confidence to the generated set of priors. Chateauneuf *et al.* (2007) suggest new capacities (neo-additives) obtained from an α -max-min expected utility with a set of priors generated by ϵ -contamination. All these models require the identification of a reference prior, which practically is not a simple task, if it is possible at all.

In some sense shadow theory can be viewed as a special case of the *phantom probability* framework suggested by Izhakian and Izhakian (2009a) and its implementation to decision theory by Izhakian and Izhakian (2009b), where the authors suggest that observable real outcomes and real probabilities are projections of consequences in a generalized multidimensional space called *phantom space*. The norm over this space, which maps phantom probabilities to real probabilities, can be considered as derived by preferences concerning ambiguity, which is equivalent to the sake function in shadow theory.

This is the point to emphasize. Shadow theory differs from all of the models mentioned above in one major aspect: ambiguity is applied directly to probabilities and not to any element of utility. That is, it is not applied to expected utility, certainty equivalent or event-wise utility, which are driven by preference concerning risk. In shadow theory no need arises to identify a reference probability distribution. It provides a formal way to compare the choices of two DMs who have different attitudes toward ambiguity, or different degrees of ambiguity; for example, two DMs who share the same information and the same attitude toward risk but have different levels of ambiguity sensitivity, or two DMs who share the same attitude toward

³⁸Relative entropy is the expected log Radon-Nikodym derivative. Technically, all alternative models have to be absolutely continuous with respect to the reference model for an entropy measure to exist.

risk and ambiguity but possess different information (different degrees of ambiguity). The ability to conduct this type of comparative static is of primary importance, as it allows for the identification of the pure effect of introducing ambiguity and attitude toward it into a model.

The behavioral decision literature documents different preferences concerning ambiguity when facing gains, as compared to facing losses: the preferences are usually ambiguity aversion for gains and ambiguity loving for losses (see, for example, Einhorn and Hogarth (1986), Ho *et al.* (2009) and Chakravarty and Roy (2009)). Our model supports different preferences for ambiguity for losses and gains. Interestingly, *overconfidence* is also linked to ambiguity. Brenner *et al.* (2011) show that when exposed to ambiguity individuals are less overconfident about the likelihood of outperforming a benchmark portfolio. In terms of shadow theory, these results imply that individuals assign lower subjective probabilities of gain when exposed to a higher degree of ambiguity. Practically, their results provide experimental evidence supporting our model.

Several approaches to estimating ambiguity have been suggested in the literature. Dow and Werlang (1992) measure uncertainty as the sum of the probability of an event and the probability of its complement event. Assuming a normal distribution with an unknown mean, Ui (2011) measures ambiguity by the difference between the minimal possible mean and the true mean. Assuming a second-order belief, Maccheroni *et al.* (2011) measure ambiguity by the variance of an unknown mean. Assuming a multiple-prior setting, Bewley (2011) measures ambiguity by a critical confidence interval. Assuming mean-variance preferences, with known variance and unknown mean, Boyle *et al.* (2011) also measure ambiguity using the confidence interval $\alpha = \left\{ \mu : \frac{(\mu - \hat{\mu})^2}{\sigma_{\hat{\mu}}^2} \leq \alpha^2 \right\}$, where $\hat{\mu}$ is the estimated value of the excess mean return, μ , and $\sigma_{\hat{\mu}}$ is its standard error. All these studies assume that the variance of consequences is known. Our measure of ambiguity, \mathcal{U}^2 , is broader; it assumes an unknown variance and allows all other parameters that characterize probabilities to be unknown.

In their empirical analysis, Anderson *et al.* (2009) and Drechsler (2011) proxy for uncertainty via the degree of disagreement of professional forecasters, attributing different weights to each forecaster. Jewitt and Mukerji (2011) investigate the ranking of ambiguous acts as revealed by the DM's preferences. These methods can be classified as subjective ordering by ambiguity, while \mathcal{U}^2 can be considered as providing an objective ordering.

The implications of ambiguity regarding the equity premium have been studied mainly by focusing on the theoretical aspects. Izhakian and Benninga (2011) add an ambiguity premium

to the conventional risk premium and show that increasing risk aversion might result in a lower uncertainty premium. Maccheroni *et al.* (2011) adjust the mean-variance model for ambiguity and extract the ambiguity premium. Ui (2011) proves that, due to changes in market participation, changes in risk premium and changes in ambiguity premium may have opposite signs. Epstein and Schneider (2008) employ the max-min model to show that the ambiguity premium depends on the idiosyncratic risk in fundamentals.³⁹ Unlike Theorem 5.1, in these papers a complete separation between preferences and beliefs and between risk and ambiguity is not obtained.

7 Conclusion

In reality almost any decision involves ambiguity. It is only too natural to look for a simple measure of ambiguity that allows for ordering uncertain alternatives by their degree of ambiguity. The current paper satisfies this need by providing an *ambiguity measure* which can easily be employed in empirical studies. Brenner and Izhakian (2011), for example, use this measure to inquire into the question of whether stock prices are affected by ambiguity. Their empirical study shows that ambiguity has a significant negative impact on stock returns. To the best of our knowledge, this study is the first to measure ambiguity from market data, rather than in laboratory experiments or calibrations.

