Econometrica, Vol. 70, No. 4 (July, 2002), 1403–1443

AMBIGUITY, RISK, AND ASSET RETURNS IN CONTINUOUS TIME

BY ZENGJING CHEN AND LARRY EPSTEIN[1]

Models of utility in stochastic continuous-time settings typically assume that beliefs are represented by a probability measure, hence ruling 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. In 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.

KEYWORDS: Ambiguity, asset pricing, backward stochastic differential equations, recursive utility, continuous-time.

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 'probabilistically sophisticated.' Because continuous-time modeling has *universally* assumed probabilistic sophistication, 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 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 (Duffie and Epstein (1992a)) whereby the usual single prior is replaced by a set of priors, as in the atemporal model of Gilboa and Schmeidler (1989). We call the resulting model *recursive multiple-priors utility*.[2]

[1] 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 benefited from discussions with Rose-Anne Dana, Jerome Detemple, 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 valuable suggestions from three very patient referees and an editor. First version: December 1998.

[2] To explain this nomenclature, note that stochastic differential utility is the continuous-time counterpart of recursive utility (Epstein and Zin (1989)).

Our model of utility is the continuous-time counterpart of that in Epstein and Wang (1994, 1995). It is well known that continuous-time modeling affords considerable analytical advantages. These 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 1.2) 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 the cited papers by Epstein and Wang. 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.

The paper proceeds as follows. The rest of this introduction elaborates on potential applications. Section 2 specifies recursive multiple-priors utility. This is accomplished in stages, beginning with an outline of the essential ingredients of the atemporal model. Section 3 provides some examples. Ambiguity and ambiguity aversion are examined in Section 4 and the application to asset pricing is provided in Section 5. Proofs are collected in appendices.

1.2. *Ambiguity in 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 (Mehra and Prescott (1985)—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, then admitting this possibility into the model may help to resolve the puzzle. This approach has been developed, with some success, in Epstein and Miao (2001).[3]

Further applications of our model are suggested by the related work on robust decision-making; see Hansen and Sargent (2000) and Anderson, Hansen, and Sargent (2000), for example. Though these authors refer to 'model uncertainty' rather than 'ambiguity,' their model is also motivated in part by the Ellsberg Paradox and it is proposed as an intertemporal version of the Gilboa-Schmeidler model. The utility function specification supporting the robust control approach is described in Hansen and Sargent (2001) and a detailed comparison of the two models of utility is provided in Epstein and Schneider (2001a).[4] In spite of the substantial differences between the models described there, the commonality in motivation and spirit suggests that the macroeconomic applications discussed by these authors (see also Hansen, Sargent, and Tallarini (1999) and Cagetti et al. (2002)) are potential applications of our model as well. These include also normative applications, for example, to optimal monetary policy in a setting where the monetary authority does not know precisely the true model describing the environment (Onatski (2000) and Onatski and Stock (2002)).

To provide some perspective on the above applications, 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 the true probability law. This approach is adopted in Abel (2002) and Cecchetti, Lam, and Mark (2000) 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. That is, there is an observational equivalence if one restricts attention to a single dynamic equilibrium. Nevertheless, our approach has several advantages.

First, the observational equivalence fails once one connects the dynamic equilibrium to behavior in other settings. For example, the equity premium 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. Such transfers are inappropriate, however, under our working hypothesis that prospects faced in an asset market are qualitatively different than hypothetical lotteries where prizes are determined by the outcome of a coin flip, for example. In this way, even though

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

[4] One difference is that the robust control model violates the usual notion of dynamic consistency. Another difference is that its underlying updating rule is such that conditional preference at time $t > 0$ depends on what might have happened in other unrealized parts of the event tree.

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, reinterpretation of the equity premium as due partly to ambiguity has potential empirical significance.

Second, 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. Finally, 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.

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 the nth draw is made from the nth urn. If the agent views the urns as 'identical and independent,' then one would *not* expect ambiguity to vanish. Indeed, Marinacci (1999) and Epstein and Schneider (2001b) prove LLN results 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*. The latter paper adopts the discrete-time counterpart of recursive multiple-priors utility and thus is directly relevant. 'Identical and independent' is modeled there by conditional one-step-ahead beliefs that are independent of history and time. The continuous-time counterpart corresponds to the special case of our model called IID ambiguity (Section 3.4), where, roughly speaking, the increments $\{dW_t : t \geq 0\}$ of the driving state process (W_t) constitute the counterpart of the set of Ellsberg urns. (See the end of Section 2.4 for further discussion of learning.)

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:

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

where $u\colon \mathscr{X} \to \mathscr{R}^1$ is a von Neumann-Morgenstern utility index and $\mathscr{P}$ is a subjective *set* of probability measures on $(\Omega, \mathscr{F})$.[5] The subjective expected utility model is obtained when the set of priors $\mathscr{P}$ is a singleton. Intuitively, the multiplicity of priors in the general case models ambiguity about 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 $\mathscr{P}$ are uniformly absolutely continuous with respect to some P in $\mathscr{P}$.[6] Then $\mathscr{P}$ may be identified with its set H of densities with respect to P, where $H \subset L^1_+(\Omega, \mathscr{F}, P)$ is weakly compact. The identification is via

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

For further details and behavioral implications of this added structure, see Epstein and Wang (1995, Section 2).

2.2. *Discrete Time*

The essence of our continuous-time model can be described by considering first a discrete-time setting.[7]

Let time vary over $t = 0, \dots, T$ and let the state space and filtration be given by $(\Omega, \{\mathscr{F}_t\}_0^T)$. The objects of choice are adapted consumption processes. To formulate a dynamic version of the multiple-priors model, it is natural to consider the process of conditional preferences and furthermore to assume that each such conditional preference satisfies the Gilboa-Schmeidler axioms (where the outcome set $\mathscr{X}$ consists of consumption streams). Suppose further that conditional preferences are dynamically consistent. Epstein and Schneider (2001a) show that these axioms, plus some 'auxiliary' ones, deliver the following representation:[8] the time t conditional utility of a consumption process $c = (c_t)$ is

$$V_t(c) = \min_{Q \in \mathscr{P}} E_Q\left[\sum_{s=t}^{T} \beta^{s-t} u(c_s) \mid \mathscr{F}_t\right], \tag{2.2}$$

where β and u are as usual and where $\mathscr{P}$ is the agent's set of priors on $(\Omega, \mathscr{F}_T)$. The set $\mathscr{P}$ satisfies the regularity conditions described above for the atemporal

[5] The set $\mathscr{P}$ is required to be weakly compact (the weak topology is that induced by the set of bounded measurable functions) and convex. Because $\mathscr{P}$ and its closed convex hull generate the identical utility function, closedness and convexity are normalizations that ensure uniqueness. See Gilboa and Schmeidler (1989) for further details. Note, however, that probability measures are assumed there to be only finitely additive, while we assume countable additivity.

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

[7] This perspective on Epstein and Wang (1994) does not appear there; it has emerged only with the benefit of hindsight and Epstein and Schneider (2001a).

[8] As in Gilboa and Schmeidler (1989), an Anscombe-Aumann style domain is adopted where the axioms are both necessary and sufficient for the representation. The auxiliary axioms deal primarily with the ranking of objective lotteries rather than ambiguous prospects.

model and also a property that (following earlier versions of this paper) is called rectangularity. Because of rectangularity, utilities satisfy the recursive relation

$$(2.3) \qquad V_t(c) = \min_{Q\in\mathcal{P}} E_Q\left[\sum_{s=t}^{\tau-1}\beta^{s-t}u(c_s) + \beta^{\tau-t}V_\tau(c) \mid \mathcal{F}_t\right]$$

for all $\tau > t$, which in turn delivers dynamic consistency if $\mathcal{P}$ is updated by applying Bayes' Rule prior by prior. It merits emphasis that such a recursive relation is not valid for general sets of priors because the minimization destroys the additivity available in the standard model.

To understand the meaning of rectangularity, observe that the recursive relation for utility depends on $\mathcal{P}$ only via the sets of one-step-ahead conditional measures that it induces at each (t, ω). Thus it must be that in a suitable sense $\mathcal{P}$ is completely determined by these one-step-ahead conditionals. To see how, think of a discrete-time event tree that represents $(\Omega, \{\mathcal{F}_t\})$, where nature determines motion through the tree and where $\mathcal{F}_T$ describes the set of terminal states or events. At each node, $\mathcal{P}$ induces a set of conditional probability measures over the state next period. Conversely, the sets of conditional-one-step-ahead measures for all time-event 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-event pair. In general, $\mathcal{P}'$ is strictly larger than $\mathcal{P}$ though they induce the identical sets of one-step-ahead measures. Call $\mathcal{P}$ rectangular if $\mathcal{P}' = \mathcal{P}$. It is apparent that rectangularity ensures an equivalence between global minimization over $\mathcal{P}$, as in (2.2), and repeated local minimization over the set of one-step-ahead conditional measures, as in (2.3).

For later use, note that if a reference probability measure P is given on $\mathcal{F}_T$, then conditional measures can be expressed in terms of their densities with respect to the conditional measures induced by P. Thus the preceding sketch can be reformulated in terms of sets of one-step-ahead densities. Further, these can be taken to be primitives and specified arbitrarily. Then the above construction delivers a rectangular set $\mathcal{P}$ and *any* rectangular set can be generated in this way.

