Foundations for Learning How to Invest when Returns are Uncertain

Thomas A. Knox
The University of Chicago Graduate School of Business

August 23, 2003*

Abstract

Most asset returns are uncertain, not merely risky: investors do not know the probabilities of different possible future returns. A large body of evidence suggests that investors are averse to uncertainty, as well as to risk. This paper builds up an axiomatic foundation for the dynamic portfolio and consumption choices of an uncertainty-averse (as well as risk-averse) investor who tries to learn from historical data. The theory developed, *model-based multiple-priors*, generalizes existing theories of dynamic choice under uncertainty aversion by relaxing the assumption of consequentialism. Examples are given to show that consequentialism, the property that counterfactuals are ignored, can be problematic when combined with uncertainty aversion. An analog of de Finetti's statistical representation theorem is proven under model-based multiple-priors, but consequentialism combines with multiple priors to rule out prior-by-prior exchangeability. A simple dynamic portfolio choice problem illustrates the contrast between a model-based multiple-priors investor and a consequentialist multiple-priors investor.

*This paper is a revision of the first chapter of my doctoral thesis; I am indebted to Gary Chamberlain and John Campbell for guidance and encouragement throughout the writing of my thesis. I am also grateful for the insights of Brian Hall, Lars Hansen, Parag Pathak, Jacob Sagi, Tom Sargent, Jeremy Stein, and James Stock, and for the helpful comments of seminar participants at the University of California, Berkeley, the University of Chicago Graduate School of Business, the Fuqua School of Business at Duke University, Harvard Business School, Harvard University, the Kellogg School of Management at Northwestern University, New York University Stern School of Business, Princeton University, and the Wharton School of the University of Pennsylvania. I benefitted from detailed suggestions made by Larry Epstein and Martin Schneider. I am solely responsible for the remaining shortcomings of this paper.

1 Introduction

Most asset returns are uncertain, not merely risky: investors do not know the probabilities of different possible future returns. An aversion to uncertainty, or a preference for bets with known odds, has been used in recent studies, including Anderson, Hansen, and Sargent (2000), Chen and Epstein (2002), and Maenhout (2001), to explain the equity premium puzzle. However, as Maenhout (2001) notes, these explanations often rely on investors ignoring data and dogmatically expecting the worst. Whether uncertainty aversion remains a plausible explanation for the equity premium puzzle when learning is accounted for is an open question. Uncertainty aversion has also been used by Liu, Pan, and Wang (2003) to explain option smirk effects, and by Routledge and Zin (2001) to model the dynamics of asset market liquidity. In both cases, the impact of learning might also be of interest. Models of learning in portfolio choice have received significant attention recently from authors including Barberis (2000), Brennan (1998), Brennan and Xia (2001), Kandel and Stambaugh (1996), Pástor (2000), Pástor and Stambaugh (2000), and Xia (2001). However, these papers do not account for uncertainty aversion. Since uncertainty aversion leads to behavior compatible with "extreme" priors, incorporating uncertainty aversion may change portfolio choices significantly.

The intersection of learning and uncertainty aversion has only recently begun to receive attention in portfolio choice and asset pricing with the work of Cagetti, Hansen, Sargent, and Williams (2002), Epstein and Schneider (2002), Hansen, Sargent, and Wang (2002), and Miao (2001). This paper lays the foundations for a novel approach to learning and uncertainty aversion in portfolio choice, and contrasts this approach with others in the literature.

First, a theory of *model-based multiple-priors* is built up from axioms on preferences. The model-based multiple-priors approach can incorporate learning and uncertainty aversion. It generalizes existing theories of dynamic choice under uncertainty aversion by relaxing the assumption of consequentialism, the property that counterfactuals are ignored. Model-based multiple-priors is then compared to consequentialist approaches, with a special focus on the most obvious alternative to the model-based multiple-priors approach: the work of Epstein and Schneider (2002) on learning within the (consequentialist) recursive multiple-priors framework. Examples are given which suggest that consequentialism may be unattractive when combined with multiple priors. Under model-based multiple-priors, a multiple-priors statistical representation theorem is proven which provides an analog of the usual single-prior de Finetti theorem; such an analog does not currently appear to be available for consequentialist multiple-priors theories. The central role of the de Finetti theorem in single-prior subjective expected utility is discussed in Chapter 11 of Kreps (1988) and in Section 7 of Chapter 3 of Savage (1954), and a large part of these discussions also applies to the multiple-priors analog proven here. Finally, model-based multiple-priors is contrasted with consequentialist multiple-priors methods in the context of a simple portfolio choice problem. A companion paper, Knox (2003), builds on the foundation established here and solves in closed form a class of portfolio and consumption choice problems with learning and uncertainty aversion in continuous time, including problems in which the investor has uncertainty about the accuracy of an asset pricing model.

The most famous experimental demonstration of uncertainty aversion, and the motivation for many economic applications of uncertainty aversion, is the Ellsberg Paradox (Ellsberg (1961)): Consider two urns, each containing 100 balls. Each ball is either red or black. The first ("known") urn contains 50 red balls (and thus 50 black balls). The second ("ambiguous") urn contains an unknown number of red balls. One ball is drawn at random from each urn, and four bets based on the results of these draws are to be ranked; winning a bet results in a 100 dollar cash prize. The first bet is won if the ball drawn from the known urn is red; the second bet is won if the ball drawn from the known urn is black; the third bet is won if the ball drawn from the ambiguous urn is red; the fourth bet is won if the ball drawn from the ambiguous urn is black. Many decision makers are indifferent between the first and second bets and between the third and fourth bets, but strictly prefer either the first or second bet to either the third or fourth bet. Since no distribution on the number of red balls in the ambiguous urn can support these preferences through expected utility, this ranking of gambles violates the axioms of subjective expected utility theory.

In a seminal response to the Ellsberg Paradox, Gilboa and Schmeidler (1989) provided an axiomatic foundation to support uncertainty aversion in static choice. They worked in the Anscombe and Aumann (1963) framework, and weakened the independence axiom. They then showed that, under this weakening, preferences could be represented by the minimum expected utility over a set of (prior) distributions. For example, an agent with Gilboa-Schmeidler preferences would exhibit Ellsberg-type behavior if the set of distributions on the number of red balls in the unknown urn included a distribution under which black balls were more numerous and one under which red balls were more numerous. The optimal choice under these assumptions is that which maximizes (over possible choices) the minimum (over the set of distributions) expected utility, leading to the label "maxmin expected utility." Atemporal extensions of maxmin expected utility include the work of Casadesus-Masanell, Klibanoff, and Ozdenoren (2000), who were able to obtain a maxmin expected utility representation of preferences in the Savage (1954) framework, and the smooth model of uncertainty aversion developed by Klibanoff, Marinacci, and Mukerji (2003), which nests maxmin expected utility as a special case.