To construct a useful measure of ambiguity, the paper introduces a novel model of decision making under ambiguity, called *shadow probability theory*, generalizing Schmeidler's (1989) Choquet expected utility and Tversky and Kahneman's (1992) cumulative prospect theory. Shadow theory assumes that probabilities of observable events are random and dominated by second-order unobservable events, modeled by two separated state spaces. The structure of two separated spaces allows for a complete distinction between risk and ambiguity and between preferences and beliefs. The degree of ambiguity and the decision maker's attitude toward it are then measured with respect to one space, while risk and risk attitude apply to the second space. In this model, subjective probabilities are framed by the nonlinear ways in which individuals may process probabilities. Perceived probabilities are nonadditive: ambiguity aversion results in a subadditive subjective probability measure, while ambiguity loving results in a superadditive

³⁹Chen and Epstein (2002) extend the MEU model to continuous-time recursive multiple-priors utility and demonstrate a separation between ambiguity premium and risk premium. Segal and Spivak (1990) also analyze the ambiguity premium, which they call a *premium of order 2*.

measure.

Ambiguity in shadow theory takes the form of probability perturbation with respect to a reference point that distinguishes losses from gains. This concept provides a natural ambiguity measure, \mathcal{U}^2 , which proved to be empirically testable. The measure of ambiguity, \mathcal{U}^2 , is simply four times the variance of the probability of loss (or gain). The present paper demonstrates the merits of this new model by incorporating ambiguous probabilities into the basic asset pricing model. It generalizes the Arrow-Pratt theory and clearly differentiates between the *risk premium* and the *ambiguity premium*, which can both be measured empirically. The measure of ambiguity, introduced in this paper, is an important tool that paves the way for studying the empirical and behavioral implications of ambiguity for finance and economics.

References

- Abdellaoui, M., A. Attema, and H. Bleichrodt (2010) “Intertemporal Tradeoffs for Gains and Losses: An Experimental Measurement of Discounted Utility,” *Economic Journal*, Vol. 120, No. 545, pp. 845–866.
- Abramowitz, M. and I. A. Stegun (1972) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover, 9th edition.
- Anderson, E. W., E. Ghysels, and J. L. Juergens (2009) “The Impact of Risk and Uncertainty on Expected Returns,” *Journal of Financial Economics*, Vol. 94, No. 2, pp. 233–263.
- Arrow, K. J. (1965) *Aspects of the Theory of Risk Bearing*, Helsinki: Yrjo Jahnssonin Saatio.
- Barberis, N. and M. Huang (2001) “Mental Accounting, Loss Aversion, and Individual Stock Returns,” *Journal of Finance*, Vol. 56, No. 4, pp. 1247–1292.
- Bewley, T. F. (2011) “Knightian Decision Theory and Econometric Inferences,” *Journal of Economic Theory*, Vol. 146, No. 3, pp. 1134–1147.
- Bier, V. M. and B. L. Connell (1994) “Ambiguity Seeking in Multi-Attribute Decisions: Effects of Optimism and Message Framing,” *Journal of Behavioral Decision Making*, Vol. 7, pp. 169–182.
- Boyle, P. P., L. Garlappi, R. Uppal, and T. Wang (2011) “Keynes Meets Markowitz: The Tradeoff Between Familiarity and Diversification,” *Management Science*, pp. 1–20.
- Brenner, M. and Y. Izhakian (2011) “Asset Pricing and Ambiguity; Empirical Evidence,” Stern School of Business, Finance Working Paper Series, FIN-11-10.
- Brenner, M., Y. Izhakian, and O. Sade (2011) “Ambiguity and Overconfidence,” Preprint at SSRN:1773663.
- Chakravarty, S. and J. Roy (2009) “Recursive Expected Utility and The Separation of Attitudes Towards Risk and Ambiguity: an Experimental Study,” *Theory and Decision*, Vol. 66, No. 3, pp. 199–228.
- Chateauneuf, A., J. Eichberger, and S. Grant (2007) “Choice Under Uncertainty with the Best and Worst in Mind: Neo-Additive Capacities,” *Journal of Economic Theory*, Vol. 137, No. 1, pp. 538 – 567.
- Chateauneuf, A. and J. Y. Jaffray (1989) “Some characterizations of lower probabilities and other monotone capacities through the use of Mobius inversion,” *Mathematical Social Sciences*, Vol. 17, No. 3, pp. 263–283.
- Chen, Z. and L. Epstein (2002) “Ambiguity, Risk, and Asset Returns in Continuous Time,” *Econometrica*, Vol. 70, No. 4, pp. 1403–1443.
- Chew, S. H. and J. S. Sagi (2008) “Small Worlds: Modeling Attitudes Toward Sources of Uncertainty,” *Journal of Economic Theory*, Vol. 139, No. 1, pp. 1–24.
- Coval, J. D. and T. J. Moskowitz (1999) “Home Bias at Home: Local Equity Preference in Domestic Portfolios,” *Journal of Finance*, Vol. 54, No. 6, pp. 2045–2073.
- Dobbs, I. M. (1991) “A Bayesian Approach to Decision-Making under Ambiguity,” *Economica*, Vol. 58, No. 232, p. 417.
- Dow, J. and S. R. d. C. Werlang (1992) “Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio,” *Econometrica*, Vol. 60, No. 1, pp. 197–204.
- Drechsler, I. (2011) “Uncertainty, Time-Varying Fear, and Asset Prices,” preprint at SSRN:1364902.