A (nonaxiomatic) generalization of (2.2) has the form

$$(2.4) \qquad V_t(c) = \min_{Q\in\mathcal{P}} V_t^Q(c), \qquad \text{where}$$

$$V_t^Q(c) = W\big(c_t, E_Q[V_{t+1}^Q(c) \mid \mathcal{F}_t]\big) \qquad \text{for each } Q \text{ in } \mathcal{P}.$$

Here W is an aggregator function (strictly increasing in its second argument) analogous to that appearing in Epstein and Zin (1989) and motivated there by the desire to disentangle risk aversion from other aspects of preference. As above, rectangularity for $\mathcal{P}$ delivers the recursive relation

$$(2.5) \qquad V_t(c) = W\Big(c_t, \min_{Q\in\mathcal{P}} E_Q[V_{t+1}(c) \mid \mathcal{F}_t]\Big) = \min_{Q\in\mathcal{P}} W(c_t, E_Q[V_{t+1}(c) \mid \mathcal{F}_t]).$$

We proceed shortly to formulate a continuous-time counterpart of (2.4)–(2.5). The key is the construction of rectangular sets of priors in continuous time, that is, sets constructed along the lines described above for the event tree.

2.3. *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 $W_t = (W_t^1, \ldots, W_T^d)^\top$ defined on $(\Omega, \mathcal{F}, P)$;
- the Brownian filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$, where $\mathcal{F}_t$ is generated by $\sigma(W_s\colon s \leq t)$ 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 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.[9] 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 (1992a) 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

$$V_t^P = E\left[\int_t^T f(c_s, V_s^P)\, ds \mid \mathcal{F}_t\right]. \tag{2.6}$$

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

$$V_t = E\left[\int_t^T e^{-\beta(s-t)} u(c_s)\, ds \mid \mathcal{F}_t\right]. \tag{2.7}$$

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.

2.4. *The Set of Priors*

As suggested in the discussion of the discrete-time model, construction of the set $\mathcal{P}$ of priors on $(\Omega, \mathcal{F}_T)$ is key. Ignore rectangularity for the moment and

[9] In this paper, $x = (x_t)$ denotes a *process*, by which we mean that it is: (i) progressively measurable, that is, (for each t) $x : [0, t] \times (\Omega, \mathcal{F}_t) \to R^\ell$ 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.

consider the representation of sets $\mathcal{P}$ of measures equivalent to P. This is done by specifying suitable densities.

For this purpose, define a *density generator* to be an R^d-valued process $\theta = (\theta_t)$ for which the process (z_t^θ) is a P-martingale, where

$$dz_t^\theta = -z_t^\theta \theta_t \cdot dW_t, \qquad z_0^\theta = 1,$$

that is,

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

A sufficient condition (see Duffie (1996, p. 288)) is that θ satisfy the Novikov condition

$$E\left[\exp\left(\frac{1}{2}\int_0^T |\theta_s|^2\, ds\right)\right] < \infty. \tag{2.8}$$

Then, because $1 = z_0^\theta = E[z_T^\theta]$, 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

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

In other words,

$$\frac{dQ^\theta}{dP} = z_T^\theta; \tag{2.9}$$

more generally,

$$\left.\frac{dQ^\theta}{dP}\right|_{\mathcal{F}_t} = z_t^\theta \qquad \text{for each } t.$$

Thus, given a set Θ of density generators, the corresponding set of priors is

$$\mathcal{P}^\Theta = \{Q^\theta\colon \theta \in \Theta \text{ and } Q^\theta \text{ is defined by (2.9)}\}. \tag{2.10}$$

Conversely, any set of equivalent measures can be generated in this way.[10]

Turn now to a further restriction on sets of density generators and hence sets of priors in order to obtain recursivity of utility. The discussion of rectangularity in the discrete-time setting pointed to the key property being that $\mathcal{P}$ is equal to the set of all measures that can be constructed via arbitrary selections from primitive sets of one-step-ahead densities. In the present setting, a density generator $\theta = (\theta_t)$ is the process that delivers the counterpart of the (logarithm of) a conditional one-step-ahead density for each time and state and the primitive

[10] Adapt the argument in Duffie (1996, p. 289).

sets of one-step-ahead densities (in logarithm form) are modeled via a process $(\Theta_t)_{t\in[0,T]}$ of correspondences from Ω into R^d; that is, for each t, let

$$\Theta_t\colon \Omega \rightsquigarrow R^d.$$

Finally, the restriction to the set of all measures that can be constructed by some selection from these sets of one-step-ahead densities corresponds to the restriction to the following set of density generators:

$$\Theta = \{(\theta_t)\colon \theta_t(\omega) \in \Theta_t(\omega)\, dt \otimes dP \text{ a.e.}\}. \tag{2.11}$$

Refer to such sets of density generators and to the corresponding sets of priors $\mathcal{P}^{\Theta}$ as *rectangular*.[11]

Assume throughout the following properties for $(\Theta_t)_{t\in[0,T]}$:

UNIFORM BOUNDEDNESS: *There is a compact subset $\mathcal{K}$ in R^d such that $\Theta_t\colon \Omega \rightsquigarrow \mathcal{K}$ each t.*

COMPACT-CONVEX: *Each Θ_t is compact-valued and convex-valued.*

MEASURABILITY: *The correspondence $(t,\omega) \mapsto \Theta_t(\omega)$, 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.

Uniform Boundedness ensures that (2.8) is satisfied by any $\theta \in \Theta$ and hence that each Q^{θ} is well-defined. Normalization ensures that the reference measure P lies in $\mathcal{P}^{\Theta}$. The roles of the other assumptions are evident.

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

$$e_t(x)(\omega) = \max_{y\in\Theta_t(\omega)} y\cdot x, \qquad x \in R^d. \tag{2.12}$$

Occasionally, we suppress the state and write simply $e_t(x)$. 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).[13] Further, by Aliprantis and Border (1994, Theorem 14.96), the above Measurability assumption is equivalent to:

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

[11] An example of a nonrectangular set is $\{(\theta)_t\colon E[\int_0^T |\theta_s|^2\, ds] \leq \ell\}$, where $\ell \geq 0$ is a parameter.

[12] That is, $\{(t,\omega) \in [0,s]\times\Omega\colon \Theta_t(\omega)\cap K \neq \varnothing\} \in \mathcal{B}([0,s])\times\mathcal{F}_s$ for each compact $K \subset \mathcal{K}$ (Aliprantis and Border (1994, Section 14.12)).

[13] The proof of Lipschitz continuity is contained in the proof of Theorem 2.2.

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.

The above assumptions on $(\Theta_t)_{t\in[0,T]}$ deliver a number of properties for the set of priors. The most important implication of rectangularity, that is, of the special structure (2.11) for Θ, is recursivity of utility as described in the next section. Here we list properties that are counterparts of those mentioned in the context of the atemporal setting.

THEOREM 2.1: *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*.[14]

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

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

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 $L^2(Q, \mathcal{F}_T, P)$. Part (e) extends the existence of a minimum to the process of conditional expectations.

Finally, consider again the issue of learning. As suggested above, $\Theta_t(\omega)$ 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 (Section 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 in part acknowledging the widely recognized fact that there is no decision theory available that serves to pin down the prior.

[14] 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.

2.5. *Definition and Existence of Utility*

Let Θ and $\mathcal{P}^{\Theta}$ be as above. In addition and in common with SDU (see (2.6)), another primitive component of the specification of utility is an aggregator $f\colon C \times R^1 \to R^1$.[15] Assume the following:

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

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

- Growth condition in consumption: $E[\int_0^T f^2(c_t,0)\,dt] < \infty$ 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 Duffie and Epstein (1992a)) to

$$V_t^Q = E_Q\left[\int_t^T f(c_s, V_s^Q)\,ds \mid \mathcal{F}_t\right], \qquad 0 \le t \le T. \tag{2.13}$$

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

$$V_t = \min_{Q \in \mathcal{P}^{\Theta}} V_t^Q, \qquad 0 \le t \le T. \tag{2.14}$$

We show shortly that (2.14) 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 $(V_t(c))$.

The definition (2.14) is the continuous-time counterpart of (2.4). One would expect, therefore, that if our construction of $\mathcal{P}^{\Theta}$ captures the appropriate notion of rectangularity, then we should obtain dynamic consistency of $(V_t(\cdot))$ by establishing a counterpart of (2.5).[16] This is achieved in the theorem to follow.

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 for utilities in differential terms. Accordingly, the theorem shows that the utility process defined by (2.14) can be characterized alternatively as the unique solution to a backward stochastic differential equation (BSDE).[17]

[15] In the terminology of Duffie and Epstein (1992a), f is a normalized aggregator. The transformations of f that lead to ordinally equivalent utility processes below are identical to those described in the cited paper.

[16] Dynamic consistency, defined as in Duffie and Epstein (1992a, p. 373), is the requirement that for all stopping times τ and all consumption processes c and c' satisfying $c' = c$ on $[0,\tau]$, $P(V_\tau(c') \ge V_\tau(c)) = 1 \Rightarrow V_0(c') \ge V_0(c)$, with strict inequality holding if $P(V_\tau(c') > V_\tau(c)) > 0$.

[17] See Appendix A for a brief outline and El Karoui, Peng, and Quenez (1997) for a comprehensive guide to the theory of BSDE's as well as to previous applications to utility theory and derivative security pricing.

To illustrate, notice that the SDU process (V_t^P) defined by (2.6) can be expressed alternatively as the unique solution to the BSDE

$$dV_t^P = -f(c_t, V_t^P)\,dt + \sigma_t^P \cdot dW_t, \qquad V_T^P = 0. \tag{2.15}$$

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 throughout. To see that the BSDE characterization follows from (2.6), observe that, by the latter,

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

which is a martingale under P. Thus the Martingale Representation Theorem delivers (2.15) for a suitable process (σ_t^P) (that depends on c). This argument may be reversed by using the fact that $\int_0^t \sigma_t^P \cdot dW_t$ is a martingale in order to establish that (2.15) implies (2.6).

