Rochester Center for Economic Research

Ambiguity, Risk and Asset Returns in Continuous Time

Zengjing Chen and Larry G. Epstein

Working Paper No. 474
July 2000

UNIVERSITY OF
ROCHESTER

AMBIGUITY, RISK AND ASSET RETURNS IN CONTINUOUS TIME

Zengjing Chen Larry Epstein*

July 2000

Abstract

Existing models of utility in stochastic continuous-time settings assume that beliefs are represented by a probability measure. As illustrated by the Ellsberg Paradox, this feature rules out a priori any concern with ambiguity. This paper formulates a continuous-time intertemporal version of multiple-priors utility, where aversion to ambiguity is admissible. When applied to a representative agent asset market setting, the model delivers restrictions on excess returns that admit interpretations reflecting a premium for risk and a separate premium for ambiguity.

1. INTRODUCTION

1.1. Outline

It is intuitive that many choice situations feature 'Knightian uncertainty' or 'ambiguity' and that these are distinct from 'risk'. The Ellsberg Paradox and related evidence have demonstrated that such a distinction is behaviorally meaningful. However, the distinction is not permitted within the subjective expected utility framework, or even more broadly, if preference is 'based on probabilities' in the precise formal sense of (a slight variation of) probabilistic sophistication [37]. Because continuous-time modeling has *universally* assumed that preference is probabilistically sophisticated, it has focussed on risk and risk aversion as the important characteristics of choice situations, to the exclusion of a role for ambiguity. This paper presents a formulation of utility in continuous-time that permits a distinction

*Chen is at the Department of Mathematics, Shandong University, Jinan, 250100, China and Epstein is at the University of Rochester, lepn@troi.cc.rochester.edu. Some of this work was completed while Epstein was visiting the HKUST and U. Paris-Dauphine and while Chen was visiting INRIA; the hospitality of these institutions is gratefully acknowledged. Financial support was provided by the National Natural Science Foundation of China, the Young Foundation of Shandong U. and by the NSF (grant SES-9972442). We have benefitted from discussions with Rose-Anne Dana, Nicole El Karoui, Lars Hansen, Xing Jin, Massimo Marinacci, Angelo Melino, Jianjun Miao, Werner Ploberger, Martin Schneider, Nizar Touzi, Tan Wang and especially Shige Peng. We are grateful also for the suggestions of two referees and Drew Fudenberg.

between risk aversion and ambiguity aversion, as well as a further distinction between these and the willingness to substitute intertemporally. This *three-way* distinction is accomplished through an extension of stochastic differential utility [14] whereby the usual single prior is replaced by a set of priors, as in the atemporal model of Gilboa and Schmeidler [28]. We call the resulting model *recursive multiple-priors utility*.[1]

Our model of utility is the continuous-time counterpart of that in [23, 24].[2] A small part of the value-added here is that, by exploiting recent advances in understanding the precise behavioral meaning of ambiguity (aversion), we are able to provide firmer foundations for claims regarding the role of ambiguity in utility and implied behavior.[3] For example, there is no counterpart in [23, 24] of Theorems 4.1 or 4.4, the latter of which formalizes the noted three-way separation and justifies our use of terms such as 'ambiguity premium'.

However, the main value-added relative to discrete-time models is in tractability for applications. It is well known that continuous-time modeling affords considerable analytical advantages, which is the reason that it is the dominant framework in finance. These analytical advantages are manifested here in our application of recursive multiple-priors utility to a representative agent asset pricing setting to study the effects of the ambiguity associated with asset returns. We show (Section 5) that excess returns for a security can be expressed as a sum of a risk premium and an ambiguity premium. We elaborate shortly (Section 1.3) on the potential usefulness of such a result and more generally, of admitting that security returns embody both risk and ambiguity, for addressing two long-standing empirical puzzles. At this point, we wish to emphasize that *none* of the asset pricing results and potential applications discussed in this paper are discussed in [23, 24]. Their focus is on the connection between ambiguity and the indeterminacy of equilibrium. In particular, a decomposition of excess returns into risk and ambiguity premia is not presented, nor is it apparent in the discrete-time framework, though it jumps off the page in the continuous-time setting.[4]

In the rest of this introduction, we revisit the Ellsberg Paradox, elaborate on potential asset market applications and then discuss a related model of robust decision-making. In Section 2, we proceed to the specification of recursive multiple-priors utility. This is accomplished in stages, beginning with an outline of the essential ingredients of the atemporal model. Section 3 provides several examples. Properties of the utility function are examined in Section 4. The application to asset pricing is provided in Section 5. Proofs are collected in appendices.

[1] To explain this nomenclature, note that stochastic differential utility is the continuous-time counterpart of recursive utility [26].

[2] A formal demonstration that the present continuous-time model is a suitable limit of the Epstein-Wang discrete-time model is the subject of work in progress [22].

[3] These recent advances include [19], [25] and [27].

[4] The potential applications of recursive multiple-priors utility are not limited to asset pricing; see Section 1.4.

1.2. Ellsberg Revisited

Though the Ellsberg Paradox is likely familiar to many readers, it may be useful to translate it into our intertemporal setting. Consider, therefore, the following scenario: An investor with horizon $[0, T]$ faces primitive uncertainty represented by a 2-dimensional state process $W_t = (W_t^1, W_t^2)$. Consider four consumption programs $\{c^i\}_{i=1}^4$. All have the structure

$$c_t^i = \begin{cases} 0 & \text{if } 0 \leq t < \tau \\ \xi^i & \text{if } \tau \leq t \leq T, \end{cases}$$

where $0 < \tau < T$ and the random variables ξ^i are given by

$$\xi^1 = 1_{\{W_\tau^1 > a_1\}}, \xi^2 = 1_{\{W_\tau^1 < a_1\}}, \xi^3 = 1_{\{W_\tau^2 > a_2\}}, \xi^4 = 1_{\{W_\tau^2 < a_2\}}.$$

In particular, for each consumption process all uncertainty is resolved at τ. Consider the following rankings:

$$c^1 \sim c^2 \succ c^3 \sim c^4. \tag{1.1}$$

A possible interpretation is that W_t^1 describes the state of the 'home' economy (measured by the level of the Dow Jones, for example) and W_t^2 describes the state of the 'foreign' economy (measured by the Hong Kong Hang Seng index). The investor feels that W_τ^i is as likely to exceed a_i as to fall short of it. At the same time, as an American, she is more familiar with the NYSE, which leads her to prefer either bet on the Dow Jones to either bet on the Hang Seng index. These preferences are impossible, (assuming that each event $\{W_\tau^i = a_i\}$ is viewed as null), if rankings are based on probabilities. That is because then $c^1 \sim c^2$ and $c^3 \sim c^4$ imply that each of the above 4 events has probability $1/2$ which leaves no room for a distinction between betting on the home stock index versus the foreign index. The general point is that while a probability measure can represent likelihood assessment, it cannot model also the other dimensions of beliefs emphasized by Knight [34] and Keynes [33], namely, confidence in that likelihood assessment or the weight of supporting evidence. We follow Ellsberg in using 'ambiguity' and 'ambiguity aversion' in referring to this added dimension. We have more to say below on the precise meaning of these terms (Section 4).

The standard expected additive utility function in continuous-time specifies time t utility by

$$V_t = E\left[\int_t^T e^{-\beta(s-t)} u(c_s)\, ds \mid \mathcal{F}_t\right], \tag{1.2}$$

where c is a consumption process, $\mathcal{F}_t$ denotes information available at time t and the expectation is computed with respect to the probability measure that represents beliefs. The representation of beliefs through a single probability measure imposes *a priori* the decision-maker's indifference to ambiguity. Many generalizations of (1.2) have been studied and applied, where intertemporal nonseparabilities are introduced in order to accommodate habit formation, learning-by-doing or a distinction between risk aversion and intertemporal substitution [14], for example. However, *all* of these models are probabilistically sophisticated (in the sense of Section 4.1) and thus exhibit indifference to ambiguity as an *a priori* feature.

1.3. Ambiguity in Asset Markets

The importance of the Ellsberg Paradox is that it is strongly suggestive of the importance of ambiguity also in nonexperimental settings. Asset markets provide an obvious instance. The risk-based models that constitute the paradigm in this literature have well documented empirical failures; and introspection suggests (at least to us) that 'ambiguity' is at least as prominent as risk in making investment decisions. An illustration of the potential usefulness of recognizing the presence of ambiguity is provided by the equity premium puzzle [41] - the failure of the representative agent model to fit historical averages of the equity premium and the risk-free rate. One aspect of the puzzle is that an implausible degree of risk aversion is needed to rationalize the observed equity premium. Naturally, the equity premium is viewed as a premium for the greater riskiness of equity. The alternative view that is suggested by our analysis is that part of the premium is due to the greater ambiguity associated with the return to equity, which reduces the required degree of risk aversion.

Another potential role for ambiguity is in addressing the home-bias puzzle, whereby investors in many countries invest 'too little' in foreign securities. Naturally, 'too little' is from the perspective of a model where securities are differentiated only via their risk characteristics. However, if foreign securities are more ambiguous than domestic ones, much as illustrated by the above Ellsberg-type example, then admitting this possibility into the model may help to resolve the puzzle. This approach has been developed, with some success, in [21].[5]

To provide some perspective on our modeling approach based on ambiguity, consider two issues that may have already occurred to readers, namely, (i) observational equivalence and (ii) learning.

For (i), consider the alternative deviation from rational expectations modeling whereby we continue to assume probabilistic sophistication (a single prior) but relax the rational expectations hypothesis that the agent knows and employs the true probability law. This approach is adopted in [1] and [8] in order to address the equity premium puzzle. Our model ultimately delivers a 'distorted probability measure', selected endogenously from the agent's set of priors, that would deliver the identical representative agent equilibrium were it adopted as a primitive specification of beliefs. In spite of this form of observational equivalence, our approach has several advantages.

First, there is an appeal to basing an explanation of asset market behavior on a phenomenon, namely ambiguity aversion, that is plausibly important in a variety of settings, rather than on a particular and invariably ad hoc specification of erroneous beliefs. Second, an agent using the wrong probability measure may plausibly be aware of this possibility and thus be led to seek robust decisions. Such self-awareness and a desire for robust decisions lead naturally to consideration of sets of priors.

Finally, while it is true that any excess return that can be generated by our model could also be delivered by a model in which equity is viewed exclusively as risky but where perceived riskiness is relative to erroneous beliefs, the latter story would still require a large degree of

[5] Once again, continuous-time plays an important role.

risk aversion and thus would not resolve the puzzle as posed by Mehra and Prescott. Put another way, the puzzle concerns not only the historical equity premium but also behavior in other settings and introspection regarding plausible choices between hypothetical lotteries - these are used to determine the range of plausible risk aversion. Implicit is that the prospects involved in all these settings are purely risky, justifying the transfer of preference parameters across settings. Our working hypothesis is that prospects faced in asset market are qualitatively different than hypothetical lotteries where prizes are determined by the outcome of a coin flip, for example.

The second natural question concerning our model is "would ambiguity not disappear eventually as the agent learns about her environment?" For example, given an Ellsberg urn containing balls of various colors in unknown proportions, it is intuitive that the true color composition would be learned asymptotically if there is repeated sampling (with replacement) from the urn. However, intuition is different for the modified setting where there is a sequence of ambiguous Ellsberg urns, each containing balls of various colors in unknown proportions, and where sampling is such that with the n^{th} draw is made from the n^{th} urn. Suppose further that the agent views the urns as 'identical and independent'.[6] Given such a prior view, one would *not* expect ambiguity to vanish. Indeed, Marinacci [38] proves a LLN result appropriate for beliefs represented by a set of priors in which the connection between empirical frequencies and asymptotic beliefs is weakened *to a degree that depends on the extent of ambiguity in prior beliefs*. Asymptotically, the decision-maker believes that the limit frequency of any given color lies in an interval, where the interval collapses to a point if there is no ambiguity in prior beliefs but not more generally. While Marinacci's framework does not accommodate our continuous-time setting, his analysis is nevertheless strongly suggestive, with the increments $\{dW_t : t \geq 0\}$ of the driving state process (W_t) constituting the counterpart of the set of Ellsberg urns.[7] (See the end of Section 2.4 for further discussion of learning.)

1.4. Robustness

A related and independently developed model, with similar motivation, is described in [3]. In describing motivation for their surrounding research agenda, Hansen and Sargent [30] translate Ellsberg-type phenomena and speak in terms of agents (consumer/investors or macroeconomic policy makers) facing *model uncertainty* and therefore seeking to make decisions that are *robust* to this uncertainty. Though their model bears some relation to the Gilboa-Schmeidler multiple-priors model and to ours, that relation is more one of spirit rather than precise form as it is here (they have no counterpart of (3.6), for example). Nevertheless,

[6] Given a set $\mathcal{P}$ of priors on a state space $\Pi_{n=1}^{N}\Omega_n$, denote by $\mathcal{P}_n$ the set of marginals on Ω_n. Following [28, p. 150], say that the Ω_n's are independent under $\mathcal{P}$ if $\mathcal{P}$ coincides with the closed convex hull of $\{p_1 \otimes ... \otimes p_n : p_n \in \mathcal{P}_n \text{ all n}\}$. For an infinite Cartesian product, require the above property to hold for $\Pi_{n=1}^{N}\Omega_n$ for all N. Marinacci [38, pp. 157-8] adopts a weaker notion of independence. Finally, define 'identical' in the obvious way, namely by $\mathcal{P}_n = \mathcal{P}_m$ for all n and m.

[7] The assumption that increments are 'identical and independent' is satisfied in the special case of our model called IID ambiguity (Section 3.4); see also Theorem 4.5.

the commonality in motivation and spirit with our model suggests that the macroeconomic applications discussed by these authors (see also [5]) are potential applications of our model as well.

We will not discuss the substantial differences in formal details but there is an important difference in emphasis between the two approaches that we would like to mention. The approach in the cited papers, and also in the related discrete-time model [31], has its roots in robust control theory rather than in decision theory. As a result, these authors give less weight to behavioral foundations, for example, in order to justify interpretations for utility parameters. These interpretations are based instead on what seems natural given the functional form or on common practice in the robust control literature. However, if the functional form is observable only indirectly through behavior, then to be empirically meaningful, a notion of robust decision-making must be expressed in terms of observable (at least in principle) choice behavior.[8] Corresponding behavioral foundations for our model are provided by [25] and [27], which underly our Theorems 4.1, 4.3, 4.4 and 4.5, all of which play a role in supporting the interpretations we suggest for our utility function and subsequent analysis.

To illustrate the importance of focussing on behavior, note that, as emphasized in [3], the model proposed there is formally a special case of stochastic differential utility, which, as has already been observed, is probabilistically sophisticated. It follows that the model of robust decision-making is inconsistent with the typical Ellsberg choices (1.1). However, if it does not correspond to Ellsberg-type behavior, then what is 'robust decision-making'? It may very well be possible to provide a satisfactory answer to this question; our point is simply that it is not addressed in [3] and the related literature.

2. MULTIPLE-PRIORS UTILITY

2.1. Atemporal Model

Consider an atemporal or one-shot choice setting where uncertainty is represented by the measurable state space $(\Omega, \mathcal{F})$. The decision-maker ranks uncertain prospects or acts, maps from Ω into an outcome set $\mathcal{X}$. According to the multiple-priors model, the utility $U(f)$ of any act f has the form:[9]

$$U(f) = \min_{Q \in \mathcal{P}} \int u(f)\, dQ, \tag{2.1}$$

[8] A pertinent analogy is that in the vNM theory of preferences over lotteries, risk aversion is defined as a property of preference - that any lottery F is ranked below the lottery that delivers the mean of F with certainty. Concavity of the vNM index is a *characterization* of risk aversion, rather than its definition, and it has significance *only* because it characterizes a property of preference.

[9] The set $\mathcal{P}$ is required to be weakly compact (the weak topology is that induced by the set of bounded measurable functions) and convex. Because $\mathcal{P}$ and its closed convex hull generate the identical utility function, closedness and convexity are normalizations that ensure uniqueness. See [28] for further details. Note, however, that probability measures are assumed there to be only finitely additive, while we assume countable additivity.

where $u : \mathcal{X} \longrightarrow \mathcal{R}^1$ is a von Neumann-Morgenstern utility index and $\mathcal{P}$ is a subjective *set* of probability measures on $(\Omega, \mathcal{F})$. The subjective expected utility model is obtained when the set of priors $\mathcal{P}$ is a singleton. Intuitively, the multiplicity of priors in the general case models the ambiguity of the likelihoods of events and the infimum delivers aversion to such ambiguity.

In anticipation of the technical requirements of continuous time, consider a specialization of the multiple-priors model for which all priors in $\mathcal{P}$ are uniformly absolutely continuous with respect to some P in $\mathcal{P}$.[10] Then $\mathcal{P}$ may be identified with its set H of densities with respect to P, where $H \subset L^1_+(\Omega, \mathcal{F}, P)$ is weakly compact. The identification is via

$$\mathcal{P} = \{h\, dP : h \in H\}.$$

For further details and behavioral implications of this added structure see [24, Section 2] and [39].

2.2. Continuous Time

Consider a finite horizon model, where time t varies over $[0, T]$. Other primitives include:

- a probability space $(\Omega, \mathcal{F}, P)$
- a standard d-dimensional Brownian motion $W_t = (W_t^1, \cdots, W_t^d)^\top$ defined on $(\Omega, \mathcal{F}, P)$
- the Brownian filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$, where $\mathcal{F}_t$ is generated by $\sigma(W_s : s \leq t)$ and the P-null sets of $\mathcal{F}$, $\mathcal{F}_T = \mathcal{F}$.