- Einhorn, H. J. and R. M. Hogarth (1986) “Decision Making under Ambiguity,” *The Journal of Business*, Vol. 59, No. 4, pp. 225–250.
- Ellsberg, D. (1961) “Risk, Ambiguity, and the Savage Axioms,” *Quarterly Journal of Economics*, Vol. 75, No. 4, pp. 643 – 669.
- Epstein, L. and M. Schneider (2008) “Ambiguity, Information Quality, and Asset Pricing,” *Journal of Finance*, Vol. 63, No. 1, pp. 197–228.
- Erbas, S. and A. Mirakhor (2007) “The Equity Premium Puzzle, Ambiguity Aversion, and Institutional Quality.”
- Ergin, H. and F. Gul (2009) “A Theory of Subjective Compound Lotteries,” *Journal of Economic Theory*, Vol. 144, No. 3, pp. 899–929.
- Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008) “Attitude Toward Imprecise Information,” *Journal of Economic Theory*.
- Gilboa, I. (1987) “Expected Utility with Purely Subjective Non-Additive Probabilities,” *Journal of Mathematical Economics*, Vol. 16, No. 1, pp. 65–88.
- Gilboa, I. and D. Schmeidler (1989) “Maxmin Expected Utility with Non-Unique Prior,” *Journal of Mathematical Economics*, Vol. 18, No. 2, pp. 141–153.
- Goetzmann, W. N. and A. Kumar (2008) “Equity Portfolio Diversification,” *Review of Finance*, Vol. 12, No. 3, pp. 433–463.
- Grabisch, M., J. luc Marichal, and M. Roubens (2000) “Equivalent Representations of Set Functions,” *Mathematics of Operations Research*, Vol. 25, pp. 157–178.
- Grant, S. and B. Polak (2007) “Generalized Variational Preferences,” Paper presented at the Risk Decision and Uncertainty (RUD), Tel-Aviv.
- Hansen, L. P. and T. J. Sargent (2001) “Robust Control and Model Uncertainty,” *American Economic Review*, Vol. 91, No. 2, pp. 60–66.
- (2007) *Robustness*, Princeton: Princeton University Press.
- Hayashi, T. and J. Miao (2011) “Intertemporal Substitution and Recursive Smooth Ambiguity Preferences,” *Theoretical Economics*.
- Ho, J. L. Y., L. R. Keller, and P. Kelytyka (2002) “Effects of Outcome and Probabilistic Ambiguity on Managerial Choices,” *Journal of Risk and Uncertainty*, Vol. 24, No. 1, pp. 47–74.
- Hong, C. S. and P. Wakker (1996) “The Comonotonic Sure-Thing Principle,” *Journal of Risk and Uncertainty*, Vol. 12, No. 1, pp. 5–27.
- Izhakian, Y. (2012) “Capital Asset Pricing under Ambiguity,” Stern School of Business, Economics Working Paper Series, ECN-12-01.
- Izhakian, Y. and S. Benninga (2011) “The Uncertainty Premium in an Ambiguous Economy,” *Quatrly Journal of Finance*, Vol. 1, pp. 323–354.
- Izhakian, Y. and Z. Izhakian (2009a) “Phantom Probability,” Preprint at arXiv:0901.0902.
- (2009b) “Decision Making in Phantom Spaces,” Preprint at SSRN:1329366.
- Jaffray, J. Y. (1989) “Linear Utility Theory for Belief Functions,” *Operations Research Letters*, No. 8, p. 107112.

- Jewitt, I. and S. Mukerji (2011) "Ordering Ambiguous Acts," University of Oxford, Department of Economics, Economics Series Working Papers.
- Ju, N. and J. Miao (2011) "Ambiguity, Learning, and Asset Returns," *Econometrica*.
- Kahneman, D. and A. Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, Vol. 47, No. 2, pp. 263–91.
- Klibanoff, P., M. Marinacci, and S. Mukerji (2005) "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, Vol. 73, No. 6, pp. 1849–1892.
- (2009) "Recursive Smooth Ambiguity Preferences," *Journal of Economic Theory*, Vol. 144, No. 3, pp. 930–976.
- Knight, F. M. (1921) *Risk, Uncertainty and Profit*, Boston: Houghton Mifflin.
- Kopylov, I. (2006) "A Parametric Model of Hedging Under Ambiguity," Mimeo, UC Irvine. [829].
- Kothiyal, A., V. Spinu, and P. P. Wakker (2011) "Prospect Theory for Continuous Distributions: a Preference Foundation," *Journal of Risk and Uncertainty*, No. 42, p. 195210.
- Levy, H., E. D. Giorgi, and T. Hens (2003) "Prospect Theory and the CAPM: A Contradiction or Coexistence?."
- Maccheroni, F., M. Marinacci, and D. Ruffino (2010) "Alpha as Ambiguity: Robust Mean-Variance Portfolio Analysis."
- Maccheroni, F., M. Marinacci, and A. Rustichini (2006) "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, Vol. 74, No. 6, pp. 1447–1498.
- Mehra, R. and E. C. Prescott (1985) "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, Vol. 15, No. 2, pp. 145–161.
- Nau, R. F. (2006) "Uncertainty Aversion with Second-Order Utilities and Probabilities," *Management Science*, Vol. 52, pp. 136–145.
- Pratt, J. W. (1964) "Risk Aversion in the Small and in the Large," *Econometrica*, Vol. 32, No. 1/2, pp. 122–136.
- Quiggin, J. (1982) "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization*, Vol. 3, No. 4, pp. 323–343.
- Quiggin, J. and R. G. Chambers (2004) "Invariant Risk Attitudes," *Journal of Economic Theory*, Vol. 117, No. 1, pp. 96–118.
- Roberts, K. W. S. (1980) "Interpersonal Comparability and Social Choice Theory," *Review of Economic Studies*, Vol. 47, No. 2, pp. 421–39.
- Rothschild, M. and J. E. Stiglitz (1970) "Increasing Risk: I. A Definition," *Journal of Economic Theory*, Vol. 2, No. 3, pp. 225 – 243.
- Rottenstreich, Y. and A. Tversky (1997) "Unpacking, Repacking, and Anchoring: Advances in Support Theory," *Psychological Review*.
- Schmeidler, D. (1989) "Subjective Probability and Expected Utility without Additivity," *Econometrica*, Vol. 57, No. 3, pp. 571–87.
- Segal, U. (1987) "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach," *International Economic Review*, Vol. 28, No. 1, pp. 175–202.