A similar reformulation is possible for the SDU process (V_t^Q) defined in (2.13) and corresponding to an agent with probabilistic beliefs given by Q in $\mathcal{P}^\Theta$. If $Q = Q^\theta$ (see (2.9)), then the Girsanov Theorem implies that (V_t^Q) solves the BSDE,

$$dV_t^Q = [-f(c_t, V_t^Q) + \theta_t \cdot \sigma_t^Q]\,dt + \sigma_t^Q \cdot dW_t, \qquad V_T^Q = 0. \tag{2.16}$$

In comparison with (2.15), 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.[18]

We are now ready to state our main theorem.

THEOREM 2.2: *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*

$$dV_t = \left[-f(c_t, V_t) + \max_{\theta \in \Theta} \theta_t \cdot \sigma_t\right] dt + \sigma_t \cdot dW_t, \qquad V_T = 0. \tag{2.17}$$

(b) *For each $Q = Q^\theta \in \mathcal{P}^\theta$, denote by (V_t^Q) the unique solution to* (2.13), *or equivalently to* (2.16). *Then (V_t) defined in (a) is the unique solution to* (2.14) *and there exists $Q^\theta \in \mathcal{P}^\Theta$ such that*

$$V_t = V_t^{Q^{\theta^*}}, \qquad 0 \le t \le T. \tag{2.18}$$

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

$$V_t = \min_{Q \in \mathcal{P}^\Theta} E_Q\left[\int_t^\tau f(c_s, V_s)\,ds + V_\tau \mid \mathcal{F}_t\right], \qquad 0 \le t < \tau \le T. \tag{2.19}$$

[18] The Girsanov Theorem and the Martingale Representation Theorem are the key tools that we employ from stochastic calculus. They are standard in finance—see Duffie (1996), for example.

Part (b) refers to the initial definition (2.14). Part (a) is the BSDE characterization that is the counterpart to (2.15). Part (c) makes explicit the recursivity of utility and justifies the name *recursive multiple-priors utility* for our model of utility. Equation (2.19) is the promised counterpart of the discrete-time relation (2.5).

Comparison of (2.17) and (2.14) yields some insight into our construction. If the volatility of utility were denoted by $-\sigma_t$ rather than σ_t, then the maximum in (2.17) would be replaced by a minimum, paralleling (2.14). 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, just as in the discrete-time setting. This equivalence is due to our construction of Θ via (2.11) as rectangular. It is easy to understand the importance of (2.11). By (B.1), the maximum in (2.17) 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 than does the global optimization problem in (2.14), 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 (2.11) imposes that Θ is the Cartesian product of its projections.

Turn to interpretation, particularly the nature of the ambiguity modeled via recursive multiple-priors utility; a more formal treatment of ambiguity is provided in Section 4.2. Only one of the measures in $\mathcal{P}^{\Theta}$, namely P, makes the driving process (W_t) a Brownian motion. Thus there is ambiguity about whether (W_t) is a Brownian motion. More specifically, Girsanov's Theorem implies that if $Q = Q^{\theta}$ is in $\mathcal{P}^{\Theta}$, then $W_t^Q \equiv W_t + \int_0^t \theta_s \, ds$ is a Brownian motion under Q. Thus ambiguity concerns (and is limited to) the drift of the driving process. The fact that ambiguity is limited to the drift is a consequence of the Brownian environment and the assumption of absolute continuity.

Several extensions of the theorem seem possible. The assumption of a Brownian filtration can be relaxed along the lines indicated in El Karoui, Peng, and Quenez (1997). The terminal value of 0 in (2.17) can be generalized and utility can be well-defined without the Lipschitz hypothesis (Lepeltier and Martin (1997)). 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 Duffie and Epstein (1992a) for stochastic differential utility. Related results for BSDE's defined on an infinite horizon may be found in Chen (1998) and Pardoux (1997).

Finally, note that BSDE's have been used to price securities in markets that feature incompleteness, short-sale constraints, or other imperfections.[19] These lead to nonlinear BSDE's characterizing (upper or lower) prices that are formally very similar to the BSDE (2.17) 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,

[19] See El Karoui, Peng, and Quenez (1997) for some references.

the multiplicity of measures arises at the level of utility and is due to ambiguity rather than features of the market.

The coming sections illustrate, interpret, and apply the recursive multiple-priors model of utility.

3. EXAMPLES

3.1. *Deterministic and Risky Consumption Processes*

It is important to keep in mind that σ_t is endogenous in the BSDE (2.17). To illustrate this endogeneity and the consequent dependence of σ_t on the consumption process, consider (2.17) for two particular consumption processes. First, suppose that c is deterministic. Then $\sigma_t = 0$ and utility is given by the ordinary differential equation

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

This is the recursive utility model for deterministic consumption processes proposed in Epstein (1987).

For the second example, let

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

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

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 (2.17). 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 (2.15) is the solution to (2.17). 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 (2.7) is

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

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 2.2(b), it is enough to have a representation for V_t^Q for each Q in $\mathcal{P}^\Theta$. However, from (2.13),

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

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.

While this functional form may seem the 'obvious' way to formulate a multiple-priors extension of the usual model (2.7), the subtlety is the rectangularity of $\mathcal{P}^\Theta$, which, as explained above, is responsible for recursivity. The latter takes the form (by Theorem 2.2(c))

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

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

3.3. κ-*Ignorance*

Fix a parameter $\kappa = (\kappa_1, \ldots, \kappa_d)$ in R_+^d and take

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

Then

$$\Theta = \{(\theta_t) : \sup\{|\theta_t^i| : 0 \leq t \leq T\} \leq \kappa_i, i = 1, \ldots, d\}.$$

The following notation will be useful. Denote by $|\sigma_t|$ the d-dimensional vector with ith component $|\sigma_t^i|$, and similarly for other d-dimensional vectors. Define

$$\operatorname{sgn}(x) \equiv \begin{cases} |x|/x & \text{if } x \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3.5}$$

and $\kappa \otimes \operatorname{sgn}(\sigma_t) \equiv (\kappa_1 \operatorname{sgn}(\sigma_t^1), \ldots, \kappa_d \operatorname{sgn}(\sigma_t^d))$.

Then

$$\max_{\theta \in \Theta} \theta_t \cdot \sigma_t = \theta_t^* \cdot \sigma_t = \kappa \cdot |\sigma_t|,$$

where

$$\theta_t^* = \kappa \otimes \operatorname{sgn}(\sigma_t), \quad \text{or} \quad \theta_t^{*i} = \begin{cases} \kappa_i |\sigma_t^i| / \sigma_t^i & \text{if } \sigma_t^i \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

Consequently, the utility process solves

$$dV_t = [-f(c_t, V_t) + \kappa \cdot |\sigma_t|]\, dt + \sigma_t \cdot dW_t, \quad V_T = 0. \tag{3.7}$$

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 SDU. We interpret the term $\kappa \cdot |\sigma_t|$ as modeling ambiguity aversion rather than risk aversion (see Section 4.2). For example, in the two-dimensional case, $\kappa_1 = 0$ and $\kappa_2 > 0$ indicate that W^2 is ambiguous but W^1 is purely risky. Further interpretation of the κ-ignorance specification is provided in the next section.

3.4. *IID Ambiguity*

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

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

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 suggested 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 shortly.

The utility process generated by this specification for (Θ_t) solves

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

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

$$e(x) = \max_{y \in K} y \cdot x, \qquad x \in R^d, \tag{3.8}$$

corresponding to a special case of (2.12) where the support function is independent of both time and the state.

By the theory of support functions (Rockafeller (1970)), the process θ^* asserted by Theorem 2.2(b) is given by $\theta^*_t \in \partial e(\sigma_t)$ for every t, where $\partial e(x)$ denotes the set of subgradients of e at x.

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

$$0_{d_1} \in \operatorname{int}(K_1) \subset R^{d_1}. \tag{3.9}$$

In words, for those process $(W^i_t)_{i \in d_1}$ 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 $x \in R^d$,[20]

$$e(x) = 0 \Longrightarrow x_i = 0 \quad \text{for all } i \text{ such that } K_i \ne \{0\}. \tag{3.10}$$

The assumption (3.9) is included in any reference below to IID ambiguity.

[20] $e(x) = 0$ iff $y \cdot x \le 0$ for all $y \in K$. Suppose there exists i such that $K^i \ne \{0\}$. (Otherwise, (3.10) is obvious.) Then it follows from (3.9) that $x_i = 0$. The reverse implication in (3.10) is evidently also true. Alternatively, (3.9) is equivalent to the assumption that the polar of K is $\{0\}$.

The special case

$$K = \{y \in R^d : |y_i| \leq \kappa_i \text{ all } i\} \tag{3.11}$$

delivers κ-ignorance. An alternative special case has

$$K = \left\{ y \in R^d : \sum_{\{i:\kappa_i \neq 0\}} \kappa_i^{-1} |y_i|^2 \leq 1 \right\}, \tag{3.12}$$

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

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

$$dc_t/c_t = \mu^c\, dt + s^c\, dW_t, \tag{3.13}$$

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

$$f(c, v) = \frac{c^\rho - \beta(\alpha v)^{\rho/\alpha}}{\rho(\alpha v)^{(\rho-\alpha)/\alpha}}, \tag{3.14}$$

for some $\beta \geq 0$ and nonzero ρ, $\alpha \leq 1$.[21] This is the continuous-time version of the so-called Kreps-Porteus functional form (Duffie and Epstein (1992a, 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

$$A_t^{\rho/\alpha} = \lambda^{-1}(1 - e^{\lambda(t-T)}),$$
$$\lambda = \beta - \rho(\mu^c - (1-\alpha)s^c \cdot s^c/2 - e(s^c)).$$

The associated volatility is

$$\sigma_t = A_t c_t^\alpha s^c. \tag{3.15}$$

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 4.2 from the interpretation

[21] This aggregator violates the Lipschitz condition for Theorem 2.2 and thus existence of utility is not ensured. See further discussion in Section 5.4.

provided there for α and $e(\cdot)$.[22] 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 $\kappa \cdot |s^c|$ in the case of κ-ignorance.