The measure P is part of our description of the consumer's preference and, for that purpose, it is significant only for defining null sets; any equivalent (mutually absolutely continuous) measure would do as well. In particular, P is not necessarily the 'true' measure (with the exception of Section 5).

Consumption processes c take values in C, a convex subset of R^ℓ.[11] Our objective is to formulate a utility function on the domain D of C-valued consumption processes. It is natural to consider a process of utility values (V_t) for each c, where V_t is the utility of the continuation $(c_s)_{s \geq t}$ and V_0 is the utility of the entire process c.

In the case of risk, where P represents the consumer's assessment of likelihoods, Duffie and Epstein [14] define stochastic differential utility (SDU). For any given c in D, the SDU process (V_t^P) is defined as the solution to the integral equation

$$V_t^P = E\left[\int_t^T f(c_s, V_s^P)\, ds \mid \mathcal{F}_t\right], \tag{2.2}$$

[10] $\mathcal{P}$ is *uniformly absolutely continuous* with respect to P if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $E \in \mathcal{F}$ and $P(E) < \delta$ imply $Q(E) < \varepsilon$, $\forall Q \in P$.

[11] In this paper, $x = (x_t)$ denotes a *process*, by which we mean that it is: (i) progressively measurable, that is, (for each t) $x : [0, t] \times (\Omega, \mathcal{F}_t) \longrightarrow \mathrm{R}^\ell$ is product measurable, and (ii) square integrable, that is, $E \int_0^T |x_s|^2 ds < \infty$. The set of all such processes is a Hilbert space under the obvious inner product. Inequalities in random variables are understood to hold P a.e., while those involving stochastic processes are understood to hold $dt \otimes dP$ a.e.

Here the function f is a primitive of the specification, called an *aggregator*. The standard expected utility specification (1.2) is obtained in the special case $f(c,v) = u(c) - \beta v$.

The limitation of SDU from the present perspective is that because all expectations are taken with respect to the single probability measure P, the consumer is indifferent to ambiguity. In the next three sections, we describe a generalization of SDU in which the consumer uses a set $\mathcal{P}$ of measures as in the atemporal multiple-priors model.

2.3. Density Generators

Our construction of the set $\mathcal{P}$ of measures on $(\Omega, \mathcal{F}_T)$ begins here. To understand the nature of the construction, think of a discrete-time event tree [12, p. 104], where nature determines motion through the tree and where $\mathcal{F}_T$ describes the set of terminal states or events. Fix a reference probability measure P on $\mathcal{F}_T$. At each time and state in the tree, the decision-maker's conditional beliefs about the state to be reached next period are represented by a set of densities with respect to the conditional measure induced by P. The set of densities determines a set of conditional probability measures over the state next period. Finally, the sets of conditional-one-step-ahead measures for all time-state pairs can be combined in the usual probability calculus way to deliver a set $\mathcal{P}$ of measures on $\mathcal{F}_T$. (In this construction, admit all possible selections of a conditional measure at each time-state pair.) In the corresponding continuous-time construction, the process that delivers the counterpart of the (logarithm of) a conditional-one-step-ahead density for each time and state is called a density generator; it is defined formally below. The reason for restricting attention to sets of priors $\mathcal{P}$ constructed in this way is explained following Theorem 2.3.

Turn to the formal details. A *density generator* is an R^d-valued process $\theta = (\theta_t)$ satisfying

$$E\left[\exp\left(\frac{1}{2}\int_0^T \mid \theta_s \mid^2 ds\right)\right] < \infty. \tag{2.3}$$

The latter condition ensures [13, p. 288] that the process (z_t^θ) is a P-martingale, where

$$dz_t^\theta = -z_t^\theta \, \theta_t \cdot dW_t, \;\; z_0^\theta = 1, \;\; \text{that is,}$$

$$z_t^\theta \equiv \exp\left\{-\frac{1}{2}\int_0^t \mid \theta_s \mid^2 ds - \int_0^t \theta_s \cdot dW_s\right\}, \; 0 \le t \le T.$$

Because $1 = z_0^\theta = E\left[z_T^\theta\right]$, z_T^θ is a P-density on $\mathcal{F}_T$. Consequently, θ generates a probability measure Q^θ on $(\Omega, \mathcal{F})$ that is equivalent to P, where

$$Q^\theta(A) = E\left[1_A \, z_T^\theta\right], \;\; \text{for all } A \text{ in } \mathcal{F}_T.$$

In other words,

$$\frac{dQ^\theta}{dP} = z_T^\theta; \text{ more generally, } \left.\frac{dQ^\theta}{dP}\right|_{\mathcal{F}_t} = z_t^\theta \text{ for each } t. \tag{2.4}$$

In the special case where $\theta_t \equiv 0$ for all t, then $z_t^\theta \equiv 1$ and $Q^\theta = P$. In order to permit the decision-maker to have a nonsingleton set of priors (containing P), we specify a set Θ of density generators (containing 0). As explained following Theorem 2.3, a central feature of our model is that we specify Θ by beginning with a set $\{\Theta_t\}_{t\in[0,T]}$ of correspondences from Ω into R^d. Thus, for each t, let

$$\Theta_t : \Omega \rightsquigarrow \mathrm{R}^d.$$

Assume the following properties:

Uniform Boundedness There is a compact subset $\mathcal{K}$ in R^d such that $\Theta_t : \Omega \rightsquigarrow \mathcal{K}$ for each t.

Compact-Convex Each Θ_t is compact-valued and convex-valued.

Measurability The correspondence $(t,\omega) \longmapsto \Theta_t(\omega)$, when restricted to $[0,s] \times \Omega$, is $\mathcal{B}([0,s]) \times \mathcal{F}_s$-measurable for any $0 < s \leq T$.[12]

Normalization $0 \in \Theta_t(\omega)\ dt \otimes dP\ a.e.$

Define the set of density generators by[13]

$$\Theta = \{(\theta_t) : \theta_t(\omega) \in \Theta_t(\omega) \;\; dt \otimes dP \; a.e.\}\,. \tag{2.5}$$

Roughly speaking, (2.5) restricts Θ to equal the Cartesian product of its projections and thus we refer to any Θ constructed in this way as being *rectangular*. An example of a set Θ that is excluded thereby is

$$\Theta = \{(\theta_t) : E\left[\int_0^T \mid \theta_s \mid^2 \; ds\right] \leq \ell\}, \tag{2.6}$$

where $\ell \geq 0$ is a parameter.

We proceed to describe the implied properties of Θ that will be used later. Say that Θ is *stochastically convex* if for any real-valued process (λ_t) with $0 \leq \lambda_t \leq 1$,

$$\theta \text{ and } \theta' \text{ in } \Theta \text{ implies that } (\lambda_t\theta_t + (1-\lambda_t)\theta'_t) \in \Theta.$$

Abbreviate $L^\infty([0,T]\times\Omega, \mathcal{B}([0,T]) \otimes \mathcal{F}_T, dt \otimes dP)$ by $L^\infty([0,T]\times\Omega)$ and similarly for L^1.

Lemma 2.1. *The set of density generators Θ satisfies:*

(a) $0 \in \Theta$ and $\sup\{\| \theta \|_{L^\infty([0,T]\times\Omega)} \colon \theta \in \Theta\} < \infty$.

(b) For any R^d-valued process (σ_t), there exists $(\theta_t^*) \in \Theta$ such that

$$\theta_t^* \cdot \sigma_t = \max_{\theta\in\Theta} \theta_t \cdot \sigma_t = \max_{y\in\Theta_t} y \cdot \sigma_t\,. \tag{2.7}$$

(c) Θ is stochastically convex and weakly compact in $L^1([0,T]\times\Omega)$.

[12] That is, $\{(t,\omega) \in [0,s]\times\Omega : \Theta_t(\omega) \cap K \neq \emptyset\} \in \mathcal{B}([0,s]) \times \mathcal{F}_s$ for each compact $K \subset \mathcal{K}$. See [2, section 14.12].

[13] Uniform Boundedness ensures that (2.3) is satisfied by any $\theta \in \Theta$ and hence that each Q^θ is well-defined.

Part (a) describes a normalization and also the norm-boundedness of Θ. Though the existence of $\max_{y\in\Theta_t(\omega)} y \cdot \sigma_t$ is apparent for each (t,ω) pair, (b) ensures that the maximizers $\theta_t^*(\omega)$ can be chosen to satisfy the measurability needed in order that $\theta^* = (\theta_t^*)$ constitute a process. Then the single process θ^* achieves the first maximum in (2.7) for every t and there is equality between the two maximizations shown.

The primitive $\{\Theta_t\}$ can be represented in an alternative way that is sometimes more convenient. Because each Θ_t is convex-valued, we can use the theory of support functions to provide a reformulation of the preceding structure. Define

$$e_t(x)(\omega) = \max_{y\in\Theta_t(\omega)} y \cdot x, \quad x \in \mathrm{R}^d. \tag{2.8}$$

Occasionally, we suppress the state and write simply $e_t(x)$. (In this notation, each of the expressions in (2.7) is equal to $e_t(\sigma_t)$.) It is well-known that, for each (t,ω), $e_t(\cdot)(\omega)$ provides a complete description of $\Theta_t(\omega)$ in that the latter can be recovered from $e_t(\cdot)(\omega)$. Characterizing properties of $e_t(\cdot)(\omega)$ include (Lipschitz) continuity, convexity, linear homogeneity and non-negativity (because of Normalization).[14] Further, by [2, Theorem 14.96], the above Measurability assumption is equivalent to:

$$(t,\omega) \longmapsto e_t(x)(\omega) \text{ is } \mathcal{B}([0,s]) \times \mathcal{F}_s\text{-measurable on } [0,s]\times\Omega \text{ for all } (s,x) \in (0,T]\times \mathrm{R}^d.$$

We use the support function primarily in the special case described in Section 3.4, where $e_t(\cdot)(\omega)$ is independent of both time and the state.

2.4. The Set of Priors

Given a set of density generators Θ, the corresponding set of priors is

$$\mathcal{P}^\Theta = \{\, Q^\theta : \theta \in \Theta \text{ and } Q^\theta \text{ is defined by (2.4)}\,\}.$$

The set of priors inherits the following properties, many of which are counterparts of those mentioned in Section 2.1 in the context of the atemporal multiple-priors model.

Theorem 2.2. *The set of priors* $\mathcal{P}^\Theta$ *satisfies:*

(a) $P \in \mathcal{P}^\Theta$.

(b) $\mathcal{P}^\Theta$ *is uniformly absolutely continuous with respect to* P *and each measure in* $\mathcal{P}^\Theta$ *is equivalent to* P.

(c) $\mathcal{P}^\Theta$ *is convex.*

(d) $\mathcal{P}^\Theta \subset ca^1_+(\Omega,\mathcal{F}_T)$ *is compact in the weak topology.*[15]

[14] The proof of Lipschitz continuity is contained in the proof of Theorem 2.3.

[15] Let $ba(\Omega,\mathcal{F}_T)$ denote the normed space of finitely additive real-valued functions on $\mathcal{F}_T$ with the total variation norm. The weak topology on $ba(\Omega,\mathcal{F}_T)$ is that induced by the set $B(\Omega,\mathcal{F}_T)$ of all bounded measurable real-valued functions. $ca^1_+(\Omega,\mathcal{F}_T)$ denotes the subset of (countably additive) probability measures; it inherits the above weak topology. Weak compactness is a stronger property than compactness in the induced weak*-topology.

(e) *For every* $\xi \in L^2(\Omega, \mathcal{F}_T, P)$, *there exists* $Q^* \in \mathcal{P}^\Theta$ *such that*

$$E_{Q^*}[\xi \mid \mathcal{F}_t] = \min_{Q \in \mathcal{P}^\Theta} E_Q[\xi \mid \mathcal{F}_t], \quad 0 \leq t \leq T.$$

(f) *For every deterministic* $\tau \in [0,T]$ *and every* $B \in \mathcal{F}_\tau$,

$$\mathcal{P}^\Theta = \left\{ Q(\cdot) = \int \left[Q^1(\cdot \mid \mathcal{F}_\tau) 1_B + Q^2(\cdot \mid \mathcal{F}_\tau) 1_{B^c} \right] dQ^3_\tau : \{Q^i\}_{i=1}^3 \subset \mathcal{P}^\Theta \right\}, \tag{2.9}$$

where Q^3_τ *denotes the restriction of* Q^3 *to* $\mathcal{F}_\tau$.

Parts (a)-(d) are self-explanatory. By (d), $\min_{Q \in \mathcal{P}^\Theta} E_Q \xi$ exists for any ξ in $L^1(\Omega, \mathcal{F}_T, P)$, *a fortiori* in $L^2(\Omega, \mathcal{F}_T, P)$. Part (e) extends the existence of a minimum to the process of conditional expectations.

Finally, consider (f), which is due to the rectangularity of Θ, as defined by (2.5). The direction '$\subset$' is trivial, because any Q in $\mathcal{P}^\Theta$ can be represented as an element of the set on the right by choosing each Q^i equal to Q and then applying the ordinary calculus for combining marginals and conditionals of a given measure. The essential content of (f) is the reverse direction '$\supset$', whereby suitable combinations of marginals and conditionals of *different* measures in $\mathcal{P}^\Theta$ yield another measure in $\mathcal{P}^\Theta$. In this sense, $\mathcal{P}^\Theta$ is suitably large. Part (f) can be strengthened in the obvious way to finitely many stopping times rather than the single deterministic time τ as stated.

It is natural to ask which properties characterize the sets of priors that can be generated via sets of density generators satisfying our regularity conditions. We provide only a partial response.

A related question is how the process θ underlying some measure Q can be derived from Q (reversing the direction of the argument surrounding (2.4)). The argument in [13, p. 289] may be adopted in response: Let Q be any measure equivalent to P. Denote by ξ its Radon-Nikodym density and define the martingale $z_t = E[\xi \mid \mathcal{F}_t]$. By the Martingale Representation Theorem, there exists an $\mathbb{R}^d$-valued process α_t such that $dz_t = \alpha_t \cdot dW_t$. Because Q is equivalent to P, (z_t) is a strictly positive process. Take $\theta_t = -\alpha_t / z_t$. Then Q is generated from θ as in (2.4). Expressed alternatively, the process θ underlying Q satisfies

$$d\left(dQ_t/dP\right) / \left(dQ_t/dP\right) = -\theta_t \cdot dW_t, \tag{2.10}$$

where Q_t is the restriction of Q to $\mathcal{F}_t$ and dQ_t/dP is its density with respect to P.

Consequently, if we begin with a set $\mathcal{P}$ of equivalent measures containing P, and define Θ as the set of all processes θ obtained in the way just described, then Θ is a candidate to satisfy $\mathcal{P} = \mathcal{P}^\Theta$. It remains only to examine whether Θ is consistent with the construction (2.5) and the preceding assumptions regarding the underlying primitive correspondences Θ_t. Because (2.10) permits one to recover Θ from $\mathcal{P}$, one could determine restrictions on $\mathcal{P}$ that would deliver the required properties for Θ. However, the restrictions obtained in this way, beyond those expressed in the Theorem, are not very illuminating and thus we do not describe them. We content ourselves with pointing out that if $\mathcal{P}$ satisfies the appropriate form of (2.9), then Θ satisfies the following condition related to the structure embodied in (2.5): For each τ, $B \in \mathcal{F}_\tau$ and $\{\theta^i\}_{i=1}^3 \subset \Theta$, then $(\theta_t) \in \Theta$, where

$$\theta_t \;=\; \left(1_{\{t\geq\tau\}\times B}\,\theta_t^1 \;+\; 1_{\{t\geq\tau\}\times B^c}\,\theta_t^2\right) \;+\; \theta_t^3\,1_{\{t<\tau\}}.$$

Finally, consider again the issue of learning. As suggested in the previous section, $\Theta_t(\omega)$ can be thought of as the set of conditional one-step-ahead densities (in logarithm) at (t,ω). Because this set depends on data (through ω), our general model permits learning. On the other hand, the responsiveness to data permitted by our model is very general and we do not yet have any compelling structure to add, for example, in order to illustrate the response of ambiguity to observation. Thus our principle examples below (Sections 3.3 and 3.4) exclude learning.

It may be useful to translate the preceding into the single-prior (and discrete-time) context. Typically, the prior is over the full state space and learning amounts to Bayes Rule. However, the Savage theory does not restrict this prior and its conditional one-step-ahead updates are similarly unrestricted. We adopt the equivalent approach of beginning with the updates and using them to construct the prior. In saying that we do not yet have an interesting structure to suggest for conditional one-step-ahead updates, we are simply acknowledging the widely recognized fact that there is no decision theory available that serves to pin down the prior.

2.5. Definition and Existence of Utility

Let Θ and $\mathcal{P}^\Theta$ be as above. In addition and in common with SDU (see (2.2)), another primitive component of the specification of utility is an aggregator $f : C \times \mathrm{R}^1 \longrightarrow \mathrm{R}^1$. Assume the following:

- f is Borel measurable.
- Uniform Lipschitz in utility: There exists a positive constant k such that

$$|f(c,v) - f(c,w)| \leq k\,|v-w|, \quad \text{for all } (c,v,w) \in C \times \mathrm{R}^2.$$

- Growth condition in consumption: $E\left[\int_0^T f^2(c_t,0)\,dt\right] < \infty$ for all $c \in D$.