- Segal, U. and A. Spivak (1990) "First Order Versus Second Order Risk Aversion," *Journal of Economic Theory*, Vol. 51, No. 1, pp. 111–125.
- Shiller, R. J. (1981) "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, Vol. 71, No. 3, pp. 421–36.
- Siniscalchi, M. (2009) "Vector Expected Utility and Attitudes Toward Variation," *Econometrica*, Vol. 77, No. 3, pp. 801–855.
- Tversky, A. and D. Kahneman (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, Vol. 5, No. 4, pp. 297–323.
- Tversky, A. and D. J. Koehler (1994) "Support Theory: A Nonextensional Representation of Subjective Probability," *Psychological Review*, Vol. 101, pp. 547–567.
- Ui, T. (2011) "The Ambiguity Premium vs. the Risk Premium under Limited Market Participation," *Review of Finance*, Vol. 15, No. 2, pp. 245–275.
- Uppal, R. and T. Wang (2003) "Model Misspecification and Under Diversification," *Journal of Finance*, Vol. 58.
- Wakker, P. and A. Tversky (1993) "An Axiomatization of Cumulative Prospect Theory," *Journal of Risk and Uncertainty*, Vol. 7, No. 2, pp. 147–75.
- Wakker, P. (2010) *Prospect Theory: For Risk and Ambiguity*: Cambridge University Press.
- Wang, M. and P. S. Fischbeck (2004) "Incorporating Framing into Prospect Theory Modeling: A Mixture-Model Approach," *Journal of Risk and Uncertainty*, Vol. 29, No. 2, pp. 181–197.
- Weil, P. (1989) "The Equity Premium Puzzle and The Risk-Free Rate Puzzle," *Journal of Monetary Economics*, Vol. 24, No. 3, pp. 401–421.
- Weymark, J. A. (2001) "Generalized Gini Indices of Equality of Opportunity," Working Papers 0114, Department of Economics, Vanderbilt University.
- Yaari, M. E. (1987) "The Dual Theory of Choice under Risk," *Econometrica*, Vol. 55, No. 1, pp. 95–115.

A Appendix

Lemma A.1. Assume a sake function $\psi(\cdot)$ satisfying

$$\frac{1}{2} \left(\frac{\psi''(p_A)}{\psi'(p_A)} \zeta_A^2 - \frac{\psi''(p_{A \cup B})}{\psi'(p_{A \cup B})} \zeta_{A \cup B}^2 \right) \leq p_B,$$

for any events $A, B \subseteq \mathbb{S}$. If $A \subset B$ then $Q(A) \leq Q(B)$.

Proof of Lemma A.1. Writing $C = A \cup B$, then by Theorem 3.5

$$\begin{aligned} Q(C) - Q(A) &= p_A + p_B + \frac{1}{2} \frac{\psi''(p_C)}{\psi'(p_C)} \zeta_C^2 - p_A - \frac{1}{2} \frac{\psi''(p_A)}{\psi'(p_A)} \zeta_A^2 \\ &= p_B + \frac{1}{2} \frac{\psi''(p_{A \cup B})}{\psi'(p_{A \cup B})} \zeta_{A \cup B}^2 - \frac{1}{2} \frac{\psi''(p_A)}{\psi'(p_A)} \zeta_A^2, \end{aligned}$$

which, by the assumption of the Lemma, is non-negative. \square

Lemma A.2. Assume two equably symmetric directing acts $y, z \in \widehat{\mathcal{F}}$ with identical means. Let μ_y^k and μ_z^k be the k th moment around 0 of y and z respectively, then $\mu_y^k (\mu_y^2)^{\frac{k}{2}} = \mu_z^k (\mu_z^2)^{\frac{k}{2}}$ for any even k .

Proof of Lemma A.2. Without loss of generality, assume $E[y] = E[z] = 0$.⁴⁰ Let $Y = y_{i+1} - y_i$ and $Z = z_{i+1} - z_i$, and recall that, since y and z share the same probability space, $\chi(y_i) = \chi(z_i)$, $\forall i = 1, \dots, m$. The $2k$ th moments of y and z can then be written as $\sum \chi_i (iY)^{2k}$ and $\sum \chi_i (iZ)^{2k}$, respectively, where χ_i is the probability of the ε_i . Now, $(\mu_y^2)^{\frac{2k}{2}}$ can be written as $(\sum \chi_i (iY)^2)^k$ and $(\mu_z^2)^{\frac{2k}{2}}$ can be written as $(\sum \chi_i (iZ)^2)^k$. Finally

$$\begin{aligned} \mu_y^k (\mu_y^2)^{\frac{k}{2}} &= \sum \chi_i (iY)^{2k} (\sum \chi_i (iZ)^2)^k = (YZ)^{2k} 2^{k+1} \sum \chi_i i^{2k} (\sum \chi_i i^2)^k \\ &= \sum \chi_i (iZ)^{2k} (\sum \chi_i (iY)^2)^k = \mu_z^k (\mu_y^2)^{\frac{k}{2}}. \end{aligned}$$

\square

Lemma A.3. Assume two equably symmetric directing acts $y, z \in \widehat{\mathcal{F}}$ with an identical mean, $E[z] = E[y]$. If $\lambda = \frac{\sigma_z}{\sigma_y}$ then $z =_d \lambda y$.