In extending the pioneering atemporal work of Gilboa and Schmeidler (1989) to a dynamic setting, however, there has been little consensus. A number of important recent studies, including Chamberlain (2000), Chamberlain (2001), Chen and Epstein (2002), Epstein and Schneider (2003), Epstein and Schneider (2002), Epstein and Wang (1994), Hansen, Sargent, and Tallarini (1999), Hansen and Sargent (2001), Hansen, Sargent, Turmuhambetova, and Williams (2001), Klibanoff (1995) (who appears to have given the first axiomatization of an explicitly dynamic maxmin expected utility theory), Siniscalchi (2001), and Wang (2003) have attacked the problem of dynamic choice under uncertainty aversion (in addition to risk aversion). In the literature, a debate is in progress over which method is to be preferred: the recursive multiple-priors method (Chen and Epstein (2002), Epstein and Schneider (2003), Epstein and Schneider (2002), and Epstein and Wang (1994)) or the robust control method (Anderson, Hansen, and Sargent (2000), Hansen and Sargent (1995), Hansen, Sargent, and Tallarini (1999), Hansen and Sargent (2001), Hansen, Sargent, Turmuhambetova, and Williams (2001) and, with an important variation, Maenhout (2001)). Although Chamberlain's work has been more econometrically focused, he is evidently aware of the portfolio-choice implications of his research, and his approach is a third angle of attack on the problem. Of these three approaches, model-based multiple-priors is closest to Chamberlain's, although he does not build an axiomatic framework or study the investment implications of his approach.

The recursive multiple-priors approach began nonaxiomatically in discrete time with the work of Epstein and Wang (1994). In Chen and Epstein (2002) the approach was brought into a continuous-time framework, the portfolio choice problem was considered generally and solved analytically in some cases, and the separate effects of "ambiguity" (uncertainty) and risk were shown in equilibrium. The recursive multiple-priors approach was given axiomatic foundations in Epstein and Schneider (2003), and learning was explicitly incorporated by Epstein and Schneider (2002). Despite the obvious importance of this strand of the literature, it is argued in Section 6 that model-based multiple-priors enjoys some significant advantages over the approach of Epstein and Schneider (2002).

Hansen and Sargent (1995) first used the robust control approach for economic modeling, although a large literature on robust control in engineering and optimization theory predates their work. The development of the model continued with the discrete-time study of Hansen, Sargent, and Tallarini (1999), and was then brought into a continuous-time setting by Anderson, Hansen, and Sargent (2000). Filtering, which may be regarded as learning about an ever-changing state, was combined with robust control in discrete time by Hansen, Sargent, and Wang (2002), and was analyzed in continuous time by Cagetti, Hansen, Sargent, and Williams (2002). The work of Hansen and Sargent (2001) and Hansen, Sargent, Turmuhambetova, and Williams (2001) responded to criticisms of the robust control approach made in some studies using the recursive multiple-priors approach (notably Epstein and Schneider (2003)). Finally, Maenhout (2001) modified the robust control "multiplier preferences" to obtain analytical solutions to a number of portfolio choice problems.

The debate between the recursive multiple-priors school and the robust control school has focused on the "constraint preferences" generated by robust control. However, in applications the robust control school typically uses the "multiplier preferences" generated by robust control, which are acknowledged by both schools to differ from the constraint preferences (though they are observationally equivalent to the constraint preferences in any single problem; see Epstein and Schneider (2003) and Hansen, Sargent, Turmuhambetova, and Williams (2001)).

The continuing debate regarding how to extend atemporal maxmin expected utility to intertemporal situations is essentially a debate about the structure of the set of prior distributions with which expected utility is evaluated. This paper shows the implications of the existence of general consistent conditional preferences for the set of distributions used to represent utility. In this connection, a novel *restricted independence axiom* is introduced, and a class of distributions, termed *prismatic*, is characterized. The set of prismatic distributions is a strict superset of the set of rectangular distributions introduced by Epstein and Schneider (2003) (see below), and allows the consequentialism assumption to be relaxed.

The remainder of the paper proceeds as follows. Section 2 lays out the domain of preferences, that is, the set of bets or gambles over which the decision maker is to choose. Section 3 presents the axioms used to formalize the decision maker's preference structure. Section 4 delivers basic results concerning the existence of consistent conditional preferences, their link to the restricted independence axiom, and the associated shape of the set of priors used by the decision maker. Section 5 specializes these basic results to justify model-based multiple-priors in a dynamic domain which includes consumption at various points in time. Section 6 puts model-based multiple-priors in perspective by comparing and contrasting it with consequentialist multiple-priors theories. Section 7 contrasts model-based multiple-priors with consequentialist multiple-priors theories in a simple dynamic portfolio choice problem. Section 8 concludes. All proofs are placed in an appendix which follows the text of the paper.

2 The Domain of Preferences

The domain of preferences used here is similar to that used by Gilboa and Schmeidler (1989), which in turn is based on the setting of Anscombe and Aumann (1963). Let X be a non-empty set of *consequences*, or *prizes*. In applications of the theory X will often be the set of possible consumption bundles. Let Y be the set of probability distributions over X having finite supports, that is, the set of *simple* probability distributions on X.

Let S be a non-empty set of *states*, let Σ be an algebra of subsets of S, and define L_0 to be the set of all Σ-measurable finite step functions from S to Y. Note that L_0 is a set of mappings from states to simple probability distributions on consequences, rather than mappings from states directly to consequences, and that each $f \in L_0$ takes on only a finite number of different values (since it is a finite step function). This is the same L_0 used by Gilboa and Schmeidler (1989). Let L_c denote the constant functions in L_0. Following Anscombe and Aumann (1963), elements of L_0 will be termed "horse lotteries" and elements of L_c will be termed "roulette lotteries." It is important to note that convex combinations in L_0 are to be performed pointwise, so that $\forall f, g \in L_0, \quad \alpha f + (1-\alpha) g$ is the function from S to Y whose value at $s \in S$ is given by $\alpha f(s) + (1-\alpha) g(s)$. In turn, convex combinations in Y are performed as usual: if the probability mass function (not the density function, since all distributions in Y have finite support) of $y \in Y$ is $p_y(x)$ and the probability mass function of $z \in Y$ is $p_z(x)$, then the probability mass function of $\alpha y + (1-\alpha) z$ is $\alpha p_y(x) + (1-\alpha) p_z(x)$.

3 Axioms

The decision maker ranks elements of L_0 using the preference ordering $\succsim$. The axioms below are placed on $\succsim$ and on the strict preference ordering, $\succ$, derived from it by: $\forall f, g \in L_0, \quad f \succ g \Leftrightarrow f \succsim g$ and not $f \precsim g$. The indifference relationship, $\sim$, is defined by: $\forall f, g \in L_0, \quad f \sim g \Leftrightarrow f \succsim g$ and $f \precsim g$.