We wish to generalize SDU by allowing the agent to employ the set $\mathcal{P}^\Theta$ of priors rather than the single measure P. On purely formal grounds, one is led to consider the following structure: Fix a consumption process c in D. Then for each measure Q in $\mathcal{P}^\Theta$, denote by (V_t^Q) the SDU utility process for c computed relative to beliefs given by Q, that is, (V_t^Q) is the unique solution (ensured by [14]) to

$$V_t^Q = E_Q\left[\int_t^T f(c_s, V_s^Q)\,ds \mid \mathcal{F}_t\right], \quad 0 \leq t \leq T. \tag{2.11}$$

The structure of the atemporal multiple-priors model suggests *defining* utility as the lower envelope

$$V_t = \min_{Q \in \mathcal{P}^\Theta} V_t^Q, \quad 0 \leq t \leq T. \tag{2.12}$$

We will show shortly that (2.13) admits a unique solution (V_t) for each c in D. Thus we can vary c and obtain the utility function $V_0(\cdot)$, or simply $V(\cdot)$ or V. When we wish to emphasize the underlying consumption process, we write $(V_t(c))$.

There are at least two important concerns regarding such a definition. First, it seems ad hoc. For example, on purely formal grounds, one might adopt instead of (2.12) the following alternative generalization of SDU:[16]

$$V_t = \min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^T f(c_s, V_s)\, ds \mid \mathcal{F}_t \right], \quad 0 \leq t \leq T. \tag{2.13}$$

In fact, we show shortly that this alternative yields an equivalent definition of utility. Section 3.2 demonstrates further that this definition delivers the natural extension of the atemporal multiple-priors model to a temporal setting.

A more practical concern regarding the above definition of utility is tractability, beginning with dynamic consistency of $(V_t(\cdot))$.[17] The next theorem shows that dynamic consistency is satisfied. This is done in the usual fashion, namely by showing that the utility process satisfies a recursive relation. Not surprisingly, the way to exploit fully the analytical power afforded by continuous-time (both in order to prove dynamic consistency and for subsequent analysis) is to express the recursive relation in differential terms. Accordingly, we show that the utility process defined by (2.12) can be characterized alternatively as the unique solution to a backward stochastic differential equation (BSDE).[18]

To illustrate, notice that the SDU process (V_t^P) defined by (2.2) can be expressed alternatively as the unique solution to the BSDE

$$dV_t^P = -f(c_t, V_t^P)\, dt + \sigma_t^P \cdot dW_t, \ \ V_T^P = 0. \tag{2.14}$$

In fact, because the volatility σ_t^P is endogenous and is part of the complete solution to the BSDE, it is more accurate to say that " (V_t^P, σ_t^P) is a (unique) solution". However, as our focus is on the utility component of the solution, we abbreviate and write " (V_t^P) is a (unique) solution"; similar abbreviated terminology is adopted for most other BSDE's that arise in the paper. To see that the BSDE characterization follows from (2.2), observe that, by the latter,

$$V_t^P + \int_0^t f(c_s, V_s^P)\, ds = E\left[\int_0^T f(c_s, V_s^P)\, ds \mid \mathcal{F}_t \right],$$

[16] Note that (2.12) can be restated as $V_t = \min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^T f(c_s, V_s^Q)\, ds \mid \mathcal{F}_t \right]$.

[17] Dynamic consistency, defined as in [14, p. 373], is the requirement that for all stopping times τ and all consumption processes c and c' satisfying $c' = c$ on $[0, \tau]$, $P(V_\tau(c') \geq V_\tau(c)) = 1 \implies V_0(c') \geq V_0(c)$, with strict inequality holding if $P(V_\tau(c') > V_\tau(c)) > 0$.

[18] See Appendix A for a brief outline and [18] for a comprehensive guide to the theory of BSDE's, as well as to previous applications to utility theory and derivative security pricing.

which is a martingale under P. Thus the Martingale Representation Theorem delivers (2.14) for a suitable process (σ_t^P) (that depends on c). This argument is readily reversed, using the fact that $\int_0^t \sigma_t^P \cdot dW_t$ is a martingale, to establish that (2.14) implies (2.2).

A similar reformulation is possible for the SDU process (V_t^Q) defined in (2.11) and corresponding to an agent with probabilistic beliefs given by Q in $\mathcal{P}^\Theta$. If $Q = Q^\theta$ (see (2.4)), then the Girsanov Theorem implies that (V_t^Q) solves the BSDE

$$dV_t^Q = \left[-f(c_t, V_t^Q) + \theta_t \cdot \sigma_t^Q\right] dt + \sigma_t^Q \cdot dW_t, \;\; V_T^Q = 0. \tag{2.15}$$

In comparison with (2.14), the drift is adjusted by the addition of $\theta_t \cdot \sigma_t^Q$ in order to account for the fact that (W_t) is not a Brownian motion under Q.[19]

We are now ready to state our main theorem.

Theorem 2.3. *Let Θ and f satisfy the preceding assumptions. Fix c in D. Then:*
(a) There exists a unique (continuous) process (V_t) solving the BSDE

$$dV_t = \left[-f(c_t, V_t) + \max_{\theta \in \Theta} \theta_t \cdot \sigma_t\right] dt + \sigma_t \cdot dW_t, \;\; V_T = 0. \tag{2.16}$$

(b) For each $Q = Q^\theta \in \mathcal{P}^\Theta$, denote by (V_t^Q) the unique solution to (2.11), or equivalently to (2.15). Then (V_t) defined in (a) is the unique solution to (2.12) and there exists $Q^{\theta^*} \in \mathcal{P}^\Theta$ such that

$$V_t = V_t^{Q^{\theta^*}}, \quad 0 \le t \le T. \tag{2.17}$$

(c) The process (V_t) is the unique solution to $V_T = 0$ and

$$V_t = \min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^\tau f(c_s, V_s)\, ds + V_\tau \mid \mathcal{F}_t \right], \;\; 0 \le t < \tau \le T. \tag{2.18}$$

Part (b) refers to the initial definition (2.12). Part (a) is the BSDE characterization that is the counterpart to (2.14). Setting $\tau = T$ in (2.18) delivers (2.13) and thus the promised equivalence between (2.12) and (2.13). More generally, (c) makes explicit the recursivity of utility and justifies the name *recursive multiple-priors utility* for our model of utility.

Comparison of (2.16) and (2.13) yields some insight into our construction. If the volatility of utility were denoted by $-\sigma_t$ rather than σ_t, then the essential supremum in (2.16) would be replaced by an essential infimum, paralleling (2.13). With this change of notation in mind, the integral and differential characterizations reveal an equivalence between the global minimization over $\mathcal{P}^\Theta$ and the continual instantaneous optimization over Θ.

This equivalence is due to our construction of Θ via (2.5) as rectangular. It is easy to understand the importance of (2.5). By (2.7), the maximum in (2.16) is equal to $\max_{y_t \in \Theta_t} y_t \cdot \sigma_t$, the solution of which at every t and ω in general permits the optimizer more freedom

[19] The Girsanov Theorem and the Martingale Representation Theorem are the key tools that we employ from stochastic calculus. They are standard in finance - see [13], for example. Section 4.2 explains further the role of the Girsanov Theorem in our model.

than does the global optimization problem in (2.12), where a single measure, or equivalently, a single θ, must be chosen at time 0. Thus if one begins with a general nonrectangular set Θ of density generators, local and global optimization would yield different results. There is equivalence here because (2.5) imposes that Θ is the Cartesian product of its projections.

Given a nonrectangular Θ, such as (2.6), then parts (a) and (b) of Theorem 2.3 would yield distinct definitions of utility. Adopting the BSDE definition in (a) would deliver recursivity and hence dynamic consistency, but not the multiple-priors form (2.12). Alternatively, choosing the latter by defining utility via (b) would violate recursivity. Though either option may be worth pursuing, we have chosen to construct a model having both properties and to opt for greater specificity rather than generality.

The coming sections illustrate, interpret and apply the recursive multiple-priors model of utility. First, we mention some extensions. The assumption of a Brownian filtration can be relaxed along the lines indicated in [18, Section 5.1]. The terminal value of 0 in (2.16) can be generalized and utility can be well-defined without the Lipschitz hypothesis [35]. Finally, we suspect that the extension from a finite horizon to an infinite horizon can be carried out in much the same way as it is done in [14] for stochastic differential utility. Related results for BSDE's defined on an infinite horizon may be found in [9] and [42].

Finally, we note that BSDE's have been used to price securities in markets that feature incompleteness, short-sale constraints or other imperfections.[20] These lead to nonlinear BSDE's characterizing (upper or lower) prices that are formally very similar to the BSDE (2.16) used here to define intertemporal utility. The similarity is suggested by the fact that, with imperfect markets, no-arbitrage delivers a nonsingleton set of equivalent martingale measures. In our setting, the multiplicity of measures arises at the level of utility and is due to ambiguity rather than features of the market.

3. EXAMPLES

3.1. Deterministic and Risky Consumption Processes

It is important to keep in mind that σ_t is endogenous in the BSDE (2.16). To illustrate this endogeneity and the consequent dependence of σ_t on the consumption process, consider (2.16) for two particular consumption processes. First, suppose that c is deterministic. Then $\sigma_t = 0$ and utility is given by the ordinary differential equation

$$dV_t \;=\; -f(c_t, V_t)\,dt, \quad V_T = 0.$$

This is the model of recursive utility for deterministic consumption processes proposed in [20].

For the second example, let

$$R \;=\; \{i : 1 \leq i \leq d,\, (\theta_t^i) = 0 \;\text{ for all } \theta \text{ in } \Theta\} \;\text{ and} \tag{3.1}$$

[20] See [18] for some references.

$$\mathcal{F}_t^R = \sigma\left(W_s^i : i \in R, \, s \le t\right). \tag{3.2}$$

Then all measures in $\mathcal{P}^\Theta$ agree with P for events that are $\mathcal{F}_T^R$-measurable and it is natural to view such events as unambiguous or purely risky. We elaborate upon this interpretation in Section 4.2. Here we wish merely to clarify the mechanics of the BSDE (2.16). Accordingly, let c be adapted to the filtration $\{\mathcal{F}_t^R\}$. Then $\sigma_t^i = 0$ for $i \notin R$ and $\max_\theta \theta_t \cdot \sigma_t = 0$, implying that the SDU utility process V_t^P defined in (2.14) is the solution to (2.16). That is because the consumption process c just described is viewed by the consumer as being purely risky.

3.2. Standard Aggregator

The aggregator underlying the expected additive utility model (1.2) is

$$f(c, v) = u(c) - \beta v, \quad \beta \ge 0. \tag{3.3}$$

For this aggregator, there exists a closed-form representation for recursive multiple-priors utility, as we now show (assuming the appropriate measurability for u). By Theorem 2.3(b), it is enough to have a representation for V_t^Q for each Q in $\mathcal{P}^\Theta$. However, from (2.11),

$$V_t^Q = E_Q\left[\int_t^T e^{-\beta(s-t)} u(c_s)\, ds \mid \mathcal{F}_t\right].$$

Conclude that

$$V_t = \min_{Q \in \mathcal{P}^\Theta} E_Q\left[\int_t^T e^{-\beta(s-t)} u(c_s)\, ds \mid \mathcal{F}_t\right], \tag{3.4}$$

which is the desired closed-form expression.

Two polar opposite reactions to this functional form seem possible. Some readers may find it to be the 'obvious' way to formulate a multiple-priors extension of the usual model (1.2), perhaps with $\mathcal{P}^\Theta$ generalized to some set of priors $\mathcal{P}$, and therefore may wonder why this paper is so long and complicated. Other readers may feel that the utility process delivered by (3.4) will be of limited use because it violates dynamic consistency. Neither view is accurate for reasons given in the discussion of the equivalence of parts (a) and (b) of Theorem 2.3. It may be useful to elaborate slightly on that discussion in the context of the specific functional form available here.

While (3.4) is a natural guess as to what might work, dynamic consistency is a concern. In the usual model, one can apply additivity of the expectations operator $E[\cdot]$ to deduce the recursive relation

$$V_t = E\left[\int_t^\tau e^{-\beta(s-t)} u(c_s)\, ds + e^{-\beta(\tau-t)} V_\tau \mid \mathcal{F}_t\right], \; t \le \tau.$$

Dynamic consistency follows immediately. In contrast, the operation of taking an infimum (over a set of measures) of expected values is not additive, leading one to suspect that dynamic consistency may be violated by (3.4). In fact, the corresponding recursive relation

$$V_t = \min_{Q \in \mathcal{P}^\Theta} E_Q\left[\int_t^\tau e^{-\beta(s-t)} u(c_s)\, ds + e^{-\beta(\tau-t)} V_\tau \mid \mathcal{F}_t\right], \; t \le \tau,$$

is valid (by Theorem 2.3(c)) and this is sufficient for dynamic consistency. However, this recursivity is due to the specific sets of priors $\mathcal{P}^\Theta$ that we have defined; it is not true more generally.

Finally, the explicit expression (3.4) for utility clarifies the connection between our model of intertemporal utility and the atemporal multiple-priors model. First, because intertemporal utility equals a lower envelope of a collection of expected utility functions, there is a more direct parallel to the atemporal model (2.1) than is possible for general aggregators. At the more important level of behavioral content, consider consumption processes c of the form

$$c_\tau = \begin{cases} \overline{x} & \text{if } \tau < t \\ \xi & \text{if } \tau \geq t, \end{cases} \tag{3.5}$$

where $\overline{x}$ is the deterministic and constant consumption level on $[0, t]$ and $\xi \in L^2(\Omega, \mathcal{F}_t, P)$ describes the constant consumption level on the interval $[t, T]$. If c' is another such process, corresponding to the same t and $\overline{x}$ but different ξ', then

$$V_0(c') \geq V_0(c) \iff \min_{Q \in \mathcal{P}^\Theta} E_Q\, u(\xi') \geq \min_{Q \in \mathcal{P}^\Theta} E_Q\, u(\xi). \tag{3.6}$$

Thus the behavioral properties of the atemporal multiple-priors model apply also to the ranking of such pairs of consumption processes.[21] A similar statement is valid for a general aggregator f, though the utility index u must be replaced by a suitable function derived from f.[22]

The remaining examples are concerned primarily with illustrative specifications for Θ.

3.3. κ-Ignorance

Fix a parameter $\kappa = (\kappa_1, ..., \kappa_d)$ in R^d_+ and take

$$\Theta_t(\cdot) = \{y \in R^d : \mid y_i \mid \leq \kappa_i \text{ for all } i\}.$$

Then

$$\Theta = \left\{ (\theta_t) : \sup \{\mid \theta^i_t \mid : \ 0 \leq t \leq T\} \leq \kappa_i, \ i = 1, ..., d \right\}.$$

The following notation will be useful. Denote by $\mid \sigma_t \mid$ the d-dimensional vector with i^{th} component $\mid \sigma^i_t \mid$; similarly for other d-dimensional vectors. Use the notation

$$sgn(x) \equiv \begin{cases} \mid x \mid / x & \text{if } x \neq 0 \\ 0 & \text{otherwise}, \end{cases} \tag{3.7}$$

and $\kappa \otimes sgn(\sigma_t) \equiv \left(\kappa_1 sgn(\sigma^1_t), ..., \kappa_d sgn(\sigma^d_t)\right)$.

[21] See [7] for an exhaustive set of behavioral implications, that is, for an axiomatization. The axiomatization in [28] is formulated in the context of the Anscombe-Aumann domain of two-stage acts and thus is not readily translated to our setting.

[22] With t and $\overline{x}$ fixed, define $U(x) \equiv \varphi(0)$ where $\varphi(T) = 0$ and $d\varphi(s)/ds = -f(x, \varphi(s))$ on $[t, T]$ and $= -f(\overline{x}, \varphi(s))$ on $[0, t]$. Then (3.6) is valid if u is replaced by U.

Then

$$\max_{\theta \in \Theta} \theta_t \cdot \sigma_t \;=\; \theta_t^* \cdot \sigma_t \;=\; \kappa \cdot \mid \sigma_t \mid \text{, where}$$

$\theta_t^* \;=\; \kappa \otimes sgn(\sigma_t)$, that is,

$$\theta_t^{*i} = \begin{cases} \kappa_i \mid \sigma_t^i \mid / \sigma_t^i & \text{if } \sigma_t^i \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

Consequently, the utility process solves

$$dV_t \;=\; [-f(c_t, V_t) \;+\; \kappa \cdot \mid \sigma_t \mid] \, dt \;+\; \sigma_t \cdot dW_t, \quad V_T = 0. \tag{3.9}$$

Though it is customary to think of a volatility such as σ_t as tied to risk, the above BSDE cannot be delivered within the risk framework of [14]. We interpret the term $\kappa \cdot \mid \sigma_t \mid$ as modeling ambiguity aversion rather than risk aversion (see Section 4.3). The connection between κ and ambiguity aversion is illustrated informally by referring back to the Ellsberg-type rankings in (1.1). They can be accommodated within the present model by taking $\kappa_1 \;=\; 0$ and $\kappa_2 \;>\; 0$.

3.4. IID Ambiguity

For a generalization of κ-ignorance, let $K \subset \mathrm{R}^d$ be a compact and convex set containing the origin and define

$$\Theta_t(\cdot) \;=\; K \quad \text{for all } t.$$

Recalling the interpretation of $\Theta_t(\omega)$ as the set of one-step-ahead conditionals, the constancy of this set indicates the lack of learning from data. As described in the introduction, there are situations in which some features of the environment remain ambiguous even asymptotically. The current specification models the agent after he has learned all that he can. The label 'IID ambiguity' is natural given the analogy with the case of a single-prior that induces one-step-ahead conditionals that are constant across time and states; further support is given in Section 4.4.

The utility process generated by this specification for $\{\Theta_t\}$ solves

$$dV_t \;=\; [-f(c_t, V_t) \;+\; e(\sigma_t)] \, dt \;+\; \sigma_t \cdot dW_t, \;\; V_T = 0,$$

where $e(\cdot)$ is the support function for K defined by

$$e(x) \;= \max_{y \in K} y \cdot x, \quad x \in \mathrm{R}^d, \tag{3.10}$$

corresponding to a special case of (2.8) where the support function is independent of both time and the state.