Proof of Lemma A.3. Without loss of generality, assume $E[y] = E[z] = 0$. To show that $=_d$ it has to be proved that y and z have an identical characteristic function. Writing the characteristic function of z

$$\begin{aligned} \phi_z(t) &= \sum_{l=1}^m e^{it z_l} \chi(z_l) dz = \sum_{l=1}^m \chi(z_l) + it \sum_{l=1}^m z_l \chi(z_l) + \frac{1}{2} (it)^2 \sum_{l=1}^m z_l^2 \chi(z_l) + \dots \\ &= \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mu_z^k = 1 + it \mu_z^1 - \frac{1}{2} t^2 \mu_z^2 - \frac{1}{3!} it^3 \mu_z^3 + \frac{1}{4!} t^4 \mu_z^4 + \dots, \end{aligned} \quad (13)$$

⁴⁰Since $E[y] = E[z]$, y and z can be adjusted such that they can be referred to as having $E[y] = E[z] = 0$.

where μ_z^k , the k th moment around 0, and $\mu_z^0 \equiv 1$ are assumed to exist and be finite. The second equality is by the power series of the exponential function. For the third equality, see Abramowitz and Stegun (1972, p. 928). Writing the k th moment of λy around 0

$$\mu_{\lambda y}^k = \sum_{l=1}^m (\lambda y_l)^k \chi(y_l) = \frac{\sigma_z^k}{\sigma_y^k} \mu_y^k.$$

Since $E[y] = E[z] = 0$ then, when k is even $\sigma_y^k = (\mu_y^2)^{\frac{k}{2}}$ and $\sigma_z^k = (\mu_z^2)^{\frac{k}{2}}$, and when k is odd $\mu_y^k = \mu_z^k = 0$. Note that the outcome spaces of z and λy are identical. Lemma A.2 implies that for an equably symmetric distribution $\mu_{\lambda y}^k = \mu_y^k \frac{(\mu_z^2)^{\frac{k}{2}}}{(\mu_y^2)^{\frac{k}{2}}} = \mu_z^k$ for any k . Therefore, by Equation (13) $\phi_z(t) = \phi_{\lambda y}(t)$, which implies $z =_d \lambda y$. \square

Proof of Theorem 3.5. The subjective probability, $Q(\mathcal{E}_j)$, of event $\mathcal{E}_j \in \Xi$ can be written

$$Q(\mathcal{E}_j) = \psi^{-1}(\psi(p_j - \varphi_j)) = \psi^{-1}\left(\sum_{i=1}^m \chi_i \psi(P_{i,j})\right), \quad (14)$$

for some φ_j . Taking the first-order Taylor approximation of $\psi(p_j - \varphi_j)$ around p_j yields

$$\psi(p_j - \varphi_j) \approx \psi(p_j) + \psi'(p_j)(p_j - \varphi_j - p_j) = \psi(p_j) - \varphi_j \psi'(p_j). \quad (15)$$

Ignoring the weighted summation in the RHS of Equation (14) for the moment, the second-order Taylor approximation of $\psi(P_{i,j})$ around p_j is

$$\psi(P_{i,j}) \approx \psi(p_j) + \psi'(p_j)(P_{i,j} - p_j) + \frac{1}{2} \psi''(p_j)(P_{i,j} - p_j)^2.$$

Since $\psi(p_j)$, $\psi'(p_j)$ and $\psi''(p_j)$ are constants, applying the weighted sum yields

$$\sum_{i=1}^m \chi_i \psi(P_{i,j}) \approx \psi(p_j) + \frac{1}{2} \psi''(p_j) \zeta_j^2. \quad (16)$$

Equating (15) to (16) and organizing terms yields

$$\varphi_j = -\frac{1}{2} \frac{\psi''(p_j)}{\psi'(p_j)} \zeta_j^2.$$

Substituting φ_j into Equation (14) proves the theorem. \square

Proof of Proposition 3.7. By definition

$$\zeta_{1\dots J}^2 = \sum_{i=1}^m \chi_i (P_{i,1\dots J} - p_{1\dots J})^2.$$

Since P is additive and so is p ,

$$\zeta_{1\dots J}^2 = \sum_{i=1}^m \chi_i [(P_{i,1\dots T} - p_{1\dots T}) + (P_{i,T+1\dots J} - p_{T+1\dots J})]^2.$$

Therefore,

$$\begin{aligned}
\zeta_{1\dots J}^2 &= \sum_{i=1}^m \chi_i (P_{i,1\dots T} - p_{1\dots T})^2 + \sum_{i=1}^m \chi_i (P_{i,T+1\dots J} - p_{T+1\dots J})^2 + \\
&\quad 2 \sum_{i=1}^m \chi_i (P_{i,1\dots T} - p_{1\dots T}) (P_{i,T+1\dots J} - p_{T+1\dots J}) \\
&= \zeta_{1\dots T}^2 + \zeta_{T+1\dots J}^2 + 2\zeta_{1\dots T, T+1\dots J}.
\end{aligned}$$