3.1 The Restricted Independence Axiom

The key, novel axiom presented here is the *restricted independence axiom.* It is stated relative to a subset $A \in \Sigma$ (recall that Σ is the algebra of subsets of S with respect to which the functions in L_0 are measurable), and is thus referred to as restricted independence relative to A.

Axiom 1 *For all $f, g \in L_0$, if $f(s) = g(s) \quad \forall s \in A^C$, and if $h(s)$ is constant on A, then*

$$\forall \alpha \in (0,1), \quad f \succ g \quad \Leftrightarrow \quad \alpha f + (1-\alpha) h \succ \alpha g + (1-\alpha) h.$$

Suppose two gambles give the same payoff for each state in some set of states. Descriptively, a decision maker may find it relatively easy to consider changing each of these gambles in the same way on that set of states (so that they continue to agree on that set). The decision maker's preferences might be preserved by changing each of these gambles in the same way on their set of agreement. The restricted independence axiom (relative to the set on which the two gambles differ) goes slightly beyond this, by saying that preferences are still preserved if, in addition to being changed on their set of agreement as described above, each gamble is also mixed with the same roulette lottery (or gamble that does not depend on the state) on the set on which they disagree. Still, this seems quite plausible descriptively. In contrast, the Ellsberg (1961) paradox shows the descriptive failings of the usual (or unrestricted) independence axiom.

Preferences that satisfy the restricted independence axiom relative to a collection of sets will be of special interest. Preferences will be said to satisfy the restricted independence axiom (or Axiom 1) relative to $\mathcal{A}$ if $\mathcal{A} = \{A_1, \ldots, A_k\}$, where preferences satisfy Axiom 1 relative to A_i for each $i \in \{1, \ldots, k\}$.

In working with partitions of S, the following axiom is often useful. It is the "roulette lottery partition-independence" analog of the "state-independence" axiom used to obtain an expected utility representation from an additively separable ("state-dependent expected utility") representation in the Anscombe and Aumann (1963) framework (see Kreps (1988), page 109).

Axiom 2 *Given* **roulette lotteries** *$l, q \in L_c$, a finite partition $\mathcal{A} = \{A_1, \ldots, A_k\} \subset \Sigma$ of S, and $h \in L_0$, define*

$$(l; h)_i = \begin{cases} l & \text{for } s \in A_i, \\ h & \text{for } s \in A_i^C, \end{cases} \tag{1}$$

and

$$(q; h)_i = \begin{cases} q & \text{for } s \in A_i, \\ h & \text{for } s \in A_i^C. \end{cases} \tag{2}$$

Then $\forall i, j \in \{1, \ldots, k\}, \quad (l; h)_i \succsim (q; h)_i \Leftrightarrow (l; h)_j \succsim (q; h)_j$.

3.2 The Gilboa-Schmeidler Axioms

The axioms of Gilboa and Schmeidler (1989) are grouped together into the following axiom.

Axiom 3

Weak Order:	$\succsim$ *is complete and transitive.*
Certainty Independence:	$\forall f, g \in L_0, \forall l \in L_c,$ *and* $\forall \alpha \in (0,1)$, $f \succ g \Leftrightarrow \alpha f + (1-\alpha) l \succ \alpha g + (1-\alpha) l.$
Continuity:	$\forall f, g, h \in L_0,\ f \succ g \succ h \Rightarrow \exists \alpha, \beta \in (0,1)$ *such that* $\alpha f + (1-\alpha) h \succ g \succ \beta f + (1-\beta) h.$
Monotonicity:	$\forall f, g \in L_0,\ f(s) \succsim g(s)\ \forall s \in S \Rightarrow f \succsim g.$
Uncertainty Aversion:	$\forall f, g \in L_0,\ f \sim g \Rightarrow \alpha f + (1-\alpha) g \succsim g$ $\forall \alpha \in (0,1).$
Non-degeneracy:	$\exists f, g \in L_0$ *such that* $f \succ g.$

The labels attached to each portion of the axiom are those used by Gilboa and Schmeidler (1989). Of these portions of the axiom, Weak Order, Continuity, Monotonicity, and Non-degeneracy are completely standard in the axiomatic literature on choice under uncertainty. The usual independence axiom is strictly stronger than the Certainty Independence and Uncertainty Aversion portions of the axiom above. Descriptively, it seems more plausible that Certainty Independence would hold than full-blown independence: Decision makers may find it easier to work through the implications of mixing with a roulette lottery, which yields the same probability distribution over consequences in every state, than to see the implications of mixing with a horse lottery, which generally yields different probability distributions over consequences in different states.

These axioms are the standard "maxmin expected utility" axioms. The theory developed here is built on the base they form.

3.3 The Consistent Conditioning Axiom

In developing a theory of conditional preferences, the notion of a *null set* will be useful.

Definition 1 *A set* $B \in \Sigma$ *is a* **null set** *if and only if* $\forall\ f, g \in L_0,\ f(s) = g(s)\ \ \forall s \in B^C \Rightarrow f \sim g.$

The axiom below formalizes the notion that conditional preferences should be genuinely conditional; that is, if two horse lotteries (elements of L_0) have identical payoffs on some set of states, then a preference relation conditional on the state being in that set ought to display indifference between the two horse lotteries. This property is termed *focus*, and when conditioning is based in a particular way on the filtration describing the information structure of a dynamic problem it specializes to *consequentialism*, which has been extensively investigated by Hammond (1988).

In Skiadas (1997a) and Skiadas (1997b), preferences satisfying the focus property are referred to as "separable."

There should also be some minimal link between conditional and unconditional preference relations: if every conditional preference relation in an exhaustive set displays a weak preference for one horse lottery (element of L_0) over another, then the unconditional preference relation ought to display a weak preference for the first horse lottery, too. If, in addition, one of the conditional preference relations, conditional on a set of states that is not null, displays a strict preference for the first horse lottery, then the unconditional preference relation ought to display a strict preference for the first horse lottery, too. This is a notion of *consistency*, and is part of the axiom below. Consistency is closely related to the "coherence" property introduced by Skiadas (1997a) and used by Skiadas (1997b).

Finally, the following axiom allows conditional preferences to display uncertainty aversion by assuming that each conditional preference relation satisfies an appropriate modification of Axiom 3.

Axiom 4 *Given a finite partition $\mathcal{A} = \{A_1, \ldots, A_k\} \subset \Sigma$ of S, $\succsim$ admits consistent conditioning relative to $\mathcal{A}$ if and only if there exists a conditional preference ordering $\succsim_i$ on L_0 for each $A_i \in \mathcal{A}$, and these conditional preference orderings satisfy:*

Focus: $\forall i \in \{1, \ldots, k\}$, $f(s) = g(s)$ $\forall s \in A_i \Rightarrow f \sim_i g$.