Finally, we clarify the meaning of IID ambiguity and κ-ignorance by describing properties of the utility processes that they deliver. Because (W_t) is a P-Brownian motion, P induces the following properties: (i) (W_t) is Markovian with identically and independently distributed increments; (ii) increments are contemporaneously independent and (iii) normally distributed with the familiar means and variances. A natural question is which of these properties, *suitably reformulated*, survive under $\mathcal{P}^\Theta$, that is, in spite of ambiguity. The next theorem shows that (i) survives under IID ambiguity, while (ii) is also valid under κ-ignorance. Thus in the latter case, $\mathcal{P}^\Theta$ models ambiguity about (iii) alone.

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 2.2.

THEOREM 3.1: *Suppose that Θ is given by IID ambiguity. For each r in $[0, T]$, let $\mathcal{G}_t^r = \sigma(W_s - W_r\colon t \geq s \geq r)$ for $t \geq r$ and $= \{\varnothing, \Omega\}$ 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*

$$V_0^T(c') \geq V_0^T(c) \Longleftrightarrow V_0^{T-r}({}^r c') \geq V_0^{T-r}({}^r c). \tag{3.16}$$

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

(d) *If Θ is given by κ-ignorance and if c is adapted to the filtration $\{\mathcal{F}_t^{\leq \ell}\}$, where $\ell \leq d$ and $\mathcal{F}_t^{\leq \ell} = \sigma(W_s^i\colon s \leq t, i \leq \ell)$, then $(V_t(c))$ is also adapted to $\{\mathcal{F}_t^{\leq \ell}\}$.*

For interpretation, consider each of these statements when $\Theta = \{0\}$ and thus when 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.

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 under IID ambiguity.

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.

[22] For further interpretation of $e(\cdot)$, see Lemma A.2.

For a general IID model, $V_t(c)$ can depend on W_t^2 even if c is adapted to $\sigma(W_s^1\colon s \leq t)$; this may happen because of a contemporaneous dependence between components of W_t. Part (d) states that this is impossible, however, given κ-ignorance. Thus (d) expresses a form of contemporaneous independence between components of the driving process. This reflects the fact that under κ-ignorance, the set $\{\theta_t^i \in R^d : |\theta_t^i| < \kappa_i\}$ of admissible distortions in coordinate i is independent of the distortions in other coordinates.

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) and (d).

4. AMBIGUITY

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 Duffie and Epstein (1992a) or El Karoui, Peng, and Quenez (1997, Prop. 3.5). As noted prior to Theorem 2.2, dynamic consistency is an immediate consequence of the recursive construction of utility via (2.17).[23]

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

4.1. *Behaviorally Distinct*

This subsection supports earlier claims that probabilistic sophistication (suitably defined) provides a behavioral distinction between recursive multiple-priors utility and *all* other continuous time intertemporal utility functions in the current literature.[24] Probabilistic sophistication implies indifference to ambiguity, both informally in that it is contradicted by Ellsberg-style behavior, the canonical illustration of nonindifference to ambiguity, and also on formal grounds (Epstein and Zhang (2001) and Ghirardato and Marinacci (2002)). Thus only recursive multiple-priors can accommodate a concern with ambiguity.

The formulation of our result is complicated by the fact that the Machina-Schmeidler (1992) notion of probabilistic sophistication is not appropriate in a dynamic setting. It requires 'primarily' that there exists a probability measure $\overline{P}$ on $(\Omega, \mathcal{F}_T)$ such that the utility of any c depends only on the probability distribution induced by c: $\Omega \to R^{\ell[0,T]}$ and $\overline{P}$. 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 the sense of Machina and Schmeidler

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

[24] These include, for example, SDU and utility functions with intertemporal nonseparabilities due to habit formation or learning-by-doing.

only in the special case of the standard intertemporally additive expected utility function. Thus we describe a variation of probabilistic sophistication that excludes such extraneous restrictions and isolates the property that preference is based on probabilities.

Denote by $D_t \subset D$ the set of consumption processes c such that (i) c_τ is deterministic for $0 \le \tau < t$ and (ii) c_τ is $\mathcal{F}_t$-measurable for each $t \le \tau \le T$. Processes in D_t are such that all uncertainty is resolved at the single instant t and thus we refer to elements in $\bigcup_{t=0}^{T} D_t$ as *timeless prospects*. Call the utility function V: $D \to R^1$ *probabilistically sophisticated for timeless prospects* if V restricted to $\bigcup_{t=0}^{T} D_t$ is probabilistically sophisticated in the sense of Machina and Schmeidler.

The Machina-Schmeidler axiomatization 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.[25]

THEOREM 4.1: *The recursive multiple-priors utility function V defined in Theorem 2.2 is probabilistically sophisticated for timeless prospects if and only if it conforms to SDU.*

It is well known that in an atemporal setting, there exist probabilistically sophisticated multiple-priors preferences where priors do not agree on all events. Thus, Marinacci (2000) establishes such global agreement only under the supplementary assumption that there exists an 'interior' event where all measures in the set of priors agree. It is noteworthy that in our setting where the set of priors is rectangular, corresponding to dynamic consistency of preference, no supplementary assumption is needed.

4.2. *Ambiguity, Ambiguity Aversion, and Risk Aversion*

We attempt now to treat ambiguity and ambiguity aversion more formally.

At a formal level, ambiguity (or unambiguity) is most naturally defined as a property of events. Identify the class $\mathcal{U}$ of unambiguous events as consisting of those events where all measures in $\mathcal{P}^\Theta$ agree, that is,

$$\mathcal{U} = \{B \in \mathcal{F}_T : Q(B) = P(B) \text{ for all } Q \text{ in } \mathcal{P}^\Theta\}. \tag{4.1}$$

Call all other events ambiguous.

Foundations for this identification are provided by the two behavioral or preference-based definitions of ambiguity in the literature, namely Ghirardato

[25] Continuous time is not important for this result or for Theorem 4.4 below. Similar results can be proven in the discrete-time model of Epstein and Wang (1994) assuming nonatomic priors, though that was not apparent at the time that paper was written. We now have the benefit of Marinacci (2002) as well as of recent advances (Epstein and Zhang (2001) and Ghirardato and Marinacci (2002), for example) in understanding the behavioral meaning of ambiguity.

and Marinacci (2002) and Epstein and Zhang (2001). For the former definition, it is immediate that (4.1) characterizes unambiguous events. For the latter definition, the characterization (4.1) is valid for 'most' events under the assumption of IID ambiguity (see Lemma E.1).[26]

To obtain a further characterization in terms of the primitive density generators, let

$$\Theta^i = \{\theta^i = (\theta^i_t) \colon \theta \in \Theta\} \qquad (i = 1, \ldots, d).$$

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

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,*

$$(4.2) \qquad \Theta^i = \{0\} \quad \text{or} \quad P(F \mid \mathcal{F}^i_T) = 0 \quad \text{or} \quad P(F \mid \mathcal{F}^i_T) = 1.$$

Consequently,

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

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

Given the preceding designation of unambiguous events, we adopt the approach advocated in Epstein (1999) and Epstein and Zhang (2001) to define the distinct notions of ambiguity aversion and risk aversion. Roughly, the approach is to identify consumption processes that are adapted to $\{\mathcal{U} \cap \mathcal{F}_t\}$ as the unambiguous processes. 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.

To proceed more precisely, 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.

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

$$(4.3) \qquad \mathcal{U} \supset \mathcal{U}^* \quad \text{and}$$

$$(4.4) \qquad V(c^{ua}) \geq (>) V(c) \Longrightarrow V^*(c^{ua}) \geq (>) V^*(c),$$

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

[26] We suspect, but have not been able to prove, that (4.1) characterizes all unambiguous events.

unambiguous for V, then so should the more ambiguity averse V^*. The weak nesting condition (4.3) ensures that the more ambiguity averse decision-maker views more events as ambiguous.

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

$$\mathcal{U} \subset \mathcal{U}^* \tag{4.5}$$

and

$$V(\bar{c}) \geq (>) V(c^{ua}) \Longrightarrow V^*(\bar{c}) \geq (>) V^*(c^{ua}), \tag{4.6}$$

for all V-unambiguous consumption processes c^{ua} and deterministic processes $\bar{c}$. Symmetric with the prior definition, the more (risk) averse agent is assumed, via (4.5), 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 $\bar{c}$, should be disliked also by the more risk averse V^*.

Consider an extreme case where

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

that is, all nontrivial events are ambiguous according to V^*. Then V^* is more risk averse than V if and only if $\mathcal{U} = \{\varnothing, \Omega\}$ 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

$$V(\bar{c}) \geq (>) V(c) \Longrightarrow V^*(\bar{c}) \geq (>) V^*(c),$$

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

The definitions are best clarified by application to canonical functional forms—Kreps-Porteus aggregators (3.14) and κ-ignorance for Θ. The utility function V generated by any 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.[27]

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

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* $\kappa^* > \kappa$. *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 process.[28]

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*

$$f = f^* \quad \text{and} \quad K^* \supset K. \tag{4.7}$$

The converse is true if $K^{*i} = \{0\}$ *for some* i.

(ii) (f^*, K^*) *is more risk averse than* (f, K) *if*

$$f^*(c, h(v)) = h'(v) f(c, v), \tag{4.8}$$

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 Sections 3.3 and 5.6 of Duffie and Epstein (1992a) in order to deal with (4.8). We refer the reader to the just cited paper for clarification of (4.8) and for an alternative to (4.8) that is more intuitive (but too involved to include here). We merely note that the transformation in (4.8) 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.

The converse in (ii) is not true in general because of the presumption in (4.8) 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 implies 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 (5.22).

[28] 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.