□

Proof of Lemma 3.8. Since $P_i = P_i(\mathcal{E})$ is additive, $P_i^c = P_i(\mathcal{E}^c) = 1 - P_i$. The covariance between $P(\mathcal{E})$ and $P(\mathcal{E}^c)$ takes the form

$$\text{Cov}[P(\mathcal{E}), P(\mathcal{E}^c)] = \sum_{i=1}^m \chi_i (P_i - p) (P_i^c - p^c) = \sum_{i=1}^m \chi_i (P_i - p) (p - P_i),$$

and therefore

$$\text{Cov}[P(\mathcal{E}), P(\mathcal{E}^c)] = -\text{Var}[P(\mathcal{E})].$$

The second equality is obtained by

$$\text{Var}[P(\mathcal{E})] = \sum_{i=1}^m \chi_i (P_i - p)^2 = \sum_{i=1}^m \chi_i (P_i^c - p^c)^2 = \text{Var}[P(\mathcal{E}^c)].$$

□

Proof of Proposition 4.3. This proof considers ambiguity aversion; the proof for ambiguity loving is similar. Let $z = P_j - p_j$ and $y = P_l - p_l$. Assume that z is more ambiguous than y , then by Definition 4.2 $z =_d y + \epsilon$. The DM's preferences \succsim^2 , characterized by the sake utility $\psi : [0, 1] \rightarrow \mathbb{R}$ (see Theorem 3.2), provides

$$\mathbb{E}[\psi(z)] = \mathbb{E}[\mathbb{E}[\psi(y + \epsilon) \mid y]].$$

Ignoring the expectation on the RHS for the moment, ambiguity aversion formed by a concave ψ , implies

$$\mathbb{E}[\psi(y + \epsilon)] \leq \psi(\mathbb{E}[y + \epsilon]) = \psi(y).$$

Taking expectations yields $\mathbb{E}[\psi(z)] \leq \mathbb{E}[\psi(y)]$. Hence, $\hat{f}_l \succsim^2 \hat{f}_j$.

For the opposite direction, let $\hat{f}_l \succsim^2 \hat{f}_j$. Then $\mathbb{E}[\psi(z)] \leq \mathbb{E}[\psi(y)]$. It needs to be shown that there exists an ϵ satisfying Definition 4.2. The proof considers two directing events. The same approach can then be extended to any number of events. Let z and y take two possible values, (z_1, z_2) and (y_1, y_2) , with probabilities $(\alpha, 1 - \alpha)$ and $(\beta, 1 - \beta)$, respectively. Without loss of generality, assume that $z_1 \geq y_1 \geq y_2 \geq z_2$. The random variable ϵ can be constructed

by $\epsilon_1 = (z_1 - y_1, z_2 - y_1)$ with probabilities $\left(\frac{y_1 - z_2}{z_1 - z_2}, \frac{z_1 - y_1}{z_1 - z_2}\right)$ and $\epsilon_2 = (z_1 - y_2, z_2 - y_2)$ with probabilities $\left(\frac{y_2 - z_2}{z_1 - z_2}, \frac{z_1 - y_2}{z_1 - z_2}\right)$. It can be easily verified that the probabilities of ϵ_1 and ϵ_2 are all positive, $E[\epsilon_1 | y_1] = 0$ and $E[\epsilon_2 | y_2] = 0$. Therefore, y and ϵ are mean-independent and $E[z] = E[y + \epsilon] = 0$. The probability that $y + \epsilon = z_1$ is

$$\beta \frac{y_1 - z_2}{z_1 - z_2} + (1 - \beta) \frac{y_2 - z_2}{z_1 - z_2}. \quad (17)$$

Since y and z have the same expectation, $\beta = \frac{z_2 - y_2 + \alpha(z_1 - z_2)}{y_1 - y_2}$. This implies that the probability that $y + \epsilon = z_1$ in Equation (17) is equal to α , and the probability that $y + \epsilon = z_2$ is equal to $1 - \alpha$. That is, $z =_d y + \epsilon$. \square

Proof of Proposition 4.5. Let $z = P_j - p_j$ and $y = P_l - p_l$.

(i) Assume that Definition 4.2 holds. The condition $P_j - p_j =_d P_l - p_l + \epsilon$ implies that $\text{Var}[P_j] = \text{Var}[P_l] + \text{Var}[\epsilon]$, which means that $\zeta_j^2 \geq \zeta_l^2$.

For the opposite direction assuming that Definition 4.4 holds such that $\zeta_j^2 \geq \zeta_l^2$. One can define $\lambda = \frac{\zeta_j}{\zeta_l} = \frac{\sigma_z}{\sigma_y} \geq 1$. Since z and y are equably symmetric distributed and $E[z] = E[y] = 0$, λy is also equably symmetric distributed with zero mean, $\lambda^2 \text{Var}[y] = \text{Var}[z]$ and by Lemma A.3 $z =_d \lambda y$. Write $x + y = \alpha(x + \lambda y) + (1 - \alpha)x$, where $\alpha = \frac{1}{\lambda}$ and x is a random variable satisfying $E[x] = 0$. By the Jensen inequality $\psi(x + y) \geq \alpha\psi(x + \lambda y) + (1 - \alpha)\psi(x)$. Applying expectation for both sides yields

$$E[\psi(x + y)] \geq \alpha E[\psi(x + \lambda y)] + (1 - \alpha) E[\psi(x)]. \quad (18)$$

Since $\lambda \geq 1$, By Definition 4.2, $x + \lambda y$ is more ambiguous than x , thus $E[\psi(x)] \geq E[\psi(x + \lambda y)]$. Together with (18), this implies $E[\psi(x + y)] \geq E[\psi(x + \lambda y)]$. Let $x = 0$. Then $E[\psi(y)] \geq E[\psi(\lambda y)] = E[\psi(z)]$, which means $y \succsim^2 z$.