Consistency: $f \succsim_i g$ $\forall i \in \{1, \ldots, k\} \Rightarrow f \succsim g$.
If, in addition, $\exists A_i \in \mathcal{A}$ such that $f \succ_i g$ and A_i is not null, then $f \succ g$.

Multiple Priors: $\forall i \in \{1, \ldots, k\}$, $\succsim_i$ *satisfies Axiom 3, but with A_i substituted for S in the definition of monotonicity and non-degeneracy holding only for A_i that are not null.*

The concept of consistency is most familiar in the form of dynamic consistency, but consistency seems to be a desirable property in any situation involving a set of conditional preferences relative to a partition. Below, consistency is examined in the context of preferences conditional on the value of a parameter (broadly defined so as to include high-dimensional parameters, structural breaks, and the like). While the descriptive merits of consistency, and especially dynamic consistency, are not uncontroversial, the normative appeal of consistency is difficult to question.

Of course, the appeal of consistency as a property of conditional preferences does not speak to the appeal of the "focus" section of the axiom above. If the partition involved is linked to a filtration (that is, represents the revelation of information to the decision maker over time), then the "focus" property is better known as "consequentialism." When the full independence axiom is assumed, the normative appeal of consequentialism seems clear; however, when the independence axiom is relaxed, consequentialism can become very unattractive. This issue is discussed extensively in Section 6, where it is shown that conditional preferences that have the focus property but are not consequentialist (because the partition is not linked to the revelation of information) can avoid many of the difficulties of consequentialist conditional preferences.

As with the restricted independence axiom, it is sometimes desirable to strengthen the consistent conditioning axiom by imposing an axiom linking conditional preferences over roulette lotteries. This is intuitively quite sensible: it seems natural that preferences over constant acts should not depend on the element of the partition the decision maker finds herself in.

Axiom 5 *Given a finite partition $\mathcal{A} \subset \Sigma$ of S and roulette lotteries $l, q \in L_c$,*

$$\forall i, j \in \{1, \ldots, k\}, l \succsim_i q \Leftrightarrow l \succsim_j q.$$

4 General Results

4.1 The Relation of Restricted Independence to Consistent Conditional Preferences

Theorem 1 *Assume Axiom 3, and consider a finite partition $\mathcal{A} = \{A_1, \ldots, A_k\} \subset \Sigma$ of S. Then the restricted independence axiom holds relative to each $A_i \in \mathcal{A}$ if and only if preferences admit consistent conditioning relative to $\mathcal{A}$. That is, Axiom 1 holds if and only if Axiom 4 holds (each being relative to $\mathcal{A}$).*

Like all other results in this paper, the proof of this theorem has been placed in a separate appendix. This theorem allows one to precisely gauge the strength of the assumption that a given set of consistent conditional preferences exists (with respect to some partition). It links preferences over *strategies*, or contingent plans formed before information arrives and is conditioned upon, to conditional preferences. It is of particular interest because it exposes the relationship between the independence axiom, which has been the focus of the axiomatic work on uncertainty aversion, and the existence of consistent conditional preferences.

4.2 Prismatic Sets of Priors

Definition 2 *A set $\mathcal{P}$ of priors will be said to be* **prismatic** *with respect to the finite partition $\mathcal{A}$ if and only if there exist closed convex sets $\mathcal{C}_i$, $i \in \{1, \ldots, k\}$ of finitely additive probability measures, where each measure $P_i \in \mathcal{C}_i$ has $P_i(A_i) = 1$, and a closed convex set of finitely additive measures $\mathcal{Q}$, where each $Q \in \mathcal{Q}$ has $Q(A_i) > 0 \ \forall i \in \{1, \ldots, k\}$, such that*

$$\mathcal{P} = \left\{ \begin{array}{c} P : \forall B \in \Sigma, P(B) = \sum_{i=1}^{k} P_i(B) Q(A_i) \\ \text{for some } P_i \in \mathcal{C}_i, i \in \{1, \ldots, k\} \text{ and } Q \in \mathcal{Q} \end{array} \right\}.$$

Prismatic sets of priors generalize the rectangular sets of priors introduced in Epstein and Schneider (2003) in a straightforward way: the shape of a prismatic set of priors need not be linked to how information is revealed to the decision maker, while the shape of a rectangular set of priors is tightly linked to the order in which information is revealed to the decision maker. If a set of priors were prismatic with respect to the set of partitions induced by the flow of information to the decision maker over time, that set of priors would also be rectangular. However, a set of priors may be prismatic without being rectangular; this will occur if the partition with respect to which the set of priors is prismatic is unrelated to the filtration describing the revelation of information to the decision maker over time. Situations in which it is natural to break the link between the filtration and the shape of the set of priors are examined in Section 6. In general, if all of the uncertainty in a decision problem is concentrated in the parameters of a model, prismatic sets of priors may lead to behavior that is more intuitive and appealing than the behavior that would result from rectangular sets of priors.

The defining property of a prismatic set of priors is that any conditional may be chosen from the set $\mathcal{C}_i$, regardless of how other conditionals or the marginals are selected. This freedom in the selection of different components of a prior has important behavioral implications which are laid out in Theorem 2 below.

4.3 The Basic Representation Result

Theorem 2 *Given a finite partition $\mathcal{A} = \{A_1, \ldots, A_k\} \subset \Sigma$ of S such that each $A_i \in \mathcal{A}$ is non-null, the following conditions are equivalent:*

(1) Axioms 1 and 2 (relative to the partition $\mathcal{A}$) and Axiom 3.

(2) Axioms 4 and 5 (relative to the partition $\mathcal{A}$) and Axiom 3.

(3) There exists a closed, convex set of finitely additive measures $\mathcal{P}$ that is prismatic with respect to $\mathcal{A}$ and a mixture linear and nonconstant $u : Y \to \mathbb{R}$ such that $\succsim$ is represented by

$$\min_{P \in \mathcal{P}} \left\{ \int_{s \in S} u\left(f\left(s\right)\right) dP\left(s\right) \right\}.$$

In this representation, $\mathcal{P}$ is unique and u is unique up to a positive affine transformation. Moreover, there is a set of conditional preference relations $\succsim_i$, $i \in \{1, \ldots, k\}$ relative to $\mathcal{A}$, and for each $i \in \{1, \ldots, k\}$, $\succsim_i$ is represented by

$$\min_{P_i \in \mathcal{C}_i} \left\{ \int_{s \in A_i} u\left(f\left(s\right)\right) dP_i\left(s\right) \right\}.$$

In this representation, $\mathcal{C}_i = \{P_i : \forall B \in \Sigma,\ P_i\left(B\right) = P\left(B \mid A_i\right) \text{ for some } P \in \mathcal{P}\}$, and is thus a closed convex set of finitely additive measures, and is unique by the uniqueness of $\mathcal{P}$.