5. ASSET RETURNS

5.1. *Supergradients*

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

$$V(c') - V(c) \leq E\left[\int_0^T \pi_t \cdot (c'_t - c_t)\, dt\right], \tag{5.1}$$

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 2.2(b)), we can use a suitable envelope theorem to relate $\partial V(c)$ to supergradients of $\{V^Q\colon Q \in \mathcal{P}^\Theta\}$. For each SDU function V^Q, the set of supergradients may be completely characterized, following Duffie and Skiadas (1994), under the added assumption that there exists $k > 0$ such that

$$\sup(|f_c(x, v)|, |\, f(x, 0)\, |) < k(1 + |\, x\,|), \qquad \text{for all } (x, v) \in C \times R^1. \tag{5.2}$$

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 (2.17); recall (B.1).

LEMMA 5.1: *Suppose that f is continuously differentiable and that it satisfies* (5.2) *and the assumptions of Theorem* 2.2. *Then*: (a)

$$\begin{aligned} \partial V(c) &\supset \Pi \\ &\equiv \left\{\pi : \exists\, \theta^* \in \Theta_c,\ \pi_t = \exp\left(\int_0^t f_v(c_s, V_s(c))\, ds\right) f_c(c_t, V_t(c)) z_t^{\theta^*} \text{all } t\right\}. \end{aligned} \tag{5.3}$$

(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 D for a proof. The set Π is alternatively expressed as

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

the set of of supergradients for the SDU functions V^{Q^*} where Q^* satisfies (2.18). 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 (5.3) 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.

5.2. *The Optimization Problem*

Consider a consumer with recursive multiple-priors utility. Her environment is standard (see Duffie (1996) 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

$$-d\pi_t/\pi_t = r_t dt + \eta_t \cdot dW_t, \quad \pi_0 = \mathbf{1}, \tag{5.4}$$

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

$$dX_t = ([r_t + \psi_t^\top (b_t - r_t\mathbf{1})]X_t - c_t)\, dt + X_t \psi_t^\top s_t dW_t, \quad X_0 > 0 \quad \text{given.} \tag{5.5}$$

Budget feasible consumption processes may be characterized by the inequality

$$E\left[\int_0^T \pi_t c_t\, dt\right] \le X_0. \tag{5.6}$$

First-order conditions for optimal consumption choice are expressed in the usual way in terms of the supergradient of utility at the optimum c.[29] In particular, c is optimal if

$$\exp\left(\int_0^t f_v(c_s, V_s(c))\, ds\right) f_c(c_t, V_t(c)) z_t^{\theta^*} = f_c(c_0, V_0)\pi_t, \quad \text{for all } t, \tag{5.7}$$

for some process θ^* in Θ_c, where, as mentioned earlier, we are restricting attention to supporting supergradients in the set Π defined in (5.3). 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.

To develop implications of these fist-order conditions, assume henceforth that the consumer has a Kreps-Porteus aggregator (3.14), which affords a simple

[29] See Duffie and Skiadas (1994) and Schroder and Skiadas (1999) for details regarding first-order conditions and their connection to security pricing.

parametric distinction between the effects of intertemporal substitution and risk aversion.[30] Suppose further that optimal consumption is an Ito process with time-varying drift and volatility, that is,

$$(5.8) \qquad dc_t/c_t = \mu_t^c\, dt + s_t^c \cdot dW_t.$$

Then (5.7), (5.4), and Ito's Lemma imply that for any process θ^* in Θ_c,

$$(5.9) \qquad (1-\rho)^{-1}(r_t - \beta) = \mu_t^c - \frac{(2-\rho)}{2} s_t^c \cdot s_t^c - \theta_t^* \cdot s_t^c + (\alpha-\rho)\left(\frac{\sigma_t}{\alpha V_t}\right) \cdot \left[s_t^c + 2^{-1}\rho(1-\rho)^{-1}\left(\frac{\sigma_t}{\alpha V_t}\right)\right]$$

and

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

where V_t and σ_t are the level and volatility of utility along the optimal consumption process. Write

$$(5.11) \qquad dX_t/X_t = b^M dt + s^M \cdot dW_t,$$

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

$$(5.12) \qquad \sigma_t/(\alpha V_t) = \rho^{-1}\left[s_t^M + (\rho-1)s_t^c\right],$$

along the optimal path.[31] Finally, substitution into (5.10) yields the following restriction for the market price of uncertainty:

$$(5.13) \qquad \eta_t = \rho^{-1}\left[\alpha(1-\rho)s_t^c + (\rho-\alpha)s_t^M\right] + \theta_t^*.$$

[30] Theorem 2.2 and Lemma 5.1 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 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. A second point is that the implied utility function 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.

[31] Multiply through (5.7) 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 left-hand-side 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.

The coming sections exploit this relation. Unless otherwise stated, κ-ignorance is assumed.

5.3. *Optimal Portfolio*

Examine the optimal portfolio when the risk-free rate and market price of uncertainty are deterministic constants. Then the optimal consumption process is geometric. Thus[32]

$$s_t^M = s_t^c, \tag{5.14}$$

and (5.13) implies that the market price of uncertainty satisfies

$$\eta_t = (1-\alpha)s_t^c + \theta_t^*. \tag{5.15}$$

Therefore,

$$\mathrm{sgn}(s_t^{c,i}) = \mathrm{sgn}(\eta_t^i - \kappa_i \ \mathrm{sgn}(s_t^{c,i})) \quad \text{for each } i.$$

Assume that ambiguity aversion is small in the sense that

$$0 \le \kappa_i < |\eta_t^i| \quad \text{for all } i. \tag{5.16}$$

Then

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

From (3.6), (5.12), and (5.14), infer that $\theta_t^* = \kappa \otimes \mathrm{sgn}(\eta_t)$ and

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

Finally, it follows from (5.5) and (5.14) that the optimal portfolio of risky assets is given by

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

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 addition to risk.[33] 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.

[32] 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.

[33] Alternatively, mean-variance efficiency is optimal if variance is computed using the appropriate measure Q^{θ^*}, as provided by Theorem 2.2(b). However, Q^{θ^*} depends on preferences through κ and thus the meaning of mean-variance efficiency is individual specific.

5.4. *Ambiguity and Risk Premia*

Next we view consumption described by the Ito process (5.8) as a given endowment and we focus on characterizing the risk-free rate and market price of uncertainty that support it, in the sense of satisfying (5.7), as a representative-agent equilibrium.

Re-examine briefly the first-order condition (5.7) from this perspective. 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 (Hansen and Jagannathan (1991)). 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. (See (5.20) for elaboration.)

Focus now on the implications of ambiguity for excess returns. From (5.13), we have the following model of excess returns:[34]

$$(5.17) \qquad b_t - r_t \mathbf{1} = s_t \eta_t = \rho^{-1}[\alpha(1-\rho)s_t s_t^c + (\rho - \alpha)s_t s_t^M] + s_t \theta_t^*.$$

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 is identical to that derived in Duffie and Epstein (1992b).

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

$$s_t^i \cdot \theta_t^* = -\mathrm{cov}_t\left(dR_t^i, dz_t^{\theta^*}/z_t^{\theta^*}\right),$$

for each security $i = 1, \ldots, N$, where s_t^i denotes the ith row of s_t and $z_t^{\theta^*}$ is as in (5.7). Thus the premium is positive if the asset's return has negative instantaneous covariation with $dz_t^{\theta^*}/z_t^{\theta^*}$. Recall from (2.9) 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 (5.12) 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.[35] Thus we consider three specializations of the endowment process, presented in increasing order of complexity, that permit sharper characterizations.

Geometric Consumption Process: Suppose that in (5.8) both μ_t^c and s_t^c are deterministic constants. From (3.6), (5.12), and (5.14),

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

[34] Much of what follows in the remainder of this section extends from κ-ignorance to IID ambiguity. 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.

[35] A similar criticism applies to the risk premium in (5.17) if $\alpha \neq \rho$. See Campbell (1999) and the references therein for further discussion and for proposed solutions in risk-based models.

where $K \subset R^d$ is the set corresponding to κ-ignorance defined in (3.11). If

$$s_t^{c,j} \neq 0 \tag{5.18}$$

for each component $j = 1, \ldots, d$, then we have the closed-form expressions

$$\theta_t^* = \kappa \otimes \mathrm{sgn}(s_t^c). \tag{5.19}$$

Thus (5.13) leads to the following expression for the market price of uncertainty:

$$\eta_t = (1-\alpha)s_t^c + \kappa \otimes \mathrm{sgn}(s_t^c). \tag{5.20}$$

In particular, η_t^i can be large even if consumption volatilities are small because the second term depends only on the sign of these volatilities and not on their magnitudes.

For excess returns, we have

$$b_t^i - r_t = (1-\alpha)s_t^i \cdot s_t^c + \kappa \cdot \left(s_t^i \otimes \mathrm{sgn}(s_t^c)\right). \tag{5.21}$$

The ambiguity premium (represented by the second term) for asset i is large if $s_t^{i,j}\mathrm{sgn}(s_t^{c,j})$ 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 $s_t^{c,j}$, $j = 1, \ldots, d$, large ambiguity premia can occur even if consumption is relatively smooth.

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

$$b_t^M - r_t = (1-\alpha)s_t^c \cdot s_t^c + \kappa \cdot |s_t^c|, \tag{5.22}$$

providing a decomposition of the *equity premium* in terms of risk (the first term) and ambiguity (the second term).[36] 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

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

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.

For the risk-free rate, substitute (5.12) and (5.14) into (5.9) to obtain

$$r_t - \beta = (1-\rho)\left(\mu_t^c - \frac{(1-\alpha)(2-\rho)}{2(1-\rho)} s_t^c \cdot s_t^c - \kappa \cdot |s_t^c|\right), \tag{5.23}$$

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

[36] Under IID ambiguity, the ambiguity premium on the right side of (5.22) is given by $e(s^c)$.