(ii) Let $\psi(z) = -e^{-\eta z}$. Taking a second-order Taylor approximation around $z = 0$ yields $\psi(z) \approx -1 + \eta z - \frac{1}{2}\eta^2 z^2$. Since $E[z] = 0$, taking expectation yields $E[\psi(z)] \approx -1 - \frac{1}{2}\eta^2 \text{Var}[z]$. This means that $y \succsim^2 z$ iff $\text{Var}[y] \leq \text{Var}[z]$. That is, iff $\zeta_l^2 \leq \zeta_j^2$.

(iii) Let $\psi(z) = -(z - \alpha)^2$, where $z \leq \alpha$ for some α . The expected sake function is $E[\psi(z)] = -(\text{Var}[z] + (E[z] - \alpha)^2)$. Since $E[z] = E[y] = 0$, then $y \succsim^2 z$ iff $\text{Var}[y] \leq \text{Var}[z]$. That is, iff $\zeta_l^2 \leq \zeta_j^2$. \square

Proof of Model 4.7. Assume two symmetric acts with an identical expected outcome and the same expected probabilities. The first equality is derived by Theorem 4.8, which proves that any ambiguity-averse DM prefers the act with the lower \mathcal{U}^2 over the act with the higher

\mathcal{U}^2 . The second equality is obtained by the fact that the variance of the probability of loss equals the variance of the probability of gain (Lemma 3.8). \square

Proof of Theorem 4.8. Assume that $\mathcal{U}^2[g] \leq \mathcal{U}^2[f]$. By Equation (5) the value assigned by a risk-neutral DM to act g is

$$\begin{aligned} V(g) &= \sum_{j=1}^n p_j x_j + \frac{1}{2} \sum_{j=1}^k \left[\frac{\psi''(p_{1\dots j})}{\psi'(p_{1\dots j})} \zeta_{1\dots j}^2 - \frac{\psi''(p_{1\dots j-1})}{\psi'(p_{1\dots j-1})} \zeta_{1\dots j-1}^2 \right] x_j + \\ &\quad \frac{1}{2} \sum_{j=k+1}^n \left[\frac{\psi''(p_{j\dots n})}{\psi'(p_{j\dots n})} \zeta_{j\dots n}^2 - \frac{\psi''(p_{j+1\dots n})}{\psi'(p_{j+1\dots n})} \zeta_{j+1\dots n}^2 \right] x_j. \end{aligned}$$

Since $\mathcal{U}^2[g] \leq \mathcal{U}^2[f]$ and act f is first-order stochastically dominated by act g , by Proposition 4.5, any $\mathcal{E}_j \subseteq \mathbb{S}$ satisfies $\zeta_{f,j}^2 = \zeta_{g,j}^2 + \text{Var}[\epsilon]$, where ϵ and P_j are mean-independent. The subjective probability of event $\mathcal{E}_j \subseteq \mathbb{S}$ under act f is, thus,

$$Q(\mathcal{E}_j) = p_j + \frac{1}{2} \frac{\psi''(p_j)}{\psi'(p_j)} (\zeta_j^2 + \text{Var}[\epsilon]).$$

Therefore, the value of act f is

$$\begin{aligned} V(f) &= \sum_{j=1}^n p_j x_j + \\ &\quad \frac{1}{2} \sum_{j=1}^k \left[\frac{\psi''(p_{1\dots j})}{\psi'(p_{1\dots j})} (\zeta_{1\dots j}^2 + \text{Var}[\epsilon_{1\dots j}]) - \frac{\psi''(p_{1\dots j-1})}{\psi'(p_{1\dots j-1})} (\zeta_{1\dots j-1}^2 + \text{Var}[\epsilon_{1\dots j-1}]) \right] x_j + \\ &\quad \frac{1}{2} \sum_{j=k+1}^n \left[\frac{\psi''(p_{j\dots n})}{\psi'(p_{j\dots n})} (\zeta_{j\dots n}^2 + \text{Var}[\epsilon_{j\dots n}]) - \frac{\psi''(p_{j+1\dots n})}{\psi'(p_{j+1\dots n})} (\zeta_{j+1\dots n}^2 + \text{Var}[\epsilon_{j+1\dots n}]) \right] x_j, \end{aligned}$$

which implies that

$$\begin{aligned} V(f) - V(g) &= \frac{1}{2} \sum_{j=1}^k \left[\frac{\psi''(p_{1\dots j})}{\psi'(p_{1\dots j})} \text{Var}[\epsilon_{1\dots j}] - \frac{\psi''(p_{1\dots j-1})}{\psi'(p_{1\dots j-1})} \text{Var}[\epsilon_{1\dots j-1}] \right] x_j + \\ &\quad \frac{1}{2} \sum_{j=k+1}^n \left[\frac{\psi''(p_{j\dots n})}{\psi'(p_{j\dots n})} \text{Var}[\epsilon_{j\dots n}] - \frac{\psi''(p_{j+1\dots n})}{\psi'(p_{j+1\dots n})} \text{Var}[\epsilon_{j+1\dots n}] \right] x_j. \end{aligned} \quad (19)$$

Organizing the terms of Equation (19) yields

$$\begin{aligned} V(f) - V(g) &= \frac{1}{2} \frac{\psi''(p_{1\dots K})}{\psi'(p_{1\dots K})} \text{Var}[\epsilon_{1\dots K}] x_k + \frac{1}{2} \sum_{j=1}^{k-1} \frac{\psi''(p_{1\dots j})}{\psi'(p_{1\dots j})} \text{Var}[\epsilon_{1\dots j}] (x_j - x_{j+1}) + \\ &\quad \frac{1}{2} \frac{\psi''(p_{K+1\dots n})}{\psi'(p_{K+1\dots n})} \text{Var}[\epsilon_{K+1\dots n}] x_{k+1} + \frac{1}{2} \sum_{j=k+2}^n \frac{\psi''(p_{j\dots n})}{\psi'(p_{j\dots n})} \text{Var}[\epsilon_{j\dots n}] (x_j - x_{j-1}). \end{aligned} \quad (20)$$