This theorem refines the result of Gilboa and Schmeidler (1989) by obtaining a set of distributions $\mathcal{P}$ that has a special structure. It is a natural generalization of the situation explored by Epstein and Schneider (2003), who obtained a result similar to the equivalence between (2) and (3). However, they do not consider restricted independence and the structure of the set of priors in their result is tied to the filtration that governs the revelation of information to the decision maker. In contrast, the result above holds for any partition of states of the world (subject to the stated conditions). This added generality will prove crucial in the development of model-based multiple-priors.

As noted above, the key to the structure of a prismatic set of distributions is that any conditional may be selected from the set $\mathcal{C}_i$, regardless of how other conditionals or the marginals are chosen. Intuitively, it is sensible that this "independence" in the selection of the distribution corresponds to the restricted independence of Axiom 1.

5 Model-based Multiple-priors

The theory of model-based multiple-priors is formulated in a dynamic setting. Time is discrete and the horizon is finite: $t \in \{0, 1, \ldots, T\}$.

To maintain continuity of notation, let the state space be denoted S as before (rather than the notation Ω more typical in the stochastic-process literature). Each state of the world is composed of a model parameter value and a vector of observable variables:

$$s = (\theta, z_0, \ldots, z_T), \tag{3}$$

where at time $t \in \{0, 1, \ldots, T\}$ the investor has observed $(z_0, \ldots, z_t)$. Note that the value of the model parameter θ is, in general, never observed by the investor. Let the set of model parameters be denoted Θ and the set of vectors of observable variables be denoted Z.

The variables z_t are revealed to the investor in a certain order, and may be considered as a stochastic process indexed by $t \in \{0, 1, \ldots, T\}$. Denote the filtration that the process $\{z_t\}_{t=0}^T$ generates by $\{\mathcal{F}_t\}_{t=0}^T$, where $\mathcal{F}_0$ is trivial (includes only Z and ϕ, the empty set). For any $t \in \{0, 1, \ldots, T\}$, the *atoms* of the σ-algebra $\mathcal{F}_t$ will be of special interest. The atoms of a σ-algebra are the sets in that σ-algebra such that any other set in the σ-algebra is the union of some collection (possibly empty) of atoms. The atoms of $\mathcal{F}_t$, taken together, thus make up the finest partition of S that can be formed using sets in $\mathcal{F}_t$. It is assumed that there is a finite number of atoms in each $\mathcal{F}_t$ for $t \in \{0, 1, \ldots, T\}$. This amounts to the assumption that the range of each z_t is a finite set (this assumption is made for clarity, and could easily be relaxed). Any filtration $\{\mathcal{F}_t\}_{t=0}^T$ satisfying this assumption has an event-tree representation, in which each atom of $\mathcal{F}_t$ is identified with a set of terminal nodes that originate from some node at the time-t level of the tree. The branches from a node at time t to nodes at time $t+1$ can be thought of as connecting an atom in $\mathcal{F}_t$ to the atoms in $\mathcal{F}_{t+1}$ which partition it.

Given any σ-algebra Σ_Θ on Θ, the set of model parameters, let the σ-algebra on $S = \Theta \times Z$ be defined by $\Sigma = \Sigma_\Theta \times \mathcal{F}_T$. Observe that the definition of Σ makes clear that the value of the model parameter is, in general, never revealed to the investor.

The set X of *consequences* or *prizes* will be the set of $(T+1)$-long sequences of consumption bundles, $(c_0, c_1, \ldots, c_T)$, such that each $c_t \in C$ for some set C (which might, for example, be the positive reals). It is now tempting to proceed with L_0 equal to the set of all $\mathcal{F}_T$ measurable finite step functions from Z into Y, the set of simple probability distributions on X (note that direct dependence on θ, the model parameter, is not allowed, although θ will generally have an impact through its influence on $(z_0, \ldots, z_T)$). This is problematic because such a definition would fail to account for the order in which information is revealed according to the filtration $\{\mathcal{F}_t\}_{t=0}^T$, since L_0 would then include, for example, horse lotteries in which c_0, consumption at time 0, was $\mathcal{F}_T$ measurable but not $\mathcal{F}_0$ measurable (in other words, constant, since $\mathcal{F}_0$ is trivial). This would mean that consumption at time zero was dependent on some possible outcome not "known" (according to the filtration) until some time in the future, making it difficult to use the filtration for its customary purpose: to represent the information structure of the environment.

Instead, attention is restricted to the subset of $\{\mathcal{F}_t\}_{t=0}^T$ *adapted* acts in L_0. A horse lottery $f \in L_0$ will be called $\{\mathcal{F}_t\}_{t=0}^T$ adapted if $f : Z \to Y$ is such that $f(z) = (f_0(z), \ldots, f_T(z))$ where for $t \in \{0, \ldots, T\}$, $f_t(z)$ is a simple probability distribution on C for fixed z, and is $\mathcal{F}_t$ measurable as a function of z. A roulette lottery $l \in L_c$, then, is a constant horse lottery; thus, a roulette lottery is a $(T+1)$-long sequence of simple probability distributions $(l_0, \ldots, l_T)$, where l_t is a simple probability distribution on C whose realization is c_t. Note that the random variables governed by the simple probability distributions l_s and l_t are independent if $s \neq t$.

These are similar to the set of adapted horse lotteries and the set of roulette lotteries that Epstein and Schneider (2003) work with, though θ is not part of the state of the world for them. Following their notation, the set of adapted horse lotteries is denoted $\mathcal{H}$ below.

To apply the results of Section 4, it is necessary to produce some set of consequences, denoted X^U, paired with the set of simple probability distributions on it, labelled Y^U, such that the set of $\mathcal{F}_T$-measurable finite step functions $f : Z \to Y^U$ is equivalent, from a preference perspective, to $\mathcal{H}$. This is achieved by showing that under Axiom 3, preferences over roulette lotteries are representable by a von Neumann-Morgenstern utility function, which is, moreover, additively time-separable. The vN-M utility function is also, as usual, unique up to a positive affine transformation. Then one may take $X^U \subset \mathbb{R}$ to be the set of all vN-M utility values arising from consumption lotteries. Preferences over adapted acts naturally induce a preference relation on $f : Z \to Y^U$, which then satisfies Axiom 3.

Proposition 1 *Suppose that $\succsim$, defined on $\mathcal{H}$, satisfies Axiom 3. Then, on the subset of $\mathcal{H}$ composed of roulette lotteries, $\succsim$ is represented by a mixture linear function $v(\cdot)$, which is unique up to a positive affine transformation. Moreover, v is additively time-separable.*

Proposition 1 will now be used to show that the results obtained in Section 4 continue to hold when preferences are defined on the dynamic domain of the current section.