By the theory of support functions [46], the process θ^* asserted by Theorem 2.3(b) is given by

$$\theta_t^* \;\in\; \partial e(\sigma_t) \;\; \text{for every } \; t, \tag{3.11}$$

where $\partial e(x)$ denotes the set of subgradients of e at x. This relation generalizes (3.8).

Denote by K^i the projection of K onto the i^{th} co-ordinate direction and let d_1 denote both $\{1 \leq i \leq d : K^i \neq \{0\}\}$ and its cardinality. We can decompose K into a product $\{0_{d-d_1}\} \times K_1$, where $K_1 \subset \mathrm{R}^{d_1}$. It will be convenient to add the assumption that

$$0_{d_1} \in int\,(K_1) \subset \mathrm{R}^{d_1}. \tag{3.12}$$

In words, for those processes $(W_t^i)_{i \in d_1}$ for which there is not complete confidence that P describes the underlying distribution, then P is not on the 'boundary' of the set of alternative conceivable measures. The corresponding property of e is that, for all $x \in \mathrm{R}^d$,[23]

$$e(x) = 0 \implies x_i = 0 \text{ for all } i \text{ such that } K_i \neq \{0\}. \tag{3.13}$$

The assumption (3.12) is included in any reference below to IID ambiguity.

The special case $K = \{y \in \mathrm{R}^d : \mid y_i \mid \leq \kappa_i \text{ all } i\}$ delivers κ-ignorance. An alternative special case has

$$K = \{y \in \mathrm{R}^d : \Sigma_{\{i: \kappa_i \neq 0\}}\, \kappa_i^{-1} \mid y_i \mid^2 \leq 1\}, \tag{3.14}$$

where $\kappa = (\kappa_1, ..., \kappa_d) \geq 0$, leading to $e(x) = (\kappa \cdot x^2)^{1/2}$; x^2 denotes the d-dimensional vector with i^{th} component x_i^2.

By restricting the aggregator f, we can compute utility explicitly for consumption processes of the form

$$dc_t / c_t = \mu^c\, dt + s^c\, dW_t, \tag{3.15}$$

where μ^c and s^c are constant. Suppose the aggregator is given by

$$f(c, v) = \frac{c^\rho - \beta(\alpha v)^{\rho/\alpha}}{\rho\,(\alpha v)^{(\rho-\alpha)/\alpha}}, \tag{3.16}$$

for some $\beta \geq 0$ and nonzero ρ, $\alpha \leq 1$.[24] This is the continuous-time version of the so-called Kreps-Porteus functional form [14, p. 367]. It is attractive because the degree of intertemporal substitution and risk aversion are modeled by the separate parameters ρ and α respectively. The homothetic version of the standard aggregator (3.3), with $u(c) = c^\alpha / \alpha$, is obtained when $\alpha = \rho$.

The corresponding utility process can be computed explicitly by verifying the trial solution

$$V_t = A_t\, c_t^\alpha / \alpha\ ,$$

where

$$A_t^{\rho/\alpha} = \lambda^{-1}\left(1 - e^{\lambda(t-T)}\right),$$

[23] $e(x) = 0$ iff $y \cdot x \leq 0$ for all $y \in K$. Suppose there exists i such that $K^i \neq \{0\}$. (Otherwise, (3.13) is obvious.) Then it follows from (3.12) that $x_i = 0$. (The reverse implication in (3.13) is evidently also true.) Alternatively, (3.12) is equivalent to the assumption that the polar of K is $\{0\}$.

[24] This aggregator violates the Lipschitz condition for Theorem 2.3 and thus existence of utility is not ensured. See further discussion in Section 5.2.

$$\lambda = \beta - \rho \left(\mu^c - (1-\alpha) s^c \cdot s^c / 2 - e(s^c) \right).$$

The associated volatility is

$$\sigma_t = A_t c_t^{\alpha} s^c. \tag{3.17}$$

Evidently the utility of the given consumption process is increasing in initial consumption and in $(\mu^c - (1-\alpha)s^c \cdot s^c/2 - e(s^c))$, the mean growth rate adjusted both for risk (via the second term) and ambiguity (via the third term). Support for the latter interpretation will follow in Section 4.3 from the interpretation provided there for α and $e(\cdot)$. Observe that the risk premium is quadratic in the consumption volatility s^c, whereas the ambiguity premium is linearly homogeneous in s^c. The ambiguity premium is $\kappa \cdot \mid s^c \mid$ in the case of κ-ignorance.

For further interpretation of $e(\cdot)$ see Lemma A.2.

4. PROPERTIES of UTILITY

Under suitable assumptions, the utility function we have defined has a number of classical properties, such as monotonicity, concavity and continuity. They can be proven as in [14] or [18, Proposition 3.5]. As noted prior to Theorem 2.3, dynamic consistency is an immediate consequence of the recursive construction of utility via (2.16).[25]

In the sequel, we focus primarily on properties of preference related to ambiguity.

4.1. Behaviorally Distinct from SDU

We have referred to our model of utility as accommodating ambiguity aversion and as distinct from SDU and all other models of (continuous-time) intertemporal utility. Here and in the next two sections we examine these claims more carefully.

Consider the comparison with SDU. Though the defining BSDE's (2.14) and (2.16) appear different, how do we know that the latter could not be generated by the SDU model? The argument to the contrary would be that the utility process V_t, or at least some monotonic transform $\overline{V}_t = \phi(V_t)$, could alternatively be generated by some aggregator g different from f, and by the singleton set $\Theta = \{0\}$ of density generators. In that case, utility would be an instance of the SDU model, in which only risk matters, refuting the suggestion that ambiguity aversion is necessarily introduced via a nonsingleton Θ. However, it is easily seen via Ito's Lemma that if ϕ is twice continuously differentiable, then $\overline{V}_t$ does not solve a BSDE of the form (2.16). A (smooth) monotonic transformation does not introduce into the drift a term that is linearly homogeneous in σ_t, which is a noteworthy and distinguishing feature of (2.16).[26]

An alternative and more satisfactory comparison would demonstrate that recursive multiple-priors utility and SDU are behaviorally distinct. The Ellsberg-type rankings in (1.1) illustrate

[25] More precisely, it follows from the Comparison Theorem A.1 for BSDE's stated in Appendix A.

[26] See [14, Section 3.3] for the relation between a smooth monotonic transformation of utility and the generating BSDE. For utility transformations that are not smooth but are convex, the corresponding generalization of Ito's Lemma involves the notion of "local time", and thus also fails to duplicate (2.16).

such a distinction; these rankings are consistent with some multiple-priors utility functions (see Section 3.3, for example) but not with any SDU functions. We would like to generalize from this example and to broaden the comparison to include utility functions other than SDU.

A central feature of *all* continuous-time intertemporal utility functions in the current literature is that they are 'based on probabilities.' To give precise meaning to the latter term, adopt a variation of probabilistic sophistication as defined by Machina and Schmeidler. Denote by $D_\tau \subset D$ the set of consumption processes c such that (i) c_t is deterministic for $0 \leq \tau < t$ and (ii) c_τ is $\mathcal{F}_t$-measurable for each $t \leq \tau \leq T$. Processes in D_t are such that all uncertainty is resolved at the single instant t and thus we refer to elements in $\cup_{t=0}^{T} D_t$ as *timeless prospects.*[27] Call the utility function $V : D \longrightarrow \mathrm{R}^1$ *probabilistically sophisticated for timeless prospects* if, for every $0 \leq t \leq T$, V restricted to D_t is probabilistically sophisticated in the sense of Machina and Schmeidler.

When applied to the utility function V on D, the Machina and Schmeidler notion requires 'primarily' that there exists a probability measure Q on $(\Omega, \mathcal{F}_T)$ such that the utility of any c depends only on the probability distribution induced by $c : \Omega \rightarrow \mathrm{R}^{\ell[0,T]}$ and Q. However, it imposes also, through their adoption of the Savage axiom $P3$ or the associated property of monotonicity with respect to 'first-order stochastic dominance', restrictions on intertemporal aspects of preference that have nothing to do with probabilities. For example, SDU is probabilistically sophisticated in their sense only in the special case of the standard intertemporally additive expected utility function. The restriction to timeless prospects is intended exactly to exclude such extraneous restrictions and to isolate the property that preference is based on probabilities.

Our adoption of a weaker notion of probabilistic sophistication makes the points to follow stronger. Also, the axiomatization in [37] may be adapted to deliver an axiomatization of our modified notion. Thus probabilistic sophistication for timeless prospects is a meaningful behavioral notion. Finally, it is satisfied by *all* existing models of continuous-time utility, but typically not by the multiple-priors model, as we now show.

Theorem 4.1. *Suppose that the recursive multiple-priors utility function V defined in Theorem 2.3 is probabilistically sophisticated for timeless prospects. Suppose also that*

$$\theta_t^1 = 0 \quad \text{for all } \theta \text{ in } \Theta; \tag{4.1}$$

that is, all measures in $\mathcal{P}^\Theta$ agree with P on events in $\mathcal{F}_T^1$. Then $\Theta = \{0\}$ and thus V conforms to the SDU model (2.11).

We suspect, but have not managed to prove, that the assumption (4.1) can be dropped. Modulo this assumption, the theorem shows that probabilistic sophistication for timeless prospects distinguishes previous models behaviorally from recursive multiple-priors utility.

[27]The process defined in (3.5) is one example.

4.2. Ambiguity

Begin with some informal remarks about the nature of the ambiguity modeled via recursive multiple-priors utility. For any given Q in $\mathcal{P}^{\Theta}$, we know from Girsanov's Theorem applied to (2.15) that V_t^Q solves

$$dV_t^Q = -f(c_t, V_t^Q)\, dt + \sigma_t \cdot dW_t^Q, \;\; V_T^Q = 0, \tag{4.2}$$

where W_t^Q is a Brownian motion under Q. Thus it is as though the consumer is certain how to compute utility, namely by solving a BSDE having the specific form indicated, but is not certain which driving process W_t^Q is correct. This uncertainty is evaluated through a minimization over all 'possible' processes W_t^Q as Q varies over $\mathcal{P}^{\Theta}$. The primitive process $W_t = W_t^P$ is but one of these. This interpretation can be pushed further because (by Girsanov's Theorem) if $Q = Q^{\theta}$, then $W_t^Q = W_t + \int_0^t \theta_s ds$ or

$$dW_t^Q - dW_t = \theta_t\, dt. \tag{4.3}$$

Thus ambiguity concerns (and is limited to) the drift of the driving process. The fact that ambiguity is limited to the drift may seem overly restrictive, but it is a consequence of the Brownian environment and the assumption of absolute continuity.

Alternatively, one could think in terms of a fixed driving process (W_t) and a multiplicity of measures that determine alternative distributions for (W_t). Only one of these measures, namely P, makes the driving process a Brownian motion. Thus there is ambiguity about whether (W_t) is a Brownian motion.

We attempt now to treat ambiguity more formally. At a formal level, as argued in [19] and [25], ambiguity (or unambiguity) is most naturally defined as a property of events (elements of $\mathcal{F}_T$). Consequently, ambiguity about whether (W_t) is a Brownian motion would be expressed in terms of ambiguity of events defined by the W_t's. Further, once the class $\mathcal{U} \subset \mathcal{F}_T$ of unambiguous events is given, consumption processes that are adapted to $\{\mathcal{U} \cap \mathcal{F}_t\}$ are naturally viewed as unambiguous; remaining processes are ambiguous. The restriction of utility V to unambiguous consumption processes embodies attitudes towards risk. The decision-maker's attitude towards ambiguity, on the other hand, is reflected in the way in which ambiguous processes are ranked relative to unambiguous ones (in a sense to be made precise). In this way a conceptual distinction can be achieved between attitudes towards risk and towards ambiguity. (See the next section.)

It remains to specify $\mathcal{U}$. It is tempting to think of unambiguous events as those events where all measures in $\mathcal{P}^{\Theta}$ agree, that is,

$$\mathcal{U} = \{B \in \mathcal{F}_T : \; Q(B) = P(B) \text{ for all } Q \text{ in } \mathcal{P}^{\Theta}\}. \tag{4.4}$$

All other events would be called ambiguous.

While seemingly intuitive, (4.4) falls short as a *definition* of unambiguous events. Indeed, because (4.4) amounts to a restriction on the functional form for utility, its interpretation and intuitive appeal are unclear unless this restriction can be related to behavior. As argued in Section 1.4, to be satisfactory, a definition of unambiguous events must be expressed in terms of observable (at least in principle) choice behavior.

Fortunately, there exist two behavioral or preference-based definitions in the literature and we can check whether they deliver (4.4) as the *characterization* of unambiguous events for the recursive multiple-priors utility function. The answer is immediately 'yes' for the definition proposed by Ghirardato and Marinacci [27]. Our (subjectively) preferred definition is the one proposed by Epstein and Zhang. Though we have not yet succeeded in constructing a proof, we strongly suspect that (4.4) is valid also for this definition.[28] We proceed on the basis of the evidence and suspicions just described to accept (4.4) as the designation of the class of unambiguous events.

It is convenient to have a further characterization in terms of the primitive set of density generators. To express it, let

$$\Theta^i = \{\theta^i = (\theta^i_t) : \theta \in \Theta\}, \quad i = 1, ..., d.$$

Denote by $\{\mathcal{F}^i_t\}$ the filtration generated by the i^{th} driving process (W^i_t). For IID ambiguity, we can prove:

Lemma 4.2. *Let Θ correspond to IID ambiguity. Then for any $F \in \mathcal{F}_T$, all measures in $\mathcal{P}^\Theta$ agree on F (that is, $Q(F) = P(F)$ for all Q in $\mathcal{P}^\Theta$) if and only if: For each i,*

$$\Theta^i = \{0\} \text{ or } P(F \mid \mathcal{F}^i_T) = 0 \text{ or } 1. \tag{4.5}$$

Consequently, we obtain

$$\mathcal{U} = \{F \in \mathcal{F}_T : \text{ for each } i, \Theta^i = \{0\} \text{ or } P(F \mid \mathcal{F}^i_T) = 0 \text{ or } 1\}.$$

To illustrate, in the κ-ignorance model let $\kappa_1 = 0$ and $\kappa_i > 0$ for $i > 1$. Then events that are determined by the first driving process (W^1_t) are unambiguous; all other events are ambiguous.

4.3. Risk and Ambiguity Aversion

Given the preceding designation of unambiguous events, we can proceed as outlined in the previous section. The approach advocated in [19] and [25] is adopted to define the distinct notions of ambiguity aversion and risk aversion. Further details and an expanded discussion may be found in these papers.

First, define c to be an *unambiguous consumption process* if c_t is $\mathcal{U}$-measurable for each $t < T$. When it is important to make explicit the underlying utility function, refer to c as V-unambiguous.

[28] If $\mathcal{U}$ is defined as in [25], then '$\supset$' in (4.4) is easily verified. For the converse, we have a proof (available on request) that, under κ-ignorance and for a 'large class' $\mathcal{B}$ of events, if B in $\mathcal{B}$ is unambiguous, then all measures in $\mathcal{P}^\Theta$ agree on B. The class $\mathcal{B}$ includes all events B of the form $B = \{\omega \in \Omega : \varphi(W_{t_1}, ..., W_{t_\ell}) \geq a\}$ where $a \in \mathrm{R}^m$ and $\varphi : \mathrm{R}^\ell \longrightarrow \mathrm{R}^m$ is continuous and strictly increasing.

Given utility functions V and V^* with corresponding classes $\mathcal{U}$ and $\mathcal{U}^*$ of unambiguous events, say that V^* is *more ambiguity averse* than V if both

$$\mathcal{U} \supset \mathcal{U}^* \text{ and} \tag{4.6}$$

$$V(c^{ua}) \geq (>)\, V(c) \implies V^*(c^{ua}) \geq (>)\, V^*(c), \tag{4.7}$$

for all consumption processes c and c^{ua}, the latter V^*-unambiguous. The interpretation is that if V prefers the V^*-unambiguous process c^{ua}, which is also unambiguous for V, then so should the more ambiguity averse V^*. The nesting condition (4.6) ensures that the more ambiguity averse decision-maker views more events as ambiguous. Of course, $\mathcal{U}$ and $\mathcal{U}^*$ may coincide.

Say that V^* is *more risk averse* than V if both

$$\mathcal{U} \subset \mathcal{U}^* \text{ and} \tag{4.8}$$

$$V(\overline{c}) \geq (>)\, V(c^{ua}) \implies V^*(\overline{c}) \geq (>)\, V^*(c^{ua}), \tag{4.9}$$

for all V-unambiguous consumption processes c^{ua} and deterministic processes $\overline{c}$. Symmetric with the prior definition, the more (risk) averse agent is assumed, via (4.8), to perceive more risk. Implicit is the presumption that 'unambiguous' and 'risky' are synonymous and thus that unambiguous consumption processes constitute the appropriate subdomain for exploring risk attitudes. For comparative purposes, 'unambiguous' must apply to both utility functions and hence mean 'V-unambiguous'. Finally, the intuition for the definition is that any risky process that is disliked by V relative to a riskless $\overline{c}$, should be disliked also by the more risk averse V^*.[29]

Consider an extreme case where

$$\mathcal{U}^* = \{\emptyset, \Omega\},$$

that is, all nontrivial events are ambiguous according to V^*. Then V^* is more risk averse than V if and only if $\mathcal{U} = \{\emptyset, \Omega\}$ and V^* and V agree in the ranking of deterministic processes. This may seem odd at first glance, but is a natural consequence of the fact that there is no risk according to either agent. Accordingly, differences in the 'certainty equivalents' assigned to any consumption process by V and V^* are attributed entirely to differences in ambiguity aversion. In particular, in this case V^* is more ambiguity averse than V if and only if

$$V(\overline{c}) \geq (>)\, V(c) \implies V^*(\overline{c}) \geq (>)\, V^*(c),$$

for all c and $\overline{c}$, the latter deterministic. Similarly, ambiguity aversion is uninteresting at the other extreme where $\mathcal{U}^* = \mathcal{F}_T$, that is, there is no ambiguity.