Markov Consumption Process: Assume that the drift and volatility in (5.8) are of the form

$$\mu_t^c = \hat{\mu}(c_t, t) \quad \text{and} \quad s_t^c = \hat{s}(c_t, t) \tag{5.24}$$

for functions $\hat{\mu}$ and $\hat{s}$. 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 $K \subset R^d$ is defined in (3.11), or more generally, K is the set corresponding to IID ambiguity as in Section 3.4. Under κ-ignorance, the formulae (5.19) and (5.21) are extended thereby to the present Markov specification.

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),

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

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, suppose the κ-ignorance specification satisfies

$$\kappa_i s_t^{c,i} = 0, \quad \text{for} \quad i = 1, \ldots, d. \tag{5.25}$$

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.

[37] A detailed derivation could be based on the 4-step procedure from Ma, Protter, and Yong (1994) applied to solve the FBSDE consisting of (2.17), (5.8), and (5.24). Intuitively, the point is simply that under 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).

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 (5.25) to deduce that

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

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 (3.6) and on the corresponding equilibrium.

The implied excess returns are obtained from the appropriate form of (5.17). For convenience, we reproduce the result here in the special case $\alpha = \rho$:

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

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 ith asset equals $s_t^i \cdot [\kappa \otimes \mathrm{sgn}(s_t^\omega)]$, 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.

Dept. of Mathematics, Shandong University, Jinan, China 250100; czengjing@hotmail.com,

and

Dept. of Economics, University of Rochester, Rochester, NY 14627, U.S.A; lepn@troi.cc.rochester.edu;http://www.econ.rochester.edu/Faculty/Epstein.html.

Manuscript received November, 1999; final revision received September, 2001.

A. APPENDIX: BSDE'S AND RELATED RESULTS

For the convenience of the reader, this appendix outlines *informally* some material regarding BSDE's. See El Karoui, Peng, and Quenez (1997) and Peng (1997) for further reading and formal details that are ignored here.

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

Given $\xi \in L^2(\Omega, \mathcal{F}_T, P)$ and a function g: $R^1 \times R^d \times \Omega \times [0,T] \to R^1$, 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.[38] Our definition of intertemporal utility for a given consumption process c (Theorem 2.2) 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 (El Karoui, Peng, and Quenez (1997, 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*

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

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

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

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

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

For a given Θ (satisfying our assumptions) and each t, the map $\xi \mapsto y_t$ 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* (Peng(1997)). 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*

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

where (μ_t^c) and (s_t^c) are continuous and bounded (adapted) processes. Let $e(\cdot)$ be as in (3.8). Then

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

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

[38] The analysis in El Karoui, Peng, and Quenez relies on the predictability of $(\omega, t) \mapsto g(y, \sigma, \omega, t)$. In our context, this would require predictability of consumption processes. However, the arguments in Pardoux and Peng (1990) and Peng (1997) rely only on progressive measurability of the above map for each fixed (y, σ). Thus the key existence and comparison theorems are valid for our setting.

The lemma provides the interpretation for $e(\cdot)$ promised in Section 3.4—it provides an instantaneous, per unit time premium for ambiguity. A proof may be found in the working paper version of this paper that is available on request.

B. APPENDIX: DENSITY GENERATORS AND THE SET OF PRIORS

Assume throughout that Θ is defined by (2.11), where (Θ_t) satisfies the regularity conditions in Section 2.4.

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 B.1: *The set of density generators Θ satisfies*:
(a) $0 \in \Theta$ *and* $\sup\{\|\theta\|_{L^\infty([0,T]\times\Omega)} : \theta \in \Theta\} < \infty$.
(b) *For any R^d-valued process (σ_t), there exists $(\theta_t^*) \in \Theta$ such that*

$$\theta_t^* \cdot \sigma_t = \max_{\theta \in \Theta} \theta_t \cdot \sigma_t = \max_{y \in \Theta_t} y \cdot \sigma_t. \tag{B.1}$$

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

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 $\theta^* = (\theta_t^*)$ constitutes a process. Then θ^* achieves the first maximum in (B.1) for every t and there is equality between the two maximizations shown.

PROOF: (b) The process θ^* is delivered by the Measurable Maximum Theorem (Aliprantis and Border (1994, 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 on $[0,T] \times \Omega$ (Revuz and Yor (1999, p. 44)). It ensures also that the value function for the latter problem is suitably measurable.

(c) Stochastic convexity is obvious. Weak compactness follows from Dunford and Schwartz (1958, Theorems IV.8.9, V.6.1). *Q.E.D.*

PROOF OF THEOREM 2.1: (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 El Karoui, Peng, and Quenez (1997, p. 20) and Uniform Boundedness, there exists $k > 0$ such that

$$(Q^\theta(A))^2 \leq kE(1_A) = kP(A),$$

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 (2.9). 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 $\mathcal{P}^\Theta$. Conclude similarly for other mixtures.

(d) Using the weak compactness of Θ (Lemma B.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 Cuoco and Cvitanic (1996).) Because Z is convex, it is also weakly closed. Clearly, Z is norm-bounded ($E(|z_T^\theta|) = 1$ 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 B.1(b). *Q.E.D.*

C. APPENDIX: PROOF OF EXISTENCE OF UTILITY

PROOF OF THEOREM 2.2: (a) First prove Lipschitz continuity of the support function e defined in (2.12). 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; dependence on ω has been suppressed notationally. Then

$$e_t(x) - e_t(x') \le y \cdot (x - x') \le d|y||x - x'|$$

and

$$e_t(x) - e_t(x') \ge y' \cdot (x - x') \ge -d|y'||x - x'|.$$

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

By the existence and uniqueness result in Pardoux and Peng (1990), there exist unique solutions (V_t, σ_t) and (V_t^Q, σ_t^Q) to (2.17) and (2.16) respectively.

(b) The Comparison Theorem and $\theta_t \cdot x_t \le \max_{y \in \Theta_t} y \cdot x_t$ for any x_t, imply that $V_t \le \min_{Q \in \mathcal{P}^\Theta} V_t^Q$. On the other hand, by Lemma B.1(b), there exists θ^* in Θ such that

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

in other words, $V_t = V_t^{Q^{\theta^*}} \ge \min_{Q \in \mathcal{P}^\Theta} V_t^Q$, proving equality and hence (2.14). Uniqueness is covered by the uniqueness results in Peng (1997).

(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

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

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

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

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

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

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

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

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

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

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

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

implying that $V_t = V_t' e^{-Kt}$ solves $V_t = \min_{Q \in \mathcal{P}^\Theta} E_Q[\int_t^T f(c_s, V_s)\, ds \mid \mathcal{F}_t]$. Similarly for $\tau < T$. *Q.E.D.*

D. APPENDIX: PROOFS OF PROPERTIES OF UTILITY

PROOF OF THEOREM 3.1: (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

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

Let $\overline{\mathcal{F}}_t = \sigma(W_{s+r} - W_r\colon s \leq t)$ 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 $(\bar{c}_t)_{0<t<T-r} = (c_{t+r})_{0 \leq t \leq T-r}$ is $\{\overline{\mathcal{F}}_t\}$-adapted. Thus there is a unique solution $(V_t^{T-r}(\bar{c}, \overline{W}), \sigma(\bar{c}))$ to

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

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

$$V_{l-r}^{T-r}(\bar{c}, \bar{W}) = \int_{l-r}^{T-r} [f(\bar{c}_s, V_s^{T-r}(\bar{c}, \bar{W})) - e(\sigma_s(\bar{c}))]\, ds - \int_{l-r}^{T-r} \sigma_s(\bar{c}) \cdot d\bar{W}_s,$$

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

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

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

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

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

Rewrite (D.1) 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 $(\bar{W}_t)_{0 \leq t \leq T-r}$ is $\{\bar{\mathcal{F}}_t\}_{0 \leq t \leq T-r}$-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 $\bar{\mathcal{F}}_t$-measurable relative to the probability space $(\Omega, \bar{\mathcal{F}}_{T-r}, \{\bar{\mathcal{F}}_t\}, P(\cdot \mid \sigma(W_r)))$. 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.

(d) For notational simplicity, let $\ell = 1$ and $d = 2$. Because c is $\{\mathcal{F}_t^{-1}\}$-adapted, there exists a unique $\{\mathcal{F}_t^1\}$-adapted solution (V_t, σ_t^1) to

$$dV_t = [-f(c_t, V_t) + \kappa_1 |\sigma_t^1|]\, dt + \sigma_t^1 dW_t^1, \quad V_T = 0.$$

Then $(V_t, \sigma_t^1, 0)$ is the unique $\{\mathcal{F}_t\}$-adapted solution to the corresponding 2-dimensional BSDE. *Q.E.D.*

Proof of Lemma 5.1: (a) Theorem 2.2(b) delivers Q^* in $\mathcal{P}^\Theta$ such that $V_t(c) = V^{Q^*}(c)$. Therefore, for any other c',

$$\begin{aligned} V(c') - V(c) &= V(c') - V^{Q^*}(c) = \min_{Q \in \mathcal{P}^\Theta} V^Q(c') - V^{Q^*}(c) \leq V^{Q^*}(c') - V^{Q^*}(c) \\ &\leq \int_0^T E_{Q^*}\left[\exp\left(\int_0^t f_v(c_s, V_s^{Q^*}(c))\, ds\right) f_c(c_t, V_t^{Q^*}(c))(c_t' - c_t)\right] dt \\ &= E_P\left[\int_0^T \pi_t(c_t' - c_t)\, dt\right]. \end{aligned}$$

The second inequality is due to the nature of supergradients for the stochastic differential utility function $V^{Q^*}(\cdot)$, as established in Duffie and Skiadas (1994).