Theorem 3 *Theorems 1 and 2 continue to hold when $\succsim$ is defined on the dynamic domain of this section. Moreover, the function u in Theorem 2 is additively time-separable.*

To obtain an axiomatic foundation for model-based multiple-priors, consider (for clarity) the case in which Θ is finite and $\Sigma_\Theta = 2^\Theta$, and choose the partition in Theorem 2 such that $A_i = \{\theta_i\} \times Z$. This partition is model-based: states of the world are partitioned according to the values of an economic model's parameters. If one assumes Axiom 1, the restricted independence axiom, and Axiom 2 with respect to this partition, and also assumes Axiom 3, then Theorem 2 implies that there is a maxmin expected utility representation for preferences, and that the set of distributions on the state of the world is prismatic with respect to the partition $\{A_1, \ldots, A_k\}$. (In place of Axioms 1 and 2, one could assume Axioms 4 and 5 with respect to the partition.) The prismatic structure of the set of distributions on S is, in fact, a "multiple-priors multiple-likelihoods" structure: there is a set of distributions on the parameters of the economic model and, given any value of the parameters of the model, there is a set of distributions on the vector of data.

To obtain a model-based multiple-priors representation, the set of likelihoods (distributions of the data given the model parameters) must be reduced to a single likelihood. This can be done in one of two (equivalent) ways: a stronger version of Axiom 1 may be assumed in which the mixing horse lottery h can be arbitrary, or, equivalently, Axiom 4 may be strengthened by assuming that each conditional preference ordering satisfies not only certainty independence, but the full independence axiom. Either one of these (equivalent) strengthened assumptions will deliver many priors (distributions on the parameters of the economic model) but only one likelihood (the distribution of the data given the parameters of the economic model). It is important to note that there are consistent conditional preferences, conditional on the *parameters of an economic model*, in model-based multiple-priors.

Finally, it is worth contrasting model-based multiple-priors with an approach which takes Θ alone as the state space and applies maxmin expected utility to horse lotteries defined on Θ. The key difference between these two approaches is that in model-based multiple-priors, the subjective nature of the distribution of z given θ (the likelihood) is acknowledged and incorporated, while in a maxmin approach that takes Θ alone as its state space, the likelihood must be assumed to be objectively given. In most economic settings, the assumption that the likelihood is objectively given appears unrealistic. In that sense, the relation between model-based multiple-priors and the "only Θ" maxmin approach is analogous to the relation between Anscombe and Aumann (1963) (or Savage (1954)) subjective expected utility and von Neumann-Morgenstern expected utility.

6 A Comparison of Uncertainty Aversion Frameworks

Using the dynamic framework that has been developed above, it is possible to compare and contrast model-based multiple-priors and the approaches to uncertainty aversion that have been used in the asset pricing literature. The work of Cagetti, Hansen, Sargent, and Williams (2002) and Hansen, Sargent, and Wang (2002) is not examined in detail, because these studies are focused on filtering (which might be thought of as learning about an infinite-dimensional parameter), and do not offer the sort of general theory of learning about a model parameter that this paper seeks to provide. The paper of Miao (2001) is also not discussed in depth, because his work deals with what might be called a "single-prior, multiple-likelihoods" framework. This does incorporate both learning and uncertainty aversion, but there is no uncertainty (in a multiple-priors sense) about the model parameters in Miao (2001): there is a single prior on the model parameters. The focus here is on situations in which there is uncertainty about the model parameters. The study of Epstein and Schneider (2002) offers an alternative to model-based multiple-priors, and their work is discussed in detail below.

In model-based multiple-priors, as in the work of Chamberlain (2000), the state of the world consists of both the parameters of an economic model and a vector of data: $s = (\theta, z)$, where s is the state of the world, θ is a vector of model parameters, and z is a vector of data. While the formal axiomatic development above treats finite partitions, it might be taken as motivation for the use of the model-based multiple-priors approach when the parameter space is a subset of finite-dimensional Euclidean space, or even when the parameter space is infinite-dimensional (so that the problem is of the type typically referred to as "nonparametric" or the type usually termed "semiparametric"). In any event, the model-based multiple-priors approach can certainly be implemented in such settings, though working in a nonparametric framework may incur a significant computational cost.

Some of the recent work of Epstein and Schneider (Epstein and Schneider (2003) and Epstein and Schneider (2002)), on the other hand, uses a set of partitions. Indeed, in any event tree, the Epstein and Schneider (2003) assumptions state that consistent, consequentialist conditional preferences exist conditional on any node in the tree. One could prove the Epstein-Schneider theorem by repeatedly applying Theorem 2. In fact, the condition on the set of distributions that they term "rectangularity" is a special case of the prismatic condition, in which the basic partition is linked to the way in which information is revealed to the decision maker. This link to the information structure of the decision problem and the resulting consequentialism of the preferences of a recursive multiple-priors decision maker are key features that distinguish recursive multiple-priors from model-based multiple-priors.

A crucial point here is that the nonexistence of a set of consistent, consequentialist conditional preferences (for example, conditioned on the nodes at some level of an event tree) does *not* imply dynamic inconsistency or a need for "committed updating." Rather, a single (least favorable) prior will be selected, and that prior will be updated in the usual, Bayesian way. The use of an economic model in model-based multiple-priors, for instance, does not indicate that dynamic inconsistency arises. Far from it: the decision maker selects a prior on the model parameters and updates it using Bayes rule. In model-based multiple-priors, it is only at the beginning of the observation and decision process that many priors are considered. Once the least favorable among them has been isolated, it is used. A natural question would be: "What if the decision maker were confronted with a choice between gambles at some later date, after the selection of the least favorable prior?" The decision maker would then step back to time zero and rank the gambles at that point, using the full set of priors. There is absolutely nothing *inconsistent* about this behavior; rather, it is not what one ordinarily thinks of as *conditional* choice behavior, since the decision maker takes into account counterfactuals when weighing alternatives. (There is a large literature on why one might want to take counterfactuals into account; Machina (1989) is a good review of the earlier portion of this literature, and argues strongly against consequentialism from both a normative perspective and a descriptive perspective.) Thus, it is not so much the consistency part of Axiom 4 that is of concern, but the focus part of the axiom (which can lead to consequentialism, depending on the partitions with respect to which it holds). This makes very good sense; it is well known that consequentialism is fundamentally linked to the independence axiom, and Theorem 1 makes the link explicit in the current setting. If one does not wish to impose the restricted independence axiom with respect to a particular set of partitions (linked to the information structure of the problem) on preferences at date zero, one must refrain from assuming the existence of consequentialist, consistent conditional preferences.

Three examples are now given to illustrate why consequentialism may be significantly less appealing once the independence axiom has been relaxed. This point has been made very persuasively by Machina (1989) in the context of risk. In the course of one of these examples, a statistical representation theorem is proven which is the model-based multiple-priors version of the classical de Finetti theorem (see Kreps (1988) and Savage (1954)). Since such a statistical representation theorem is apparently not currently available for consequentialist multiple-priors theories, this would seem to increase the relative appeal of model-based multiple-priors.