The definitions are best clarified by application to canonical functional forms - Kreps-Porteus aggregators (3.16) and κ-ignorance for Θ. The utility function V generated by any

[29] Absolute notions of ambiguity and risk aversion could be formulated as well, but they are less important and are thus omitted in the interest of brevity.

such pair (f, Θ) can be identified with a quartet of parameters $(\beta, \rho, \alpha, \kappa)$. The temptation, to which we have yielded above, is to interpret β and ρ as describing time preference and willingness to substitute intertemporally given deterministic processes, α as a risk aversion parameter and to view κ as modeling ambiguity aversion. Partial support is provided by the facts that the ranking of deterministic processes uniquely determines β and ρ and that it is unaffected by α and κ. Additional support for the above interpretations is described next.[30]

Theorem 4.3. *(i)* $(\beta^*, \rho^*, \alpha^*, \kappa^*)$ *is more ambiguity averse than* $(\beta, \rho, \alpha, \kappa)$ *if* $(\beta^*, \rho^*, \alpha^*) = (\beta, \rho, \alpha)$ *and* $\kappa^* \geq \kappa$. *The converse is true if* $\kappa_i^* = 0$ *for some* i.
(ii) $(\beta^*, \rho^*, \alpha^*, \kappa^*)$ *is more risk averse than* $(\beta, \rho, \alpha, \kappa)$ *if* $(\beta^*, \rho^*) = (\beta, \rho)$, $\alpha^* \leq \alpha$ *and for each* i, $\kappa_i = 0$ *implies* $\kappa_i^* = 0$. *The converse is true if* $\kappa_i = 0$ *for some* i.

Ambiguity aversion alone is increased by increasing the ignorance parameter, while risk aversion alone is increased by reducing α. In this comparative sense these two aspects of preference are modeled by separate parameters, and separately from properties of the ranking of deterministic processes.[31]

The above theorem generalizes in a straightforward way to general aggregators and IID ambiguity.

Theorem 4.4. *Consider aggregators* f *and* f^* *and let* Θ *and* Θ^* *correspond to IID ambiguity with corresponding sets* K *and* K^* *as described in Section 3.4. Then:*
(i) (f^*, K^*) *is more ambiguity averse than* (f, K) *if*

$$f = f^* \text{ and } K^* \supset K. \tag{4.10}$$

The converse is true if $K^{*i} = \{0\}$ *for some* i.
(ii) (f^*, K^*) *is more risk averse than* (f, K) *if*

$$f^*(c, h(v)) = h'(v)\, f(c, v), \tag{4.11}$$

for some transformation h *with* $h' > 0$ *and* $h'' \leq 0$; *and*
for each i, $K^i = \{0\}$ *implies* $K^{*i} = \{0\}$.

The proof is similar to that of the preceding theorem, with reliance also on [14, Sections 3.3, 5.6] in order to deal with (4.11). We refer the reader to the just cited paper for clarification of (4.11) and for an alternative to (4.11) that is more intuitive (but too involved to include here). We merely note that the transformation in (4.11) implies that $V^* = h(V)$ when restricted to deterministic consumption processes. Thus they rank such processes identically, which is a necessary condition for their risk attitudes to be comparable.

[30] We continue to ignore the existence and uniqueness issues for the Kreps-Porteus aggregator.

[31] It is not possible to change the two forms of aversion simultaneously, because the change from α to α^* makes ambiguity attitudes noncomparable. This parallels the inability within the Kreps-Porteus functional form to change simultaneously the elasticity of intertemporal substitution $(1-\rho)^{-1}$ and the degree of risk aversion; the change from ρ to ρ^* makes risk attitudes noncomparable.

The converse in (ii) is not true in general because of the presumption in (4.11) that the function h relating V^* and V is twice differentiable. However, if we restrict attention to this case and if $K^i = \{0\}$ for some i, implying that there is some nontrivial risk common to both utility functions, then V^* more risk averse than V does imply the conditions stated in the theorem.

We use the preceding theorems to justify the interpretation of various expressions as capturing the effects of risk aversion or of ambiguity aversion. An example is the lognormal consumption process described in Section 3.4, where we suggested that $(1-\alpha)s^c \cdot s^c/2$ represents a premium for the riskiness of c and that $e(s^c)$ represents a premium for its ambiguity. A later example is a decomposition of the equity premium (5.19).

4.4. Aspects of the Temporal Structure of Preference

While any specification for Θ models ambiguity, it is desirable to understand precisely what it is about the environment that is perceived as ambiguous or unambiguous. This section provides an answer for Θ's conforming to IID ambiguity. A rationale for the latter term is delivered as a by-product.

We begin with the following elementary observation: For each r in $[0,T]$,

$$c'_t = c_t \text{ on } [r,T]\times\Omega \implies V_t(c') = V_t(c) \text{ on } [r,T]\times\Omega, \tag{4.12}$$

confirming that $V_r(c)$ depends only on the continuation of c beyond r, that is, on ${}^r c = (c_t)_{r\leq t\leq T}$. In particular, ambiguity does not introduce dependence on the past.

When we wish to vary the length T of the horizon and want to make explicit the particular horizon being discussed, we write $V_0^T(\cdot)$ for the utility function defined as in Theorem 2.3.

Theorem 4.5. *Suppose that Θ is given by IID ambiguity (Section 3.4). For each r in $[0,T]$, let $\mathcal{G}_t^r = \sigma(W_s - W_r : t \geq s \geq r)$ for $t \geq r$ and $= \{\emptyset, \Omega\}$ otherwise.*

(a) If c is adapted to the filtration $\{\mathcal{G}_t^r\}$, then $(V_t(c))$ is deterministic for $t \leq r$.

(b) If $c' = c$ on $[0,r)\times\Omega$, and both processes are adapted to $\{\mathcal{G}_t^r\}$, then

$$V_0^T(c') \geq V_0^T(c) \iff V_0^{T-r}({}^r c') \geq V_0^{T-r}({}^r c). \tag{4.13}$$

(c) If c_t is $\sigma(W_t)$-measurable for each t in $[0,T]$, then so is $V_t(c)$.

For interpretation, consider each of these statements when $\Theta = \{0\}$ and thus beliefs are represented by the single prior P. Suppose as in (a) that consumption is deterministic until time r and thereafter depends only on increments $W_s - W_r$. Because (W_t) is Brownian motion relative to P, such increments are independent of $\mathcal{F}_t$ for any $t < r$. Thus the time t conditional utility $V_t(c)$ is deterministic until r. We are led to interpret (a) as expressing a form of independence in beliefs about future increments even when Θ is a nonsingleton (but conforms to the IID specification).

Part (b) implies that calendar time matters only because it implies a different length for the remaining horizon. Under P, this is due to the stationarity of Brownian motion (the unconditional distribution of $W_s - W_r$ is identical to that of W_{s-r}). Accordingly, we interpret (b) as expressing a form of stationarity in beliefs even given IID ambiguity.

Finally, with regard to (c), let $\tau > t$. Then the Markov property of Brownian motion implies that, under P, time t conditional beliefs about W_τ and hence also conditional utility at t depend only on W_t. Part (c) asserts that this Markov-type property is preserved under IID ambiguity.

It merits emphasis that each of the properties in the theorem has *behavioral significance*. The latter is explicit for (b). Given two consumption processes c' and c as in part (c), their conditional ranking at t depends on time t information only via W_t. Similarly, for the significance of (a).

In summary, the Theorem gives precise meaning to the statement that an IID specification for Θ models the decision-maker as viewing the driving process (W_t) as "Markov with independent and identical increments". This view is modeled also by the standard model with singleton prior P, in that (a)-(c) are valid also then. The primary difference is that in the standard model but not under IID ambiguity, the decision-maker views increments as being normally distributed with the familiar means and variances. Complete confidence in this aspect of the environment is relaxed via IID ambiguity.

Incomplete confidence in the other aspects noted above can be modeled by the more general specifications for Θ that are permitted by (2.5), that is, by relaxing the assumption made in the IID model that the correspondences Θ_t are constant. We do not elaborate but the general point merits emphasis - our model permits determination of the behavioral meaning of alternative specifications for Θ.

The theorem may be illustrated as follows: Take IID ambiguity, the simple linear aggregator with zero discounting ($f(c,v) = c$) and the consumption process

$$c_t = \begin{cases} a \cdot (W_s - W_r) & \text{if } t \geq s \\ 0 & \text{if } t < s, \end{cases}$$

where r and s are fixed times, $r < s$, $a \in \mathrm{R}^d$ and where we allow negative consumption levels for the sake of an explicit solution. Then $(V_t(c))$ solves

$$dV_t(c) = [-c_t + e(\sigma_t)]\, dt + \sigma_t \cdot dW_t, \quad V_T(c) = 0.$$

The unique solution is

$$\sigma_t = a\,(T-s) 1_{[r,s]}(t), \text{ and}$$

$$V_t(c) = \begin{cases} a \cdot (W_s - W_r)\,(T-t) & \text{if } s \leq t \leq T \\ a \cdot (W_t - W_r)\,(T-s) - e(a)\,(s-t), & \text{if } r \leq t < s \\ -e(a)\,(s-r), & \text{if } 0 \leq t < r. \end{cases}$$

In particular, when $r = 0$, then

$$V_t(c) = a \cdot W_t\,(T-s) - e(a)(s-t), \quad \text{for } t < s,$$

illustrating part (c) of the theorem.

4.5. Supergradients

The asset pricing application to follow makes use of the notion of supergradients for utility. A *supergradient* for V at the consumption process c is a process (π_t) satisfying

$$V(c') - V(c) \leq E\left[\int_0^T \pi_t \cdot (c'_t - c_t)\, dt\right], \tag{4.14}$$

for all c' in D. Denote by $\partial V(c)$ the set of supergradients at c.

Because V is a lower envelope of SDU functions V^Q (Theorem 2.3(b)), we can use a suitable envelope theorem to relate $\partial V(c)$ to supergradients of $\{V^Q : Q \in \mathcal{P}^\Theta\}$. For each SDU function V^Q, the set of supergradients may be completely characterized under suitable additional assumption (see [16]). The added assumption is that there exists $k > 0$ such that

$$\sup\left(\mid f_c(x,v) \mid, \mid f(x,0) \mid\right) < k\ (1+ \mid x \mid), \text{ for all } (x,v) \in C \times \mathrm{R}^1. \tag{4.15}$$

The above reasoning leads immediately to the following characterization of $\partial V(c)$. It uses the notation

$$\Theta_c = \{\theta^* \in \Theta : \theta^*_t \in \arg\max_{y \in \Theta_t} y \cdot \sigma_t \text{ all } t\},$$

for any $c \in D$, where (σ_t) is the (unique) volatility of utility defined by (2.16); recall (2.7).

Lemma 4.6. *Suppose that f is continuously differentiable and that it satisfies (4.15) and the assumptions of Theorem 2.3. Then: (a)*

$$\partial V(c) \supset \Pi \equiv \left\{\pi : \exists\, \theta^* \in \Theta_c,\ \pi_t = \exp\left(\int_0^t f_v(c_s, V_s(c))\, ds\right) f_c(c_t, V_t(c))\ z^{\theta^*}_t \text{ all } t\right\}. \tag{4.16}$$

(b) Suppose further that V is concave and that c lies in the interior of the domain D. Then $\partial V(c) = \Pi$.

See Appendix E for a proof. The set Π is alternatively expressed as

$$\Pi = \left\{\partial V^Q(c) : Q = Q^{\theta^*} \text{ and } \theta^* \in \Theta_c\right\},$$

the set of supergradients for the SDU functions V^{Q^*} where Q^* satisfies (2.17). Evidently, $\partial V(c)$ is a nonsingleton in general. For example, if c is deterministic, then $\Theta_c = \Theta$ because the appropriate σ_t vanishes. Under the conditions in (b), the containment in (4.16) can be strengthened to equality. The scope of (b) is limited, however, by the fact that the non-negative orthant of the Hilbert space of square integrable processes has empty interior. Thus the interiority assumption can be satisfied only if V is well-defined for some processes where consumption may be negative. We include (b) in order to show a sense in which divergence between $\partial V(c)$ and Π can be viewed as 'pathological'. In the asset pricing application that follows, we restrict attention to supergradients lying in Π and thus possibly to a proper subset of equilibria.

5. ASSET RETURNS

5.1. The Environment

There is a representative agent with recursive multiple-priors utility. Otherwise, the environment is standard (see [13] for elaboration and supporting technical details). There is a single consumption good, a riskless asset with return process r_t and d risky securities, one for each component of the Brownian motion W_t. Returns R_t to the risky securities are described by

$$dR_t \;=\; b_t\,dt \;+\; s_t\,dW_t,$$

where s_t is a $d \times d$ volatility matrix. Assume that markets are *complete* in the usual sense that s_t is invertible almost surely for every t. Market completeness delivers a (strictly positive) *state price* process π_t. Let

$$-d\pi_t \,/\, \pi_t \;=\; r_t\,dt \;+\; \eta_t \cdot dW_t,\; \pi_0 = 1, \tag{5.1}$$

where $\eta_t \;=\; s_t^{-1}(b_t - r_t\mathbf{1})$ and is typically referred to as the market price of risk. We refer to it as the *market price of uncertainty* to reflect the fact that security returns embody both risk and ambiguity.

Denote time t wealth by X_t and the trading strategy by ψ_t, where ψ_t^i is the proportion of wealth invested in risky security i. Thus $1 - \psi_t \cdot \mathbf{1}$ equals the proportion invested in the riskless asset. The law of motion for wealth is

$$dX_t \;=\; \left([r_t + \psi_t^\top(b_t - r_t\mathbf{1})]X_t \;-\; c_t\right)\;dt + X_t\,\psi_t^\top s_t\,dW_t,\;\; X_0 > 0 \text{ given.} \tag{5.2}$$

Budget feasible consumption processes may be characterized by the inequality

$$E\left[\int_0^T \;\pi_t c_t\,dt\right] \;\leq\; X_0. \tag{5.3}$$

First-order conditions for optimal consumption choice are expressed in the usual way in terms of the supergradient of utility at the optimum c.[32] In particular, c is optimal if

$$\exp\left(\int_0^t f_v(c_s, V_s(c))ds\right)\; f_c(c_t, V_t(c))\;\, z_t^{\theta^*} \;=\; f_c(c_0, V_0)\pi_t, \quad \text{for all } t, \tag{5.4}$$

for some process θ^* in Θ_c, where, as mentioned earlier, we are restricting attention to supporting supergradients in the set Π defined in (4.16). The multiple-priors model is reflected in the presence of the factor $z_t^{\theta^*}$ on the left side; $z_t^{\theta^*}$ is identically equal to 1 if beliefs are represented by P.

Suppose for the moment that the aggregator is the standard one (3.3), implying the first-order conditions

$$e^{-\beta t}u'(c_t)\;\, z_t^{\theta^*} \;=\; u'(c_0)\pi_t, \quad \text{for all } t. \tag{5.5}$$

[32] See [16] (and also [48]) for details regarding first-order conditions and their connection to security pricing.

A difficulty in fitting aggregate time-series data to this relation when $z_t^{\theta^*} = 1$, is that the observed volatility of consumption is too small relative to that of state prices to be consistent with this equation [29]. The presence of the factor $z_t^{\theta^*}$ has the potential to increase the variability of the left side and thus come closer to fitting observed moments.

Alternatively, we could view the term $z_t^{\theta^*}$ as transforming state prices. That is, rewrite the equation as

$$e^{-\beta t} u'(c_t) = u'(c_0) \left(\pi_t \, / \, z_t^{\theta^*}\right) \equiv u'(c_0) \, \widehat{\pi}_t \, , \tag{5.6}$$

which comprise the first-order conditions for an expected additive utility maximizer, but one who faces the 'effective' state price process $\widehat{\pi}_t$, rather than the original one π_t. For some specifications, $\widehat{\pi}_t$ is less variable than π_t, permitting a closer fit to a smooth consumption process.

It is convenient to write

$$dX_t / X_t = b^M \, dt + s^M \cdot dW_t, \tag{5.7}$$

where b^M is the mean return to the market portfolio and s^M is its (relative) volatility.

5.2. Ambiguity and Risk Premia

Let the consumer have a Kreps-Porteus aggregator (3.16), which affords a simple parametric distinction between the effects of intertemporal substitution and risk aversion.[33]

Suppose that the optimal consumption process satisfies

$$dc_t / c_t = \mu_t^c \, dt + s_t^c \cdot dW_t. \tag{5.8}$$

For the most part, we view consumption as a given endowment and we focus on characterizing the risk-free rate and market price of uncertainty that support (in the sense of satisfying (5.4)) the endowment as an optimum or representative-agent equilibrium.