(b): The argument is virtually identical to the proof of Lemma 2.2 in Epstein and Wang (1995). *Q.E.D.*

E. APPENDIX: AMBIGUITY

PROOF OF THEOREM 4.1: Assume probabilistic sophistication relative to $\overline{P}$. Then $\overline{P}$ is nonatomic.

Fix $0 < t_1 < t_2 < T$. Let B_1^* and $A_1^* \subset R^d$ be Borel sets such that

$$0 < \overline{P}(B_1 \cap A_1) = \overline{P}(B_1^c \cap A_1) < 1, \quad \text{where} \tag{E.1}$$

$$B_1 \equiv \{\omega\colon W_{t_1} \in B_1^*\} \quad \text{and} \quad A_1 = \{\omega\colon W_{t_2} \in A_1^*\}.$$

Such sets exist by the nonatomicity of $\overline{P}$. We show that all measures in $\mathcal{P}^\Theta$ agree on $B_1 \cap A_1$. Then Marinacci (2002) implies that they agree on all events in $\mathcal{F}_{t_2}$. Since $t_2 < T$ is arbitrary, they agree on $\mathcal{F}_T$.

Define $A_2 = A_1$, $A_3 = A_4 = A_1^c$, $B_3 = B_1$, and $B_2 = B_4 = B_1^c$. Then $\{B_i\}_{i=1}^2$ and $\{B_i \cap A_i\}_{i=1}^4$ form partitions of Ω.

A real-valued and $\mathcal{F}_{t_2}$-measurable random variable (or act) f can be associated with a consumption process c^f such that $c_t^f = 0$ for $t < t_2$ and such that $f(\omega) = V_{t_2}(c^f, \omega)$, the continuation utility of c^f. By the recursivity of utility, any such c^f induces the identical utility $V_0(c^f)$, which can therefore be viewed as 'the utility of f.' In this way, we can think of utility and preference as defined on acts f rather than on consumption processes. Abuse notation and write $V_0(f)$. The acts of particular relevance in what follows have the form $f = \sum_1^n x_i 1_{B_i \cap A_i}$, where the x_i's are real numbers.

The hypothesis of probabilistic sophistication implies that

$$f \equiv x_1 1_{B_1 \cap A_1} + x_2 1_{B_2 \cap A_2} + \sum_{i>2} x_i 1_{B_i \cap A_i} \sim x_2 1_{B_1 \cap A_1} + x_1 1_{B_2 \cap A_2} + \sum_{i>2} x_i 1_{B_i \cap A_i}, \tag{E.2}$$

because under (E.1) the two acts induce the same distributions over outcomes under $\overline{P}$. We establish further implications from the fact that the stated indifference applies for all x_i's.

Rectangularity of $\mathcal{P}^\Theta$ implies that the utility of f can be computed in the following two-stage manner:

$$V_0(f) = \min_{Q \in \mathcal{P}^\Theta} \begin{bmatrix} Q(B_1) \min\limits_{q \in \mathcal{P}^\Theta} (x_1 q(A_1 \mid B_1) + x_3 q(A_3 \mid B_1)) \\ +(1 - Q(B_1)) \min\limits_{q \in \mathcal{P}^\Theta} (x_2 q(A_2 \mid B_2) + x_4 q(A_4 \mid B_2)) \end{bmatrix} \tag{E.3}$$

Step 1: Let $x_4 < \min\{x_1, x_2\} < \max\{x_1, x_2\} < x_3$. Then

$$V_0(f) = \min_{Q \in \mathcal{P}^\Theta} \begin{bmatrix} Q(B_1)([\max\limits_{q \in \mathcal{P}^\Theta} q(A_1 \mid B_1)](x_1 - x_3) + x_3) \\ +(1 - Q(B_1))([\min\limits_{q \in \mathcal{P}^\Theta} q(A_2 \mid B_2)](x_2 - x_4) + x_4) \end{bmatrix} \tag{E.4}$$

$$= \min_{Q \in \mathcal{P}^\Theta} \left[Q(B_1) \left(\left[\max_{q \in \mathcal{P}^\Theta} q(A_1 \mid B_1) \right] (x_1 - x_3) - \left[\min_{q \in \mathcal{P}^\Theta} q(A_2 \mid B_2) \right] (x_2 - x_4) + x_3 - x_4 \right) \right]$$

$$+ \left[\min_{q \in \mathcal{P}^\Theta} q(A_2 \mid B_2) \right] (x_2 - x_4) + x_4.$$

For many x_i's the bracket multiplying $Q(B_1)$ is positive and thus Q is chosen to minimize $Q(B_1)$, leading to

$$V_0(f) = \left[\min_{Q \in \mathcal{P}^\Theta} Q(B_1) \max_{q \in \mathcal{P}^\Theta} q(A_1 \mid B_1) \right] x_1 + \left[\max_{Q \in \mathcal{P}^\Theta} Q(B_2) \min_{q \in \mathcal{P}^\Theta} q(A_2 \mid B_2) \right] x_2 + h(x_3, x_4),$$

for a function h whose definition does not matter for what follows. By (E.2), the right-hand-side is symmetric in x_1 and x_2. Deduce that

$$\min_{Q \in \mathcal{P}^\Theta} Q(B_1) \max_{q \in \mathcal{P}^\Theta} q(A_1 \mid B_1) = \max_{Q \in \mathcal{P}^\Theta} Q(B_2) \min_{q \in \mathcal{P}^\Theta} q(A_2 \mid B_2). \tag{E.5}$$

Similarly,

$$\max_{Q\in\mathcal{P}^{\Theta}} Q(B_1)\min_{p\in\mathcal{P}^{\Theta}} q(A_1 \mid B_1) = \min_{Q\in\mathcal{P}^{\Theta}} Q(B_2)\max_{q\in\mathcal{P}^{\Theta}} q(A_2 \mid B_2). \tag{E.6}$$

Step 2: Let $x_1 < x_3 = x_4 < x_2$. Then (E.4) is valid and calculations paralleling those above deliver

$$V_0(f) = \left[\max_{Q\in\mathcal{P}^{\Theta}} Q(B_1)\max_{q\in\mathcal{P}^{\Theta}} q(A_1 \mid B_1)\right]x_1 + \left[\min_{Q\in\mathcal{P}^{\Theta}} Q(B_2)\min_{q\in\mathcal{P}^{\Theta}} q(A_2 \mid B_2)\right]x_2 + h'(x_3, x_4).$$

Step 3: Let $x_2 < x_3 = x_4 < x_1$. For $x_4 - x_2$ sufficiently larger than $x_3 - x_1$, compute

$$V_0(f) = \left[\max_{Q\in\mathcal{P}^{\Theta}} Q(B_1)\min_{q\in\mathcal{P}^{\Theta}} q(A_1 \mid B_1)\right]x_1 + \left[\min_{Q\in\mathcal{P}^{\Theta}} Q(B_2)\max_{q\in\mathcal{P}^{\Theta}} q(A_2 \mid B_2)\right]x_2 + h''(x_3, x_4).$$

The last two equations and (E.2) imply that

$$\max_{Q\in\mathcal{P}^{\Theta}} Q(B_1)\max_{q\in\mathcal{P}^{\Theta}} q(A_1 \mid B_1) = \min_{Q\in\mathcal{P}^{\Theta}} Q(B_2)\max_{q\in\mathcal{P}^{\Theta}} q(A_2 \mid B_2).$$

Similarly,

$$\max_{Q\in\mathcal{P}^{\Theta}} Q(B_2)\max_{q\in\mathcal{P}^{\Theta}} q(A_2 \mid B_2) = \min_{Q\in\mathcal{P}^{\Theta}} Q(B_1)\max_{q\in\mathcal{P}^{\Theta}} q(A_1 \mid B_1).$$

Combine with (E.6) to deduce that $\min_{q\in\mathcal{P}^{\Theta}} q(A_1 \mid B_1) = \max_{q\in\mathcal{P}^{\Theta}} q(A_1 \mid B_1)$ and

$$\{q(A_1 \mid B_1) \colon q \in \mathcal{P}^{\Theta}\} \text{ is a singleton.} \tag{E.7}$$

Now (E.5)–(E.6) imply

$$1 \geq \frac{\min_{Q\in\mathcal{P}^{\Theta}} Q(B_1)}{\max_{Q\in\mathcal{P}^{\Theta}} Q(B_1)} = \frac{\max_{Q\in\mathcal{P}^{\Theta}} Q(B_2)}{\min_{Q\in\mathcal{P}^{\Theta}} Q(B_2)} \geq 1,$$

and thus $\{Q(B_1) \colon Q \in \mathcal{P}^{\Theta}\}$ is a singleton. Combine with (E.7) to obtain that $\{Q(B_1)Q(A_1 \mid B_1) \colon Q \in \mathcal{P}^{\Theta}\} = \{Q(A_1 \cap B_1) \colon Q \in \mathcal{P}^{\Theta}\}$ is a singleton. *Q.E.D.*

PROOF OF LEMMA 4.2: Assume (4.2). 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

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

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

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

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

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

or, in terms of the support function (3.8), $e(\sigma_t) = -e(-\sigma_t')$. By the nonnegativity of e, $e(\sigma_t) = 0$. Apply (3.10) 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. *Q.E.D.*

PROOF OF THEOREM 4.3: (i) $\Leftarrow$: 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 $\kappa^* \geq \kappa$, (3.7), and the Comparison Theorem A.1.

$\Rightarrow$: 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 $\kappa^* \gg 0$, 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 (4.4), (3.7), and the Comparison Theorem to deduce that $\kappa^* \geq \kappa$.

(ii) $\Leftarrow$: 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 Duffie and Epstein (1992a). The converse is similar to (i). *Q.E.D.*

Consider finally the relation between the designation (4.1) and the definition of ambiguous events in Epstein and Zhang (2001). It is immediate that an event where all measures agree is unambiguous in the sense of that paper. We show that, under IID ambiguity, and for 'many' events the converse is valid.