Since the DM is ambiguity averse, i.e., $\frac{\psi''(p_{j\dots n})}{\psi'(p_{j\dots n})} < 0$ for any $1 \leq j \leq n$, and $0 \leq (x_{j+1} - x_j)$, the second component in the first line of Equation (20) is positive, while the second component in the second line of Equation (20) is negative. Because acts are symmetric with $x_k \leq x_s$, the absolute value of the negative component is greater than the positive component. Thus, their

sum is negative. The first components in the first line and in the second line of Equation (20) are both negative; therefore, $V(f) - V(g) \leq 0$, which implies $g \succ f$.

For the opposite direction, assume $V(g) \geq V(f)$. Since all the parameters in the value functions $V(f)$ and $V(g)$ of acts f and g are identical, except $\text{Var}[P_L]$, and first-order stochastic dominance is satisfied, Equation (20) implies that $\text{Var}[\epsilon] \geq 0$. Thus, $\text{Var}[P_L]$ of act f is greater than $\text{Var}[P_L]$ of act g . According to Lemma 3.8 this is also true for the probability of gain P_G ; therefore, $\mathcal{U}^2[f] \geq \mathcal{U}^2[g]$. \square

Proof of Theorem 5.1. The first-order Taylor approximation of the LHS of Equation (9) with respect to $E[x]$ is

$$LHS = U(E[x] - \mathcal{K}) = \sum_{j=1}^n p_j U(E[x] - \mathcal{K}) \approx \sum_{j=1}^n p_j (U(E[x]) - \mathcal{K}U'(E[x])).$$

Writing the RHS of Equation (9) as

$$RHS = \underbrace{\sum_{j=1}^n p_j U(x_j)}_I - \underbrace{\left(\sum_{j=1}^k (\varphi_{1\dots J} - \varphi_{1\dots J-1}) U(x_j) + \sum_{j=k+1}^n (\varphi_{J\dots N} - \varphi_{J+1\dots N}) U(x_j) \right)}_{II},$$

the second-order Taylor approximation of I around $E[x]$ is

$$\begin{aligned} I &\approx \sum_{j=1}^n p_j \left(U(E[x]) + U'(E[x])(x_j - E[x]) + \frac{1}{2}U''(E[x])(x_j - E[x])^2 \right) \\ &= U(E[x]) + \frac{1}{2}U''(E[x])\text{Var}[x]. \end{aligned}$$

Writing

$$\begin{aligned} II &= \varphi_{1\dots K}U(x_k) + \varphi_{K+1\dots N}U(x_{k+1}) + \\ &\quad \sum_{j=1}^{k-1} \varphi_{1\dots J}[U(x_j) - U(x_{j+1})] + \sum_{j=k+2}^n \varphi_{J\dots N}[U(x_j) - U(x_{j-1})] \end{aligned}$$

and taking the first-order Taylor approximation of II around $E[x]$ yields⁴¹

$$\begin{aligned} II &\approx \varphi_{1\dots K}U(x_k) + \varphi_{K+1\dots N}U(x_{k+1}) + \\ &\quad \sum_{j=1}^{k-1} \varphi_{1\dots J}U'(E[x])(x_j - x_{j+1}) + \sum_{j=k+2}^n \varphi_{J\dots N}U'(E[x])(x_j - x_{j-1}). \end{aligned}$$

Since outcomes are assumed to be symmetrically distributed and x_k is relatively close to $E[x]$

$$II \approx \varphi_{1\dots K}U(x_k) + \varphi_{K+1\dots N}U(x_{k+1}).$$

Because $U(x_k) = 0$, and $U(\cdot)$ is almost linear around the reference point, x_k , then $U(x_k) \approx$

⁴¹Recall that φ_j is the probability premium, holding an order of magnitude of the variance of probability. Thus, φ_j is smaller by one order of magnitude than probabilities.

$U(x_{k+1}) \approx U'(E[x]) E[x]$.⁴² Therefore,

$$II \approx \frac{1}{4} \left[-\frac{1}{2} \frac{\psi''(p_L)}{\psi'(p_L)} - \frac{1}{2} \frac{\psi''(p_G)}{\psi'(p_G)} \right] \mathcal{U}^2[x] U'(E[x]) E[x].$$

Combining the LHS, the RHS, I and II , the uncertainty premium is

$$\mathcal{K} \approx -\frac{1}{2} \frac{U''(E[x])}{U'(E[x])} \text{Var}[x] - \frac{1}{8} \left[\frac{\psi''(p_L)}{\psi'(p_L)} + \frac{\psi''(p_G)}{\psi'(p_G)} \right] \mathcal{U}^2[x] E[x].$$

□

Proof of Corollary 5.2. CRRA implies $U'(x) = x^{-\gamma}$ and $U''(x) = -\gamma x^{-\gamma-1}$. CAAA implies $\psi'(q_i) = e^{-\eta q_i}$ and $\psi''(q_i) = -\eta e^{-\eta q_i}$. Substituting in Theorem 5.1 proves the corollary.

□

⁴²See, for example, Segal and Spivak (1990) and Levy *et al.* (2003).