Let θ^* be any process in Θ_c. From (5.4), (5.1) and Ito's Lemma, conclude that

$$\begin{aligned}(1-\rho)^{-1}(r_t - \beta) \quad = \quad & \mu_t^c - \frac{(2-\rho)}{2} s_t^c \cdot s_t^c - \theta_t^* \cdot s_t^c + \\ & (\alpha - \rho)\left(\frac{\sigma_t}{\alpha V_t}\right) \cdot \left[s_t^c + 2^{-1}\rho(1-\rho)^{-1}\left(\frac{\sigma_t}{\alpha V_t}\right)\right] \text{ and}\end{aligned} \tag{5.9}$$

[33] Theorem 2.3 and Lemma 4.6 do not apply because, for example, the Lipschitz condition is violated. We proceed assuming existence of an optimum and focus on its characterization. Schroder and Skiadas [48] provide conditions for existence given a Kreps-Porteus aggregator and no ambiguity. It remains to be seen how their analysis may be extended to accommodate ambiguity. In addition, the Kreps-Porteus aggregator is concave in (c, v) if and only if $\alpha \geq \rho$. However, V is concave and first-order conditions are sufficient for all admissible parameter values. Schroder and Skiadas prove this in the absence of ambiguity, while the multiple-priors structure 'adds concavity'.

A final point is that we have defined the Kreps-Porteus aggregator so as to exclude zero values for α or ρ. However, it can be defined for those parameter values in the usual limiting fashion and some of the results to follow remain valid in those cases.

$$(1-\rho)s_t^c = (\alpha-\rho)\left(\frac{\sigma_t}{\alpha V_t}\right) + (\eta_t - \theta_t^*), \tag{5.10}$$

where V_t and σ_t are the level and volatility of utility along the optimal consumption process. We can show that along an optimal path,[34]

$$\sigma_t / (\alpha V_t) = \rho^{-1}\left[s_t^M + (\rho-1)\, s_t^c\right]. \tag{5.11}$$

Substitution into (5.10) yields the following restriction for the market price of uncertainty:

$$\eta_t = \rho^{-1}\left[\alpha(1-\rho)\, s_t^c + (\rho-\alpha)\, s_t^M\right] + \theta_t^*.$$

We have thus obtained the following model of excess returns:[35]

$$b_t - r_t\mathbf{1} = s_t\eta_t = \rho^{-1}\left[\alpha(1-\rho)s_t\, s_t^c + (\rho-\alpha)s_t\, s_t^M\right] + s_t\,\theta_t^*. \tag{5.12}$$

The right side expresses excess returns as the sum of a *risk premium* (the first term) and the *ambiguity premium* $s_t\,\theta_t^*$. The risk premium consists of the two-factor model derived in [26] and [15], according to which systematic risk of an asset is measured by a linear combination of its covariation with consumption growth and with the market return.

For the ambiguity premium, observe that, using common notation,

$$s_t^i \cdot \theta_t^* = -\,cov_t(\,dR_t^i,\, dz_t^{\theta^*}/z_t^{\theta^*}\,),$$

for each security $i = 1,...,N$, where s_t^i denotes the i$^{\text{th}}$ row of s_t and $z_t^{\theta^*}$ is as in (5.4). Thus the premium is positive if the asset's return has negative instantaneous covariation with $dz_t^{\theta^*}/z_t^{\theta^*}$. Recall from (2.4) that $z_t^{\theta^*} = dQ_t^*/dP$, where Q_t^* is the restriction of $Q^* = Q^{\theta^*}$ to $\mathcal{F}_t$.

Alternatively, some insight into the ambiguity premium is provided by applying (5.11) to deduce that θ_t^* solves

$$\max_{y\in\Theta_t} y \cdot \left[\rho^{-1}s_t^M + (1-\rho^{-1})\, s_t^c\right].$$

This characterization of θ_t^* is not completely satisfactory because though s_t^c is exogenous in our endowment economy model, the volatility of the market return is endogenous.[36] Thus the next section describes alternative characterizations that are valid under suitable specializations of the endowment process and of Θ.

[34] Multiply through (5.4) by c_t and integrate over time and states, using $dt \otimes dP$, to obtain $E\left[\int_0^T \exp\left(\int_0^t f_v(c_s, V_s(c))ds\right) c_t\, f_c(c_t, V_t(c))\, z_t^{\theta^*}\, dt\right] = f_c(c_0, V_0)\, E\left[\int_0^T \pi_t c_t\, dt\right] = f_c(c_0, V_0)\, X_0 = c_0^{\rho-1} X_0\, (\alpha V_0)^{(\alpha-\rho)/\alpha}$, where we use the Kreps-Porteus aggregator. Given the latter, utility is homogeneous of degree α, that is, $V(\lambda c) = \lambda^\alpha\, V(c)$ for all $\lambda > 0$. Therefore, by a form of Euler's Theorem, the LHS above equals αV_0. Deduce that $\alpha V_0 = (c_0^{\rho-1} X_0)^{\alpha/\rho}$. In the same way, $\alpha V_t = (c_t^{\rho-1} X_t)^{\alpha/\rho}$ for all t. Now apply Ito's Lemma to obtain the desired expression for the volatility of V_t.

[35] Whenever we refer to 'expected returns' or other moments, the intention is expectation with respect to the reference measure P. When making connections to data, assume that P is the true probability measure.

[36] A similar criticism applies to the risk premium in (5.12) if $\alpha \neq \rho$. See [6] and the references therein for further discussion and for proposed solutions in risk-based models.

Before specializing further, we offer a final comment on the general model (5.12). From the Girsanov Theorem, we can write

$$dR_t = (b_t - s_t\theta_t^*)\,dt + s_t\,dW_t^*, \tag{5.13}$$

where (W_t^*) is a Brownian motion relative to Q^*. Thus the ambiguity premium $s_t\theta_t^*$ in (5.12) equals the adjustment in the drift needed when P is replaced by Q^*. More to the point, our prediction for $b_t - r_t\mathbf{1}$ is identical to that obtained in a model where the agent is indifferent to ambiguity and uses the 'distorted' single prior Q^*. This is an instance of the observational equivalence mentioned in the introduction. On the other hand, we would argue that specifying Q^* as a primitive is likely to seem contrived and less natural than the story provided here. Other advantages of our approach were described in Section 1.3.

5.3. Special Consumption Processes

We assume IID ambiguity throughout and consider three alternative specializations of the endowment process.

Markov consumption process: Assume that the drift and volatility in (5.8) are of the form

$$\mu_t^c = \widehat{\mu}(c_t, t) \text{ and } s_t^c = \widehat{s}(c_t, t) \tag{5.14}$$

for suitable functions $\widehat{\mu}$ and $\widehat{s}$. Then, under suitable restrictions, the corresponding utility process has the form

$$V_t = H(c_t, t),$$

for some function H.[37] If H is differentiable in the consumption argument, then the volatility of utility is simply $\sigma_t = c_t\,H_c(c_t, t)\,s_t^c$. In particular, if the noted derivative is everywhere positive, (intertemporal utility is an increasing function of current consumption), we obtain the following simple characterization:

$$\theta_t^* \text{ solves } \max_{\theta_t \in K} \theta_t \cdot s_t^c,$$

where $K \subset \mathrm{R}^d$ is the set corresponding to Θ as in Section 3.4. Under κ-ignorance, if

$$s_t^{c,j} \neq 0 \tag{5.15}$$

for each component $j = 1, ..., d$, then (3.7) and (3.8) deliver the closed-form expression

$$\theta_t^* = \kappa \otimes sgn\,(\,s_t^c). \tag{5.16}$$

[37] A detailed derivation could be based on the 4-step procedure from [36] applied to solve the FBSDE consisting of (2.16), (5.8) and (5.14). Intuitively, the point is simply that because of IID ambiguity and the Markov property for consumption, current consumption is the only state variable that is relevant for defining the utility process for (c_t).

The implied model of mean excess returns is given by substitution into (5.12). We could also use (5.9) to derive the implied risk-free rate. Such a derivation is described for the following further specialization.

Geometric consumption process: Suppose that both μ_t^c and s_t^c are deterministic constants, as in (3.15). (Deterministic time dependence is readily accommodated.) Then[38]

$$s_t^M = s_t^c. \tag{5.17}$$

Consequently, the market price of uncertainty satisfies

$$\eta_t = (1-\alpha)\, s_t^c + \theta_t^*. \tag{5.18}$$

If (5.15) and κ-ignorance are assumed for concreteness, then the expected excess return for asset i equals

$$b_t^i - r_t = (1-\alpha)\, s_t^i \cdot s_t^c + \kappa \cdot \left(s_t^i \otimes sgn(s_t^c)\right).$$

Thus the ambiguity premium (represented by the second term) for asset i is large if $s_t^{i,j}\, sgn(s_t^{c,j})$ is large and positive for components j of the driving process W_t that are very ambiguous in the sense of having large κ_j. Because the premium depends on the endowment process only via the signs of $s_t^{c,j}$, $j = 1, ..., d$, large ambiguity premia can occur even if consumption is relatively smooth.

Of special interest is the excess return to the market portfolio given by

$$b_t^M - r_t = (1-\alpha) s_t^c \cdot s_t^c + \kappa \cdot \mid s_t^c \mid, \tag{5.19}$$

providing a decomposition of the *equity premium* in terms of risk (the first term) and ambiguity (the second term).[39] The ambiguity premium for the market portfolio vanishes as s^c approaches zero. However, because it is a first-order function of volatility, it dominates the risk premium for small volatilities.

Combine the preceding to yield

$$b_t^i - r_t = \left[\frac{[(1-\alpha)\, s_t^i \cdot s_t^M + \kappa \cdot \left(s_t^i \otimes sgn(s_t^M)\right)]}{(1-\alpha)\, s_t^M \cdot s_t^M + \kappa \cdot \mid s_t^M \mid}\right] (b_t^M - r_t),$$

a variant of CAPM. Ambiguity leads to a large excess return for asset i if $s_t^{i,j} s_t^{M,j} > 0$ for components j of the Brownian motion for which κ_j is large. However, unlike the case for the risk premium, the ambiguity premium depends only on the sign of each $s_t^{M,j}$ and not on its magnitude.

[38] By the homogeneity of intertemporal utility, c_t and wealth X_t are related by $c_t = a_t X_t$ for some deterministic a_t. The claim follows by Ito's Lemma.

[39] Much of what has been derived extends to IID ambiguity, in which case the ambiguity premium on the right side of (5.19) is given by $e(s^c)$. See Lemma A.2 for an interpretation of $e(\cdot)$.

To study the relation between the risk-free rate and the drift in consumption, substitute (5.11) and (5.17) into (5.9) to obtain

$$\mu_t^c - \frac{(1-\alpha)(2-\rho)}{2(1-\rho)} s_t^c \cdot s_t^c - \theta_t^* \cdot s_t^c = (1-\rho)^{-1}(r_t - \beta). \tag{5.20}$$

Under κ-ignorance, we obtain the following expression for the risk-free rate:

$$r_t - \beta = (1-\rho)\left(\mu_t^c - \frac{(1-\alpha)(2-\rho)}{2(1-\rho)} s_t^c \cdot s_t^c - \kappa \cdot \mid s_t^c \mid\right), \tag{5.21}$$

which is decreasing in risk aversion $(1-\alpha)$ and in ambiguity aversion κ.

Stochastic drift and volatility: Generalize the Markov model by permitting more general specifications for the stochastic nature of the drift and volatility of consumption growth. Specifically, suppose that there exists an R^ℓ-valued state variable ω_t such that the joint process (c_t, ω_t) is Markovian, that is, (using slightly abused but transparent notation)

$$dc_t/c_t = \mu_t^c(c_t, \omega_t)\,dt + s_t^c(c_t, \omega_t) \cdot dW_t \text{ and}$$

$$d\omega_t = \mu_t^\omega(c_t, \omega_t)\,dt + s_t^\omega(c_t, \omega_t)\,dW_t.$$

The new twist in this model relative to the earlier one is that we exploit the auxiliary state process (ω_t) in order to model a situation in which there is ambiguity about the stochastic evolution of the drift and volatility of consumption growth but not about its conditional distribution. Formally, use the κ-ignorance specification for Θ and assume that

$$\kappa_i\, s_t^{c,i} = 0, \quad \text{for } i = 1, ..., d. \tag{5.22}$$

This suggests the decomposition $W_t = (W_t^c, W_t^\omega)$ such that consumption growth is driven by W_t^c, ω_t is driven by both W_t^c and W_t^ω and there is ambiguity only about the latter.

By arguments similar to those outlined for the Markov model, one can justify the following expression for the utility of the endowment process:

$$V_t = H(c_t, \omega_t, t),$$

for a suitable H. If the latter is differentiable, Ito's Lemma yields

$$\sigma_t = c_t\, H_c(c_t, \omega_t, t)\, s_t^c + H_\omega(c_t, \omega_t, t)\, s_t^\omega.$$

Apply (5.22) to deduce that

$$\theta_t \cdot \sigma_t = \theta_t \cdot \left(H_\omega(c_t, \omega_t, t)\, s_t^\omega\right), \quad \theta \in \Theta.$$

If some components of $H_\omega s_t^\omega$ are zero, then $\max_{y \in \Theta_t} y \cdot \sigma_t$ has many solutions. Focus on that given by (3.8) and on the corresponding equilibrium.

The implied excess returns are obtained from the appropriate form of (5.12). For convenience, we reproduce the result here in the special case $\alpha = \rho$:

$$b_t - r_t\mathbf{1} \ = (1-\alpha)s_t \ s_t^c \ + \ s_t \ [\kappa \otimes sgn\,(H_\omega \ s_t^\omega)\,] \, .$$

Three features of this result are noteworthy. First, in the standard expected utility risk-based model, mean excess returns at any time and state of the world depend on the endowment process only via its current volatility and hence via the associated conditional distribution of consumption. In contrast, ambiguity aversion leads, through s_t^ω, to a dependence also on the instantaneous change in the conditional distribution of consumption.

Second, observe that the ambiguity premium can be large even if s_t^ω is small in norm. For example, take the case where ω_t is real-valued and suppose that H_ω is everywhere positive (a globally negative sign would do as well). Then the ambiguity premium for the i^{th} asset equals $s_t^i \cdot \ [\kappa \otimes sgn\,(\ s_t^\omega)\,]$, which depends on s_t^ω only through its sign.

Finally, the ambiguity premium undergoes discrete *jumps* at points where components of $H_\omega s_t^\omega$ change sign, even though the stochastic environment is Brownian and hence continuous. For example, if $\ell = 1$ and s_t^ω is constant, then θ_t^* jumps wherever H_ω changes sign and rates of return follow a two-state switching model.[40]

5.4. Optimal Portfolio

Turn to an examination of the optimal portfolio for a consumer facing an exogenous and deterministic risk-free rate and market price of uncertainty, taken for simplicity to be constants. Then the optimal consumption process is geometric and from (5.18),

$$sgn(s_t^{c,i}) \ = \ sgn\left(\eta_i - \ \kappa_i sgn(s_t^{c,i})\right) \ \text{ for each } i.$$

Assume that ambiguity aversion is small in the sense that

$$0 \ \leq \ \kappa_i \ < \mid \eta_t^i \mid \ \text{ for all } i. \tag{5.23}$$

Then

$$s_t^{c,i} > \ (<)\, 0 \ \text{ if } \eta_t^i > \ (<)\, 0,$$

implying with (5.16) that $\theta_t^* \ = \ \kappa \otimes sgn(\eta_t)$ and

$$(1-\alpha)\, s_t^c \ = \ \eta_t - \ \ \kappa \otimes sgn(\eta_t).$$

Finally, it follows from (5.2) and (5.17) that the optimal portfolio of risky assets is given by

$$\psi_t \ = \ (1-\alpha)^{-1}\, (s_t^\top)^{-1} \ (\eta_t \ - \kappa \otimes sgn(\eta_t))\, .$$

Evidently, the optimal portfolio is not instantaneously mean-variance efficient if P is used to compute variance. Our interpretation is that this is due to ambiguity being present in

[40] Thus far we have been unable to find a parametric example where this endogenous regime-switching can be demonstrated in closed-form.

addition to risk.[41] The mutual fund separation property is valid if and only if κ is common to all agents. Though the composition of risky assets is independent of the risk aversion parameter α, it depends on preferences through κ.

A. APPENDIX: BSDE's and Related Results

For the convenience of the reader, this appendix outlines *informally* some material regarding BSDE's. See [18] and [44] for further reading and formal details that are ignored here.

The stochastic environment $\left(\Omega, \{\mathcal{F}_t\}_0^T, P\right)$ used throughout the paper is assumed.

Given $\xi \in L^2(\Omega, F_T, P)$ and a function $g : \mathrm{R}^1 \times \mathrm{R}^d \times \Omega \times [0,T] \longrightarrow \mathrm{R}^1$, consider the problem of finding processes (y_t) and (σ_t) satisfying the BSDE

$$dy_t = g(y_t, \sigma_t, \omega, t)\, dt + \sigma_t \cdot dW_t, \quad y_T = \xi. \tag{A.1}$$

The existence of a unique solution may be proven under Lipschitz and other technical conditions for g [18, Theorem 2.1].[42] Our definition of intertemporal utility for a given consumption process c (Theorem 2.3) deals with the special case

$$g(y, \sigma, \omega, t) = -f(c_t(\omega), y) + \max_{\theta \in \Theta} \theta_t(\omega) \cdot \sigma.$$

The following result [18, Theorem 2.2] was referred to in the text and is used in the sequel.

Theorem A.1. ***(Comparison)*** *Consider the BSDE above corresponding to (g, ξ) and that associated with another pair (g', ξ'). Let corresponding unique solutions be (y_t, σ_t) and (y'_t, σ'_t). Suppose that*

$$\xi' \geq \xi \text{ and } g'(y_t, \sigma_t, \omega, t) \leq g(y_t, \sigma_t, \omega, t) \quad dt \otimes dP \text{ a.e.}$$

Then $y'_t \geq y_t$ for almost every $t \in [0,T]$. Moreover, the comparison is strict in the sense that if, in addition, $y'_\tau = y_\tau$ on the event $A \in \mathcal{F}_\tau$, then $\xi' = \xi$ on A and

$$g'(y_t, \sigma_t, \omega, t) = g(y_t, \sigma_t, \omega, t) \quad \text{on } [\tau, T] \times A \;\; dt \otimes dP \text{ a.e.}$$

[41] Alternatively, mean-variance efficiency is optimal if variance is computed using the appropriate measure Q^{θ^*}, as provided by Theorem 2.3(b). However, Q^{θ^*} depends on preferences through κ and thus the meaning of mean-variance efficiency is individual specific.