LEMMA E.1: *Suppose that Θ conforms to IID ambiguity. Let E be unambiguous according to Epstein-Zhang. If also $E \in \sigma(W_s\colon t_0 < s < t)$ for some $0 < t_0 < t < T$, then $Q(E) = P(E)$ for all Q in $\mathcal{P}^\Theta$.*

PROOF: Exclude the trivial case $\Theta \equiv \{0\}$. Then there exists A in $\mathcal{F}_{t_0}$ such that

$$0 < \min_{\mathcal{P}^\Theta} Q(A^c) \neq \min_{\mathcal{P}^\Theta} Q(A) > 0. \tag{E.8}$$

(Under IID ambiguity, if all measures agree on $\mathcal{F}_{t_0}$, then Lemma 4.2 implies that they agree on $\mathcal{F}_T$ and hence $\Theta \equiv \{0\}$. This is the only point at which IID ambiguity is used; thus the lemma admits substantial generalization.) Define $E_1 = E \cap A$, $E_2 = E \cap A^c$, $A_1 = A \cap E_1^c$, and $A_2 = A^c \cap E_2^c$.

Proceed as in the proof of Theorem 4.1 to translate preference over consumption processes into preference over $\mathcal{F}_t$-measurable real-valued random variables. In terms of this derived preference order, E unambiguous implies that

$$x^* 1_{A_1} + x 1_{A_2} + z 1_E \sim x 1_{A_1} + x^* 1_{A_2} + z 1_E \quad \text{iff}$$
$$x^* 1_{A_1} + x 1_{A_2} + z' 1_E \sim x 1_{A_1} + x^* 1_{A_2} + z' 1_E.$$

The fact that this invariance is required to hold for all choices of x^*, x, z, and z', regardless of relative magnitudes, is the source of its power.

Proceed as in the proof of Theorem 4.1 to exploit rectangularity and compute utilities in two stages, delivering thereby closed-form expressions for the utilities of the above acts. Then tedious algebra and application of (E.8) deliver

$$\max_{\mathcal{P}^\Theta} Q(E_2 \mid A^c) = \min_{\mathcal{P}^\Theta} Q(E_2 \mid A^c) = \min_{\mathcal{P}^\Theta} Q(E_1 \mid A) = \min_{\mathcal{P}^\Theta} Q(E_1 \mid A),$$

which implies, by the noted two-stage calculation, that

$$\min_{\mathcal{P}^\Theta} p(E) = \min_{\mathcal{P}^\Theta} \left\{ p(A) \min_{\mathcal{P}^\Theta} Q(E_1 \mid A) + (1 - p(A)) \min_{\mathcal{P}^\Theta} Q(E_2 \mid A^c) \right\}$$
$$= \min_{\mathcal{P}^\Theta} Q(E_1 \mid A) = \max_{\mathcal{P}^\Theta} Q(E_1 \mid A) = \max_{\mathcal{P}^\Theta} p(E).$$

Hence all measures in $\mathcal{P}^\Theta$ agree on E. *Q.E.D.*

REFERENCES

ABEL, A. (2002): "An Exploration of the Effects of Pessimism and Doubt on Asset Returns," *Journal of Economic Dynamics and Control*, 26, 1075–1092.

ALIPRANTIS, C., AND K. BORDER (1994): *Infinite Dimensional Analysis*. Berlin: Springer-Verlag.

ANDERSON, E., L. P. HANSEN, AND T. J. SARGENT (2000): "Robustness, Detection and the Price of Risk," Mimeo, Stanford University.

CAGETTI, M., L. HANSEN, T. SARGENT, AND N. WILLIAMS (2002): "Robustness and Pricing when Growth is Hidden," *Review of Financial Studies*, 15, 363–404.

CAMPBELL, J., (1999): "Asset Prices, Consumption and the Business Cycle," in *Handbook of Macroeconomics I*, ed. by J. Taylor and M. Woodford. Amsterdam: North-Holland, 1231–1303.

CECCHETTI, S., P. LAM, AND N. MARK (2000): "Asset Pricing with Distorted Beliefs: Are Equity Returns too Good to be True?" *American Economic Review*, 90, 787–805.

CHEN, Z. (1998): "Existence and Uniqueness for BSDE with Stopping Time," *Chinese Science Bulletin*, 43, 96–99.

CUOCO, D., AND J. CVITANIC (1996): "Optimal Consumption Choices for a 'Large' Investor," *Journal of Economic Dynamics and Control*, 22, 401–436.

DUFFIE, D. (1996): *Dynamic Asset Pricing Theory*, 2nd ed. Princeton: Princeton U. Press.

DUFFIE, D., AND L. G. EPSTEIN (1992a): "Stochastic Differential Utility," *Econometrica*, 60, 353–394 (Appendix C with C. Skiadas).

——— (1992b): "Asset Pricing with Stochastic Differential Utility," *Review of Financial Studies*, 5, 411–436.

DUFFIE, D., AND C. SKIADAS (1994): "Continuous-Time Security Pricing: A Utility Gradient Approach," *Journal of Mathematical Economics*, 23, 107–131.

DUNFORD, N., AND J. T. SCHWARTZ (1958): *Linear Operators Part I: General Theory*. New York: John Wiley.

EL KAROUI, N., S. PENG, AND M. QUENEZ (1997): "Backward Stochastic Differential Equations in Finance," *Mathematical Finance*, 7, 1–71.

EPSTEIN, L. G. (1987): "The Global Stability of Efficient Intertemporal Allocations," *Econometrica*, 55, 329–358.

——— (1999): "A Definition of Uncertainty Aversion," *Review of Economic Studies*, 66, 579–608.

EPSTEIN, L. G., AND J. MIAO (2001): "A Two-Person Dynamic Equilibrium under Ambiguity," *Journal of Economic Dynamics and Control*, forthcoming.

EPSTEIN, L. G., AND M. SCHNEIDER (2001a): "Recursive Multiple-Priors," Mimeo, University of Rochester.

——— (2001b): "IID: Independently and Indistinguishably Distributed," Mimeo, University of Rochester.

EPSTEIN, L. G., AND T. WANG (1994): "Intertemporal Asset Pricing under Knightian Uncertainty," *Econometrica*, 62, 283–322.

——— (1995): "Uncertainty, Risk-Neutral Measures and Security Price Booms and Crashes," *Journal of Economic Theory*, 67, 40–80.

EPSTEIN, L. G., AND J. ZHANG (2001): "Subjective Probabilities on Subjectively Unambiguous Events," *Econometrica*, 69, 265–306.

EPSTEIN, L. G., AND S. ZIN (1989): "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

GHIRARDATO, P., AND M. MARINACCI (2002): "Ambiguity Made Precise: A Comparative Foundation," *Journal of Economic Theory*, 102, 251–289.

GILBOA, I., AND D. SCHMEIDLER (1989): "Maxmin Expected Utility with Nonunique Prior," *Journal of Mathematical Economics*, 18, 141–153.

HANSEN, L., AND R. JAGANNATHAN (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

HANSEN, L., AND T. SARGENT (2000): "Acknowledging Misspecification in Macroeconomic Theory," Sargent's Frisch Lecture at the 2000 World Congress of the Econometric Society.

——— (2001): "Robust Control and Model Uncertainty," *American Economic Review*, 91, 60–66.

HANSEN, L., T. SARGENT, AND T. TALLARINI (1999): "Robust Permanent Income and Pricing," *Review of Economic Studies*, 66, 873–907.

KARATZAS, I., AND E. SHREVE (1988): *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.

LEPELTIER, J. P., AND J. S. MARTIN (1997): "Backward Stochastic Differential Equations with Continuous Coefficients," *Statistics and Probability Letters*, 32, 425–430.

MA, J., P. PROTTER, AND J. YONG (1994): "Solving Forward-Backward Stochastic Differential Equations Explicitly—a Four Step Scheme," *Probability Theory and Related Fields*, 98, 339–359.

MACHINA, M., AND D. SCHMEIDLER (1992): "A More Robust Definition of Subjective Probability," *Econometrica*, 60, 745–780.

MARINACCI, M. (1999): "Limit Laws for Non-Additive Probabilities and their Frequentist Interpretation," *Journal of Economic Theory*, 84, 145–195.

——— (2002): "Probabilistic Sophistication and Multiple Priors," *Econometrica*, 70, 755–764.

MEHRA, R., AND E. C. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

ONATSKI, A. (2000): "Robust Monetary Policy under Model Uncertainty: Incorporating Rational Expectations," Harvard University.

ONATSKI, A., AND J. STOCK (2002): "Robust Monetary Policy under Model Uncertainty in a Small Open Economy," *Macroeconomic Dynamics*, 6, 85–110.

PARDOUX, E. (1997): "Generalized Discontinuous BSDE's," in *Backward Stochastic Differential Equations*, Pitman Research Notes in Mathematics 364, ed. by N. El Karoui and L. Mazliak. Essex: Addison Wesley Longman, 207–219.

PARDOUX, E., AND S. PENG (1990): "Adapted Solution of a Backward Stochastic Differential Equation," *Systems Control Letters*, 14, 55–61.

PENG, S. (1997): "BSDE and Related g-Expectation," in *Backward Stochastic Differential Equations*, Pitman Research Notes in Mathematics 364, ed. by N. El Karoui and L. Mazliak Essex: Addison Wesley Longman, 141–159.

REVUZ, D., AND M. YOR (1999): *Continuous Martingales and Brownian Motion*. Berlin: Springer-Verlag.

ROCKAFELLER, R. T. (1970): *Convex Analysis*. Princeton: Princeton University Press.

SCHRODER, M., AND C. SKIADAS (1999): "Optimal Consumption and Portfolio Selection with Stochastic Differential Utility," *Journal of Economic Theory*, 89, 68–126.