6.1 A Three-color Ellsberg Urn

Epstein and Schneider (2003) give a three-color Ellsberg urn example which demonstrates a potential hazard of consequentialism in the absence of the independence axiom. In this example, a ball is drawn from an urn which contains red (R), blue (B), and green (G) balls. In their Section 4.1, Epstein and Schneider (2003) focus on the situation in which there are 90 balls in the urn, of which 30 are known to be red and 60 are either blue or green. They note that "[a] natural state space is $\Omega = \{R, B, G\}$," and consider a decision problem in which there are three periods: $t \in \{0, 1, 2\}$. Epstein and Schneider (2003) stipulate that "the color [of the ball] is revealed to the decision-maker at $t = 2$," and that "[a]t the intermediate stage time 1, the decision-maker is told whether or not the color drawn is G." Any distribution over the possible colors of the ball can be represented by a probability vector:

$$p = (p_R, p_B, p_G), \tag{4}$$

where p_R is the probability that the ball drawn is red, p_B is the probability that the ball drawn is blue, and p_G is the probability that the ball drawn is green. Sets of priors are sets of such probability vectors.

As noted by Epstein and Schneider (2003), in order to be rectangular (in accord with recursive multiple-priors), a set of priors that admits a range of probabilities of a green ball versus a blue ball must also admit a range of probabilities of a red ball. But this is troubling: the probability of a red ball being drawn is known to be $\frac{1}{3}$, since there are 30 red balls and 90 balls in all. Even in the most favorable scenario, in which the interior of the interval of probabilities of a red ball includes the true probability $\frac{1}{3}$, this means that a recursive multiple-priors decision maker who owned the contingent claim "one hundred dollars if the ball drawn is red, otherwise zero dollars" and who had a range of priors would be willing to pay someone to trade this claim for a bet that paid one hundred dollars with probability $\frac{1}{3}$. Since the contingent claim and the bet are probabilitistically identical, this willingness to pay in order to trade one for the other is disturbing.

Note that a model-based multiple-priors decision maker who used the partition of states into $\{R\}$ and $\{G, B\}$ would not exhibit the problematic preferences discussed above.

6.2 Exchangeability and De Finetti

Consider sampling with replacement from an ordinary, two-color Ellsberg urn. Figure 1 shows what form a rectangular set must take in this context.

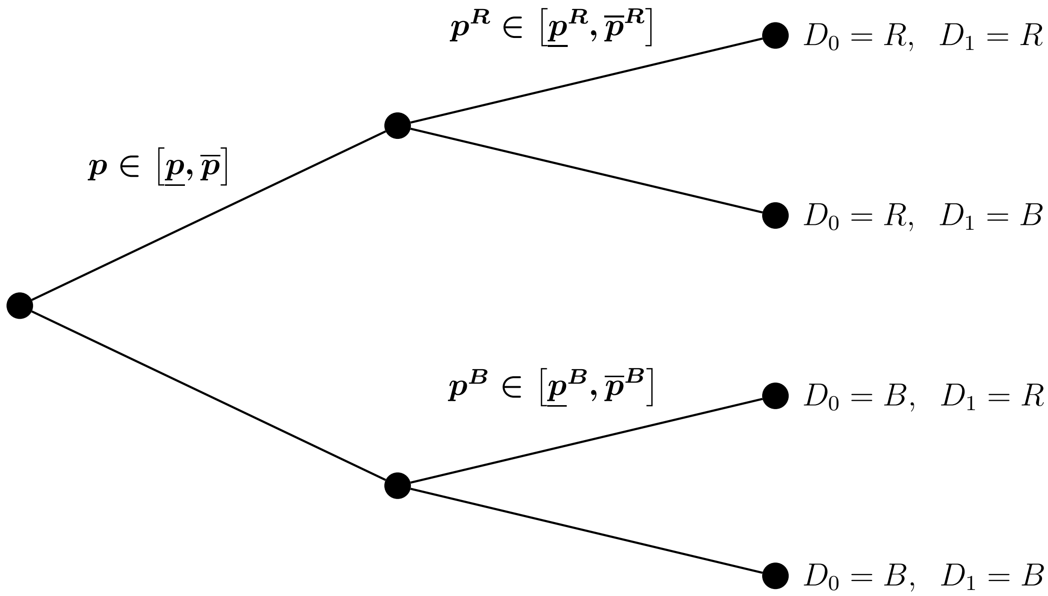

Figure 1: **A Rectangular Set of Priors for Sampling with Replacement from an Ellsberg Urn**
The event tree above applies to sampling with replacement from a (two-color) Ellsberg urn. The set of priors depicted is rectangular, and any rectangular set of priors in this context must take the form shown.

clearpage

In this setting, the *ex ante* probability of a red ball followed by a black ball should equal the *ex ante* probability of a black ball followed by a red ball:

$$p\left(1-p^{R}\right) \;=\; \left(1-p\right)p^{B}. \tag{5}$$

However, if the intervals $\left[\underline{p},\overline{p}\right]$, $\left[\underline{p}^{R},\overline{p}^{R}\right]$, and $\left[\underline{p}^{B},\overline{p}^{B}\right]$ are nontrivial in a rectangular set of priors, the typical prior in that rectangular set will not satisfy exchangeability. Indeed, if the rectangular set is thought of in three-dimensional Euclidean space, with p on the first axis, p^{R} on the second axis, and p^{B} on the third axis, the subset of priors that satisfy exchangeability will have Lebesgue measure zero. This can be seen by observing that an exchangeable prior must satisfy the equation above, and that thus the set of exchangeable priors forms a surface in the three-dimensional space described.

In contrast, in model-based multiple-priors, exchangeability is natural. In fact, a de Finetti theorem under uncertainty can be proven: any closed, convex set of distributions on an infinite sequence of Bernoulli (zero-one) random variables is exchangeable if and only if it can be represented as the set of distributions derived from a single i.i.d. binomial likelihood and a closed, convex set of priors on $p = \Pr\left(X_i = 1\right)$. The closedness of a set of distributions is under the topology of weak convergence, matching the representation result obtained by Gilboa and Schmeidler (1989) and specialized here.

To state and prove the result, one must specify the metric with respect to which the continuity of real-valued functions of infinite sequences of zeros and ones will be judged. This will determine which sets of distributions over infinite zero-one sequences are considered to be closed. Although the result is not overly sensitive to the precise choice of metric, a convenient choice is the metric

$$d\left(x,y\right) \;\equiv\; \sum_{i=1}^{\infty}\frac{1}{2^{i}}\left|x_i-y_i\right|, \tag{6}$$

where $x=\left(x_1,x_2,\ldots\right)$ and $y=\left(y_1,y_2,\ldots\right)$ are infinite sequences such that $x_i, y_i \in \{0,1\}\ \forall i$. The metric used to determine the continuity of real-valued functions on $[0,1]$ is the usual Euclidean distance.