[42] The analysis in [18] relies on the predictability of $(\omega, t) \longmapsto g(y, \sigma, \omega, t)$. In our context, this would require predictability of consumption processes. However, the arguments in [43]- [44] rely only on progressive measurability of the above map for each fixed (y, σ); recall that we have assumed progressive measurability of all (including consumption) processes. Thus the key existence and comparison theorems (see below) are valid for our setting.

A further specialization of (A.1) has $f \equiv 0$, or

$$dy_t = [\max_{\theta \in \Theta} \theta_t \cdot \sigma_t] \, dt + \sigma_t \cdot dW_t, \;\; y_T = \xi. \tag{A.2}$$

For given Θ (satisfying our assumptions) and each t, the map $\xi \longmapsto y_t$ defines a nonlinear functional from $L^2(\Omega, F_T, P)$ into $\mathcal{F}_t$-measurable random variables. Use the notation $\mathcal{E}[\xi \mid \mathcal{F}_t]$ for y_t, suggesting a form of *nonlinear conditional expectation* [44]. In fact,

$$\mathcal{E}[\xi \mid \mathcal{F}_t] = \min_{Q \in \mathcal{P}^{\Theta}} E_Q[\xi \mid \mathcal{F}_t].$$

Evidently, $E[\xi \mid \mathcal{F}_t] - \mathcal{E}[\xi \mid \mathcal{F}_t]$ is a form of premium due to ambiguity.

Lemma A.2. *Consider a consumption process c satisfying*

$$dc_t = \mu_t^c \, dt + s_t^c \cdot dW_t, \;\; 0 \leq t \leq T, \;\; c_0 \text{ given}, \tag{A.3}$$

where (μ_t^c) and (s_t^c) are continuous and bounded (adapted) processes. Let $e(\cdot)$ be as in (3.10). Then

$$\lim_{\tau \longrightarrow r^+} \frac{E[c_\tau \mid \mathcal{F}_r] - \mathcal{E}[c_\tau \mid \mathcal{F}_r]}{\tau - r} = e(s_r^c),$$

where the limit is in the sense of $L^2(\Omega, \mathcal{F}_T, P)$.

The lemma provides the interpretation for $e(\cdot)$ promised at the end of Section 3.4 - it provides an instantaneous, per unit time premium for ambiguity.

Proof. For any $0 \leq \tau \leq T$, there exists a unique solution $(\mathcal{E}[c_\tau \mid \mathcal{F}_t], \sigma_t^\tau)$ to

$$\mathcal{E}[c_\tau \mid \mathcal{F}_t] - c_\tau = -\int_t^\tau e(\sigma_u^\tau) \, du - \int_t^\tau \sigma_u^\tau \cdot dW_u, \;\; 0 \leq t \leq \tau. \tag{A.4}$$

For $t = r$ we obtain, using also (A.3), that

$$\mathcal{E}[c_\tau \mid \mathcal{F}_r] - c_r = \int_r^\tau (\mu_u^c - e(\sigma_u^\tau)) \, du + \int_r^\tau (s_u^c - \sigma_u^\tau) \cdot dW_u, \;\; 0 \leq r \leq \tau.$$

By [4, Propn. 2.2], (see also [18, Propn. 2.1]), there exist constants $\beta > 0$ and $\gamma > 0$ such that[43]

$$E\left[\frac{1}{(\tau - r)} \int_r^\tau \| s_u^c - \sigma_u^\tau \|^2 \, du \mid \mathcal{F}_r\right] \leq \gamma \frac{1}{(\tau - r)} E\left[\int_r^\tau e^{\beta u} \mid \mu_u^c - e(s_u^c) \mid du \mid \mathcal{F}_r\right]^2$$

$$\leq \gamma \, E\left[\int_r^\tau e^{2\beta u} \mid \mu_u^c - e(s_u^c) \mid^2 \, du \mid \mathcal{F}_r\right]$$

[43]Use the inequality $\frac{1}{\tau - r} \left(\int_r^\tau x_u \, du\right)^2 \leq \int_r^\tau x_u^2 \, du$.

$$\leq \gamma e^{2\beta(T-r)} (\tau - r) E \left[\sup_{r \leq u \leq T} |\mu_u^c - e(s_u^c)|^2 \mid \mathcal{F}_r \right].$$

It follows that

$$E\left[\frac{1}{(\tau - r)} \int_r^\tau \| s_u^c - \sigma_u^\tau \|^2 \, du \mid \mathcal{F}_r \right] \underset{\tau \to r^+}{\longrightarrow} 0 \text{ in } L^1(\Omega, \mathcal{F}_T, P).$$

Because $e(\cdot)$ is Lipschitz continuous and (s_t^c) is a continuous process,

$$\lim_{\tau \longrightarrow r^+} E\left[\frac{1}{\tau - r} \int_r^\tau | e(s_u^c) - e(\sigma_u^\tau) |^2 \, du \mid \mathcal{F}_r \right] = 0 \text{ and}$$

$$\lim_{\tau \longrightarrow r^+} E\left[\frac{1}{\tau - r} \int_r^\tau | e(s_u^c) - e(s_r^c) |^2 \, du \mid \mathcal{F}_r \right] = 0$$

also in $L^1(\Omega, \mathcal{F}_T, P)$. Therefore, the triangle inequality implies that

$$\lim_{\tau \longrightarrow r^+} E\left[\frac{1}{\tau - r} \int_r^\tau | e(\sigma_u^\tau) - e(s_r^c) |^2 \, du \mid \mathcal{F}_r \right] = 0$$

and hence also that

$$\lim_{\tau \longrightarrow r^+} E\left[\frac{1}{\tau - r} \int_r^\tau e(\sigma_u^\tau) \, du - e(s_r^c) \mid \mathcal{F}_r \right]^2 = 0,$$

once again in $L^1(\Omega, \mathcal{F}_T, P)$. Conclude that $E\left[\frac{1}{\tau - r} \int_r^\tau e(\sigma_u^\tau) \, du \mid \mathcal{F}_r \right]$ converges to $e(s_r^c)$ in $L^2(\Omega, \mathcal{F}_T, P)$.

Finally, set $t = r$ in (A.4) and apply the conditional expectation $E[\cdot \mid \mathcal{F}_r]$ to both sides to obtain

$$\frac{\mathcal{E}[c_\tau \mid \mathcal{F}_r] - E[c_\tau \mid \mathcal{F}_r]}{\tau - r} = -(\tau - r)^{-1} E[\int_r^\tau e(\sigma_u^\tau) \, du \mid \mathcal{F}_r],$$

which converges to $-e(s_r^c)$ in $L^2(\Omega, \mathcal{F}_T, P)$. ■

B. APPENDIX: Density Generators and the Set of Priors

Proof of Lemma 2.1: (b) The process θ^* is delivered by the Measurable Maximum Theorem [2, Theorem 14.91], which ensures that there exists a progressively measurable selection from $\arg\max_{y \in \Theta_t(\omega)} y \cdot \sigma_t(w)$. (To apply the Maximum Theorem, use the progressive σ-field [45, p. 44] on $[0, T] \times \Omega$.) It ensures also that the value function for the latter problem is suitably measurable.
(c) Stochastic convexity is obvious. Weak compactness follows from [17, Theorems IV.8.9, V.6.1]. ■
Proof of Theorem 2.2: (b) Fix $A \in \mathcal{F}_T$ and $Q^\theta \in \mathcal{P}^\Theta$. By Girsanov's Theorem, $Q^\theta(A \mid \mathcal{F}_t) = y_t$, where (y_t, σ_t) is the unique solution to

$$dy_t = \theta_t \cdot \sigma_t \, dt + \sigma_t \cdot dW_t, \quad y_T = 1_A.$$

By the bounding inequality in [18, p. 20] and Uniform Boundedness, there exists $k > 0$ such that

$$\left(Q^{\theta}(A)\right)^2 \leq kE(1_A) = k\,P(A),$$

where k is independent of θ. This delivers uniform absolute continuity. Equivalence obtains because $z_T^{\theta} > 0$ for each θ.

(c) For $i = 1, 2$, let Q^i be the measure corresponding to $\theta^i \in \Theta$ and the martingale z_t^i as in (2.4). Define $\theta = (\theta_t)$ by

$$\theta_t = \frac{(\theta_t^1 z_t^1 + \theta_t^2 z_t^2)}{z_t^1 + z_t^2}.$$

(Recall that z_t^1 and z_t^2 are strictly positive.) Then $\theta \in \Theta$ and $d(z_t^1 + z_t^2) = -(z_t^1 + z_t^2)\,\theta_t \cdot dW_t$, which implies that $(z_T^1 + z_T^2)/2$ is the density for $(Q^1 + Q^2)/2$. Conclude that the latter lies in $\mathcal{P}^{\Theta}$. Similarly for other mixtures.

(d) Using the weak compactness of Θ (Lemma 2.1), one can show that $Z = \{z_T^{\theta} : \theta \in \Theta\}$ is norm-closed in $L^1(\Omega, \mathcal{F}_T, P)$. (The argument is analogous to the proof of Lemma B.2 in [10].) Because Z is convex, it is also weakly closed. Clearly, Z is norm-bounded ($E\left(\mid z_T^{\theta} \mid\right) = 1$ for all θ.) Thus, Z is weakly compact by the Alaoglu Theorem. Finally, Z is homeomorphic to $\mathcal{P}^{\Theta}$ when weak topologies are used in both cases.

(e) Follows from Lemma 2.1(b).

(f) For '$\subset$', given Q in $\mathcal{P}^{\Theta}$, let $Q^i = Q$ for all i. Turn to the nontrivial direction '$\supset$'. Let θ^i and z^i correspond to Q^i as in (2.4). Define $\theta \in \Theta$ by

$$\theta_t = \left(1_{\{t \geq \tau\} \times B}\,\theta_t^1 + 1_{\{t \geq \tau\} \times B^c}\,\theta_t^2\right) + \theta_t^3\,1_{\{t < \tau\}}.$$

Let z be the martingale generated as in (2.4); it satisfies

$$z_t = \begin{cases} z_t^3 & \text{if } t < \tau \\ z_T\, z_t^1 \,/\, z_T^1 & \text{if } t \geq \tau,\ \omega \in B \\ z_T\, z_t^2 \,/\, z_T^2 & \text{if } t \geq \tau,\ \omega \notin B \end{cases}$$

Then, for any A in $\mathcal{F}_T$,

$$Q(A) = \int \left[Q^1(A \mid \mathcal{F}_\tau)1_B + Q^2(A \mid \mathcal{F}_\tau)1_{B^c}\right] dQ_\tau^3$$

$$= \int \left[Q^1(A \mid \mathcal{F}_\tau)1_B + Q^2(A \mid \mathcal{F}_\tau)1_{B^c}\right] z_\tau^3\, dP_\tau =$$

$$\int \left[Q^1(A \mid \mathcal{F}_\tau)1_B + Q^2(A \mid \mathcal{F}_\tau)1_{B^c}\right] z_\tau\, dP_\tau =$$

$$E\left[z_\tau \frac{E[1_A 1_B\, z^1(T) \mid \mathcal{F}_\tau]}{E[z^1(T) \mid \mathcal{F}_\tau]} + z_\tau \frac{E[1_A 1_{B^c}\, z^2(T) \mid \mathcal{F}_\tau]}{E[z^2(T) \mid \mathcal{F}_\tau]}\right] =$$

$$E\left[1_B\, z_\tau \frac{E[1_A\, z(T) \mid \mathcal{F}_\tau]}{E[z(T) \mid \mathcal{F}_\tau]} + 1_{B^c}\, z_\tau \frac{E[1_A\, z(T) \mid \mathcal{F}_\tau]}{E[z(T) \mid \mathcal{F}_\tau]}\right] =$$

$$E\left[E[1_A z(T) \mid \mathcal{F}_\tau]\right] = E[1_A z(T)],$$

where use has been made of $B \in \mathcal{F}_\tau$, the martingale nature of (z_t) and the law of iterated expectations. Conclude that $z_T = z_T^{\theta}$ is the density of Q and hence that $Q \in \mathcal{P}^{\Theta}$. ■

C. APPENDIX: Proof of Existence of Utility

Proof of Theorem 2.3: (a) First we prove Lipschitz continuity of the support function e defined in (2.8). Let x and x' be in R^d and suppose that $e_t(x) = y \cdot x$ and $e_t(x') = y' \cdot x$ for y and y' in Θ_t; the dependence on ω has been suppressed notationally. Then

$$e_t(x) - e_t(x') \leq y \cdot (x - x') \leq d \mid y \mid \mid x - x' \mid \text{ and}$$

$$e_t(x) - e_t(x') \geq y' \cdot (x - x') \geq -d \mid y' \mid \mid x - x' \mid .$$

Now use Uniform Boundedness ($\Theta_t(\omega) \subset K$ and K compact.)

By the existence and uniqueness result in [43], there exist unique solutions (V_t, σ_t) and (V_t^Q, σ_t^Q) to (2.16) and (2.15) respectively.

(b) The Comparison Theorem and $\theta_t \cdot x_t \leq \max_{y \in \Theta_t} y \cdot x_t$ for any x_t, imply that $V_t \leq \min_{Q \in \mathcal{P}^\Theta} V_t^Q$. On the other hand, by Lemma 2.1(b), there exists θ^* in Θ such that

$$dV_t = [-f(c_t, V_t) + \theta_t^* \cdot \sigma_t]\, dt + \sigma_t \, dW_t, \;\; V_T = 0; \tag{C.1}$$

in other words, $V_t = V_t^{Q^{\theta^*}} \geq \min_{Q \in \mathcal{P}^\Theta} V_t^Q$, proving equality and hence (2.12). Uniqueness is covered by the uniqueness results in [44].

(c) Case 1: Suppose that $f(c, \cdot)$ is decreasing for each c in C.

Let $\tau = T$. By Girsanov's Theorem,

$$V_t^Q = E_Q \left[\int_t^T f(c_s, V_s^Q) \, ds \mid \mathcal{F}_t \right] .$$

Thus $V_t = \min_{Q \in \mathcal{P}^\Theta} V_t^Q =$

$$\min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^T f(c_s, V_s^Q) \, ds \mid \mathcal{F}_t \right] \leq$$

$$\min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^T f(c_s, \min_{Q \in \mathcal{P}^\Theta} V_s^Q) \, ds \mid \mathcal{F}_t \right]$$

$$= \min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^T f(c_s, V_s) \, ds \mid \mathcal{F}_t \right] .$$

On the other hand, (C.1) and Girsanov's Theorem imply that

$$V_t = E_{Q^*} \left[\int_t^T f(c_s, V_s) \, ds \mid \mathcal{F}_t \right]$$

$$\geq \min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^T f(c_s, V_s) \, ds \mid \mathcal{F}_t \right] .$$

For general τ, denote V_τ by ξ. Then (V_t, σ_t) is the unique solution to the BSDE (on $[0, \tau]$)

$$dV_t = \left[-f(c_t, V_t) + \max_{\theta \in \Theta} \theta_t \cdot \sigma_t\right] dt + \sigma_t dW_t, \quad V_\tau = \xi.$$

The fact that the terminal value ξ is nonzero is of no consequence for the preceding arguments. In particular, (V_t) solves

$$V_t = \min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^\tau f(c_s, V_s)\, ds + \xi \mid \mathcal{F}_t\right].$$

Case 2: Let f be arbitrary. For the given process c, define

$$F(t, v) \equiv -Kv + e^{Kt} f(c_t, e^{-Kt} v),$$

where K is the Lipschitz constant for f. Then, $F(t, \cdot)$ is decreasing and thus by Claim 1, (the time dependence is of no consequence), there exists a unique (V_t') solving

$$V_t' = \min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^T F(s, V_s')\, ds \,|\, \mathcal{F}_t\right]. \tag{C.2}$$

For this fixed (V_t'), define further the function

$$H(t, v) \equiv -Kv + e^{Kt} f(c_t, e^{-Kt} V_t').$$

Again by Claim 1, there exists a unique $(\overline{V}_t)$ solving

$$\overline{V}_t = \min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^T H(s, \overline{V}_s)\, ds \,|\, \mathcal{F}_t\right]. \tag{C.3}$$

Comparison of (C.2) and (C.3) and the uniqueness of solutions yield the equality

$$V_t' = \overline{V}_t. \tag{C.4}$$

Furthermore, by (b), we have

$$\overline{V}_t = \min_{Q \in \mathcal{P}^\Theta} V_t^Q, \tag{C.5}$$

where (V_t^Q) solves the BSDE

$$dV_t^Q = [-H(t, V_t^Q) + \theta_t \sigma_t]\, dt + \sigma_t\, dW_t, \quad V_T^Q = 0.$$

This linear BSDE has explicit solution

$$V_t^Q e^{-Kt} = E_Q \left[\int_t^T f(c_s, e^{-Ks} V_s')\, ds \,|\, \mathcal{F}_t\right], \quad \forall Q \in \mathcal{P}^\Theta.$$

Combine with (C.4) and (C.5) to deduce that

$$V_t' e^{-Kt} = \overline{V}_t e^{-Kt} = \min_{Q \in \mathcal{P}^\Theta} V_t^Q e^{-Kt} = \min_{Q \in \mathcal{P}^\Theta} E_Q \left[\int_t^T f(c_s, e^{-Ks} V_s')\, ds \,|\, \mathcal{F}_t\right],$$

implying that $V_t = V_t' e^{-Kt}$ solves

$$V_t = \min_{Q \in \mathcal{P}^\Theta} E_Q[\int_t^T f(c_s, V_s)\, ds \,|\, \mathcal{F}_t].$$

Similarly for $\tau < T$. ■

D. APPENDIX: Ambiguity

Proof of Lemma 4.2**:** Assume (4.5). Given Q^θ in $\mathcal{P}^\Theta$, $Q^\theta(F) = y_0$ where (y_t, σ_t) is the unique $\{\mathcal{F}_t\}$-adapted solution to the BSDE

$$dy_t = \theta_t \cdot \sigma_t dt + \sigma_t \cdot dW_t, \;\; y_T = 1_F.$$

If $F \in \mathcal{F}_T$, then $\sigma_t^i = 0$ if $P(F \mid \mathcal{F}_T^i) = 0$ or 1. Thus $\theta_t \cdot \sigma_t = 0$ and the BSDE reduces to the one defining $P(F)$, namely where $\theta = 0$. Therefore, $Q(F) = P(F)$.

For the converse, suppose that all measures agree on F. Then $y_0 = y_0'$, where

$$dy_t = \max_{\theta \in \Theta} \theta_t \cdot \sigma_t dt + \sigma_t \cdot dW_t, \;\; y_T = 1_F.$$

$$dy_t' = \min_{\theta \in \Theta} \theta_t \cdot \sigma_t' dt + \sigma_t' \cdot dW_t, \;\; y_T = 1_F.$$

By the strict portion of the Comparison Theorem A.1,

$$\max_{\theta \in \Theta} \theta_t \cdot \sigma_t = \min_{\theta \in \Theta} \theta_t \cdot \sigma_t',$$

or, in terms of the support function (3.10), $e(\sigma_t) = -e(-\sigma_t)$. By the non-negativity of e, $e(\sigma_t) = 0$. Now apply (3.13) to conclude that if $K^i \neq \{0\}$, then $\sigma_t^i = 0$, which implies $P(F \mid \mathcal{F}_T^i) = 0$ or 1. ■

Proof of Theorem 4.3:(i) $\Longleftarrow$: Because $\kappa_i^* = 0$ implies $\kappa_i = 0$, it follows from Lemma 4.2 that $\mathcal{U}^* \subset \mathcal{U}$. The consumption processes that are unambiguous for V^* are those that are adapted to the filtration generated by $\{W_t^i : \kappa_i^* = 0\}$. On such processes, V and V^* coincide with V^P, the Kreps-Porteus utility having measure P and parameters (β, ρ, α). That is,

$$V(c^{ua}) = V^*(c^{ua}) = V^P(c^{ua}).$$

Therefore, it is enough to prove that $V^*(c) \leq V(c)$ for all consumption processes c. This follows from $\kappa^* \geq \kappa$, (3.9) and the Comparison Theorem A.1.

$\Longrightarrow$: The above argument is reversible. First, $\mathcal{U}^* \subset \mathcal{U}$ implies that $\kappa_i = 0$ whenever $\kappa_i^* = 0$. From the definition of 'more ambiguity averse' it follows that V and V^* agree in the ranking of V^*-unambiguous consumption processes. These processes are deterministic if $\kappa^* >> 0$, in which case we can conclude only that V and V^* agree in the ranking of deterministic processes and therefore that $(\beta^*, \rho^*) = (\beta, \rho)$. However, under the assumption that $\kappa_i^* = 0$ for some i, there exist sufficiently many stochastic processes that are V^*-unambiguous in order to conclude that the risk aversion parameters α and α^* must be equal. Finally, apply (4.7), (3.9) and the Comparison Theorem to deduce that $\kappa^* \geq \kappa$.

(ii) $\Longleftarrow$: It follows from Lemma 4.2, that on V-unambiguous processes, V^* agrees with $V^{P,(\beta^*,\rho^*,\alpha^*)}$, the Kreps-Porteus utility with measure P and parameters $(\beta^*, \rho^*, \alpha^*)$, while V agrees with $V^{P,(\beta,\rho,\alpha)}$, defined similarly. Thus the comparative risk aversion statement follows from [14]. The converse is similar to (i). ■

E. APPENDIX: Proofs of Properties of Utility

Proof of Theorem 4.1: Fix t in $[0,T]$ and consider processes c of the form (3.5) where ξ is R_+^ℓ-valued. Denote by $\overline{D}_t$ the set of all such acts. Clearly $\overline{D}_t \subset D_t$ and is isomorphic to the set of all R_+^ℓ-valued and $\mathcal{F}_t$-measurable square integrable random variables on Ω. The restriction of V to $\overline{D}_t$ is probabilistically sophisticated (in the sense of [37]) and the associated probability measure is nonatomic. Finally, because of (4.1), all measures in $\mathcal{P}^\Theta$ agree on an event A in $\mathcal{F}_t$ with common measure in $(0,1)$; indeed, any A in $\mathcal{F}_t^1$ with $0 < P(A) < 1$ will do. Under these conditions, Marinacci [40] shows that all measures in $\mathcal{P}^\Theta$ agree on all of $\mathcal{F}_t$. This argument applies for every $t \leq T$. ■

Proof of Theorem 4.5:
(a) and (b): To make explicit the dependence on the driving process (W_t), write $V_t^T(c,W)$ to denote the solution of

$$V_t^T(c,W) = \int_t^T [f(c_s, V_s^T(c,W)) - e(\sigma_s(c))]\, ds - \int_t^T \sigma_s(c) \cdot dW_s. \tag{E.1}$$

Let $\overline{\mathcal{F}}_t = \sigma(W_{s+r} - W_r : s \leq t)$ and $\overline{W}_t = W_{t+r} - W_r$ for $0 \leq t \leq T-r$. Then $(\overline{W}_t)_{0 \leq t \leq T-r}$ is $\{\overline{\mathcal{F}}_t\}$-Brownian motion under P and $(\overline{c}_t)_{0 \leq t \leq T-r} = (c_{t+r})_{0 \leq t \leq T-r}$ is $\{\overline{\mathcal{F}}_t\}$-adapted. Thus there is a unique solution $(V_t^{T-r}(\overline{c},\overline{W}), \sigma(\overline{c}))$ to

$$V_t^{T-r}(\overline{c},\overline{W}) = \int_t^{T-r} [f(\overline{c}_s, V_s^{T-r}(\overline{c},\overline{W})) \; - \; e(\sigma_s(\overline{c}))] ds - \int_t^{T-r} \sigma_s(\overline{c}) \cdot d\overline{W}_s, \tag{E.2}$$

where t varies over $[0, T-r]$. After the change of variables $l = t + r$, this can be rewritten as

$$V_{l-r}^{T-r}(\overline{c},\overline{W}) = \int_{l-r}^{T-r} [f(\overline{c}_s, V_s^{T-r}(\overline{c},\overline{W})) \; - \; e(\sigma_s(\overline{c}))]\, ds - \int_{l-r}^{T-r} \sigma_s(\overline{c}) \cdot d\overline{W}_s, \tag{E.3}$$

for $r \leq l \leq T$. The further change $u = s + r$ yields

$$V_{l-r}^{T-r}(\overline{c},\overline{W}) = \int_l^T [f(\overline{c}_{u-r}, V_{u-r}^{T-r}(\overline{c},\overline{W})) - e(\sigma_{u-r}(\overline{c}))]\, du - \int_l^T \sigma_{u-r}(\overline{c}) \cdot d\overline{W}_{u-r}.$$

Because $\overline{c}_{u-r} = c_u$ and $d\overline{W}_{u-r} = dW_u$, deduce that $(V_{t-r}^{T-r}(\overline{c},\overline{W}), \sigma_{t-r}(\overline{c}))_{r \leq t \leq T}$ solves (on $[r,T]$)

$$V_t^T(c,W) = \int_t^T [f(c_s, V_s^T(c,W)) - e(\sigma_s(c))]\, ds - \int_t^T \sigma_s(c) \cdot dW_s. \tag{E.4}$$

That is, $V_{t-r}^{T-r}(\overline{c},\overline{W}) = V_t^T(c,W)$ and $\sigma_{t-r}(\overline{c}) = \sigma_t(c)$ for $t \in [r,T]$. In particular, choosing $t = r$, we have $V_r^T(c,W) = V_0^{T-r}(\overline{c},\overline{W})$, which is deterministic.

Rewrite (E.4) as

$$V_t^T(c,W) = V_r^T(c,W) + \int_t^r [f(c_s, V_s^T(c,W) - e(\sigma_s(c)]\, ds - \int_t^r \sigma_s(c) \cdot dW_s,$$

for $0 \leq t \leq r$. By hypothesis, c_t is deterministic for $0 \leq t \leq r$. By the unicity of solutions and the fact that $V_r^T(c,W)$ is deterministic, it follows that $(V_t^T(c,W), 0)$ is the solution of the ODE

$$V_t^T(c,W) = V_r^T(c,W) + \int_t^r f(c_s, V_s^T(c,W))\, ds,$$

for $0 \leq t \leq r$, proving (a). Because a corresponding representation is valid for c', (b) follows by the Comparison Theorem (restricted to ODEs).
(c) Let $r \in [0,T]$ and adopt the other notation above. Then $(\overline{W}_t)_{0 \leq t \leq T-r}$ is $\{\overline{\mathcal{F}}_t\}_{0 \leq t \leq T-r}$-Brownian motion under $P(\cdot \mid \sigma(W_r))$. Because c_t is $\sigma(W_t)$-measurable, $c_t = g(W_t)$ for a suitable function g and thus $c_{t+r} = g(W_{t+r} - W_r + W_r)$ is $\overline{\mathcal{F}}_t$-measurable relative to the probability space $(\Omega, \overline{\mathcal{F}}_{T-r}, \{\overline{\mathcal{F}}_t\}, P(\cdot \mid \sigma(W_r))$. By arguing as above, we can show that $V_r(c)$ is deterministic relative to this probability space, implying that it is $\sigma(W_r)$-measurable. ■

Proof of Lemma 4.6: (a) Theorem 2.3(b) delivers Q^* in $\mathcal{P}^\Theta$ such that $V_t(c) = V^{Q^*}(c)$. Therefore, for any other c',

$$V(c') - V(c) = V(c') - V^{Q^*}(c) =$$

$$\min_{Q \in \mathcal{P}^\Theta} V^Q(c') - V^{Q^*}(c) \leq V^{Q^*}(c') - V^{Q^*}(c)$$

$$\leq \int_0^T E_{Q^*}\left[\exp\left(\int_0^t f_v(c_s, V_s^{Q^*}(c))\, ds\right) f_c(c_t, V_t^{Q^*}(c))\,(c'_t - c_t)\right] dt$$

$$= E_P\left[\int_0^T \pi_t(c'_t - c_t)\, dt\right].$$

The second inequality is due to the nature of supergradients for the stochastic differential utility function $V^{Q^*}(\cdot)$, as established in [16].
(b): The argument is virtually identical to the proof of [24, Lemma 2.2]. ■

References

[1] A. Abel, An exploration of the effects of pessimism and doubt on asset returns, (1997), mimeo.

[2] C. Aliprantis and K. Border, *Infinite Dimensional Analysis*, Springer Verlag, 1994.

[3] E. Anderson, L.P. Hansen and T.J. Sargent, Robustness, detection and the price of risk, (2000), mimeo.

[4] P. Briand, F. Coquet, Y. Hu, J. Memin and S. Peng, A converse comparison theorem for BSDE's and related properties of g-expectation, Prep. 99-10, Inst. de Recherche Mathematique, U. Rennes, 1999.

[5] M. Cagetti, L. Hansen, T. Sargent and N. Williams, Robustness and pricing when growth is hidden, (2000), mimeo.

[6] J. Campbell, Asset prices, consumption and the business cycle, (1998), NBER 6485.

[7] R. Casadesus-Masanell, P. Klibanoff and E. Ozdenoren, Maxmin expected utility over Savage acts with a set of priors, *J. Econ. Theory* 92 (2000), 35-65.

[8] S. Cecchetti, P. Lam and N. Mark, Asset pricing with distorted beliefs: are equity returns too good to be true? (1998), mimeo.

[9] Z. Chen, Existence and uniqueness for BSDE with stopping time, *Chinese Science Bulletin* 43 (1998), 96-99.

[10] D. Cuoco and J. Cvitanic, Optimal consumption choices for a 'large' investor, *J. Econ. Dynamics and Control* 22 (1996), 401-436.

[11] C. Dellacherie, Sur l'existence de certains essinf et essup de familles de processus mesurables, *Sem. Probab. XII, Lecture Notes in Math.* vol 649, Springer Verlag, 1977.

[12] D. Duffie, *Security Markets,* Academic Press, 1988.

[13] D. Duffie, *Dynamic Asset Pricing Theory*, 2nd ed., Princeton U. Press, 1996.

[14] D. Duffie and L. Epstein, Stochastic differential utility, *Econometrica* 60 (1992), 353-394 (Appendix C with C. Skiadas).

[15] D. Duffie and L. Epstein, Asset pricing with stochastic differential utility, *Rev. Finan. Studies* 5 (1992), 411-436.

[16] D. Duffie and C. Skiadas, Continuous-time security pricing: a utility gradient approach, *J. Math. Econ.* 23 (1994), 107-131.

[17] N. Dunford and J.T. Schwartz, *Linear Operators Part I: General Theory*, Wiley, 1958.

[18] N. El Karoui, S. Peng and M. Quenez, Backward stochastic differential equations in finance, *Math. Fin.* 7 (1997), 1-71.

[19] L.G. Epstein, A definition of uncertainty aversion, *Rev. Econ. Stud.* 66 (1999), 579-608.

[20] L.G. Epstein, The global stability of efficient intertemporal allocations, *Econometrica* 55 (1987), 329-358.

[21] L.G. Epstein and J. Miao, A two-person dynamic equilibrium under ambiguity, (2000), mimeo, http://www.econ.rochester.edu/Faculty/Epstein.html.

[22] L.G. Epstein and W. Ploberger, Ambiguous Brownian motion, in progress.

[23] L.G. Epstein and T. Wang, Intertemporal asset pricing under Knightian uncertainty, *Econometrica* 62 (1994), 283-322.

[24] L.G. Epstein and T. Wang, Uncertainty, risk-neutral measures and security price booms and crashes, *J. Econ. Theory* 67 (1995), 40-80.

[25] L.G. Epstein and J. Zhang, Subjective probabilities on subjectively unambiguous events, mimeo, *Econometrica*, forthcoming.

[26] L.G. Epstein and S. Zin, Substitution, risk aversion and the temporal behavior of consumption and asset returns: a theoretical framework, *Econometrica* 57 (1989), 937-969.

[27] P. Ghirardato and M. Marinacci, Ambiguity made precise: a comparative foundation, Cal Tech #1026, 1999.

[28] I. Gilboa and D. Schmeidler, Maxmin expected utility with nonunique prior, *J. Math. Econ.* 18, (1989), 141-153.

[29] L. Hansen and R. Jagannathan, Implications of security market data for models of dynamic economies, *J. Pol. Econ.* 99 (1991), 225-262.

[30] L. Hansen and T. Sargent, Wanting robustness in macroeconomics, (2000), mimeo.

[31] L. Hansen, T. Sargent and T. Talliarini, Robust permanent income and pricing, *Rev. Ec. Stud.* 66 (1999), 873-907.

[32] I. Karatzas and E. Shreve, *Brownian Motion and Stochastic Calculus,* Springer-Verlag, 1988.

[33] J.M. Keynes, *A Treatise on Probability*, Macmillan, 1921.

[34] F.H. Knight, *Risk, Uncertainty and Profit*, Houghton Mifflin, 1921.

[35] J.P. Lepeltier and J.S. Martin, Backward stochastic differential equations with continuous coefficients, *Stat. Prob. Lett.* 32 (1997), 425-430.

[36] J. Ma, P. Protter and J. Yong, Solving forward-backward stochastic differential equations explicitly - a four step scheme, *Prob. Th. & Rel. Fields* 98 (1994), 339-359.

[37] M. Machina and D. Schmeidler, A more robust definition of subjective probability, *Econometrica* 60 (1992), 745-780.

[38] M. Marinacci, Limit laws for non-additive probabilities and their frequentist interpretation, *J. Econ. Theory* 84 (1999), 145-195.

[39] M. Marinacci, Monotone continuous multiple priors, (1998), mimeo.

[40] M. Marinacci, Probabilistic sophistication and multiple priors, (1999), mimeo.

[41] R. Mehra and E. C. Prescott, The equity premium: a puzzle, *J. Monet. Econ.* 15 (1985), 145-161.

[42] E, Pardoux, Generalized discontinuous BSDE's, in *Backward Stochastic Differential Equations*, N. El Karoui and L. Mazliak eds., Pitman Research Notes in Mathematics 364 (1997), 207-219.

[43] E. Pardoux and S. Peng, Adapted solution of a backward stochastic differential equation, *Systems Control Lett.* 14 (1990), 55-61.

[44] S. Peng, BSDE and related g–expectation, in *Backward Stochastic Differential Equations*, N. El Karoui and L. Mazliak eds., Pitman Research Notes in Mathematics 364 (1997), 141-159.

[45] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, Springer, 1999.

[46] R.T. Rockafeller, *Convex Analysis*, Princeton U. Press, 1970.

[47] D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica* 57 (1989), 571-587.

[48] M. Schroder and C. Skiadas, Optimal consumption and portfolio selection with stochastic differential utility, *J. Econ. Theory* 89 (1999), 68-126.