Theorem 4 *Given any closed, convex set $\mathcal{P}$ of distributions on $[0,1]$, define the set $\mathcal{E}$ of distributions on $\{0,1\}^{\infty}$ (the set of all infinite sequences of zeros and ones) as the set of all distributions E on $\{0,1\}^{\infty}$ such that $\exists\ Q \in \mathcal{P}$ with the property that, for any $N<\infty$ and for any distinct $i_1, i_2, \ldots, i_N \in \mathbb{N}$,*

$$\Pr_E\left(x_{i_1},x_{i_2},\ldots,x_{i_N}\right) \;\equiv\; \int_0^1 p^{\sum_{j=1}^{N}x_{i_j}}\left(1-p\right)^{N-\sum_{j=1}^{N}x_{i_j}}\,dQ\left(p\right). \tag{7}$$

Then every distribution in $\mathcal{E}$ is exchangeable, and $\mathcal{E}$ is closed and convex.

Conversely, given any closed, convex set $\mathcal{E}$ of distributions on $\{0,1\}^{\infty}$ such that every $E\in\mathcal{E}$ is exchangeable, there is a closed, convex set $\mathcal{P}$ of distributions on $[0,1]$ with the property that, $\forall\ E\in\mathcal{E}$, $\exists\ Q\in\mathcal{P}$ such that, for any $N<\infty$ and for any distinct $i_1,i_2,\ldots,i_N\in\mathbb{N}$,

$$\Pr_E\left(x_{i_1},x_{i_2},\ldots,x_{i_N}\right) \;\equiv\; \int_0^1 p^{\sum_{j=1}^{N}x_{i_j}}\left(1-p\right)^{N-\sum_{j=1}^{N}x_{i_j}}\,dQ\left(p\right). \tag{8}$$

This theorem provides a statistical representation of any set of exchangeable distributions on infinite sequences of zero-one random variables. It is of particular importance for model-based multiple-priors because it justifies, in terms of the properties of distributions on observables, an approach in which uncertainty is concentrated in model parameters. The importance of de Finetti's theorem in the context of standard subjective expected utility is discussed in detail by Kreps (1988) and Savage (1954). Theorem 4 provides an analog of this fundamental result in the model-based multiple-priors framework. Crucially, however, it seems difficult to prove a result of this nature for multiple-priors theories other than model-based multiple-priors; certainly, the proof given here for model-based multiple-priors could not be used in a consequentialist multiple-priors context.

6.3 Preferences over Derivative Assets

A specific set of derivative assets can reveal some counterintuitive behavior on the part of a consequentialist multiple-priors investor. Suppose there is a risky asset whose return in each of two periods is either high (H) or low (L). Label the first period "period zero" and the second period "period one." Consider one derivative, A, on the risky asset that pays off 1,000 dollars at the end of period one if the return sequence is (H, L) and otherwise pays off one dollar at the end of period one, and another derivative, B, that pays off 1,000 dollars at the end of period one if the return sequence is (L, H) and otherwise pays off one dollar at the end of period one. A and B are essentially bets on the order of the high and the low return, if one high and one low return are realized.

For a recursive multiple-priors investor, the set of distributions over possible pairs of period-zero and period-one returns is rectangular:

$$\Pr(R_0 = H) \in \left[\underline{p}, \overline{p}\right] \tag{9}$$

$$\Pr(R_1 = H | R_0 = H) \in \left[\underline{p}^H, \overline{p}^H\right] \tag{10}$$

$$\Pr(R_1 = H | R_0 = L) \in \left[\underline{p}^L, \overline{p}^L\right]. \tag{11}$$

One might expect that an investor would be indifferent between holding a portfolio of only A and a holding a portfolio of only B. Indeed, it would seem reasonable that the investor would be indifferent between A, B, and flipping a fair coin, then holding a portfolio of only A if the coin came up heads, and only B if the coin came up tails. However, if $\overline{p} > \underline{p}$, so that the investor has uncertainty aversion over the time-zero return, and if $\underline{p}^L > 0$ and $\overline{p}^H < 1$ (ruling out dogmatic beliefs about time-one returns), then a recursive multiple-priors investor cannot be indifferent between A, B, and randomizing over A and B with equal probabilities. In contrast, a model-based multiple-priors investor using an i.i.d. model for returns will always be indifferent between these three choices.

First, consider the preferences of a model-based multiple-priors investor whose model is that returns are i.i.d. It is innocuous, given the invariance of von Neumann-Morgenstern utility rankings to positive affine transformations of the utility function, to normalize the utility of one dollar to zero and the utility of 1,000 dollars to one. Then the maxmin expected utility of holding A (to a model-based multiple-priors investor) is:

$$\min_{\pi \in \Pi} \left\{ \int_0^1 p\left(1-p\right) d\pi\left(p\right) \right\}, \tag{12}$$

where p is the probability of a high return in any given period, and Π is a set of priors on that probability. But this is also the maxmin expected utility of holding B, and it is the maxmin expected utility of holding any roulette lottery whose prizes are A and B. Essentially, this is due to exchangeability; the model-based multiple-priors investor feels that (H, L) and (L, H) are equiprobable return sequences.

In contrast, the preferences of a recursive multiple-priors investor must reflect the rectangular structure of her set of priors (see Figure 3). Thus, the maxmin expected utility of holding A is, by the normalization of von Neumann-Morgenstern utility given above, just the minimized probability of a high return followed by a low return:

$$\underline{p}\left(1-\overline{p}^H\right), \tag{13}$$

while the maxmin expected utility of holding B is just the minimized probability of a low return followed by a high return:

$$\left(1-\overline{p}\right)\underline{p}^L. \tag{14}$$

Now consider the maxmin expected utility of a roulette lottery delivering A with probability $\frac{1}{2}$ and B with probability $\frac{1}{2}$ to the recursive multiple-priors investor. It is:

$$\min_{p \in \left[\underline{p}, \overline{p}\right]} \left\{ \frac{1}{2} p\left(1-\overline{p}^H\right) + \frac{1}{2}\left(1-p\right)\underline{p}^L \right\}. \tag{15}$$

The following calculation reveals that it is not possible for all three of these maxmin expected utility values to be the same if $\overline{p} > \underline{p}$ (so that there is uncertainty, as well as risk, regarding the time-zero return), $\underline{p}^L > 0$ (ruling out dogmatic beliefs after a low time-0 return), and $\overline{p}^H < 1$ (ruling out dogmatic beliefs after a high time-0 return). If the maxmin expected utilities of holding A and holding B differ, there is no more to show.

Suppose, then, that they are the same, so that

$$\underline{p}\left(1-\overline{p}^H\right) = \left(1-\overline{p}\right)\underline{p}^L. \tag{16}$$

Under this condition, it is now shown that the maxmin expected utility of the roulette lottery delivering A with probability $\frac{1}{2}$ and B with probability $\frac{1}{2}$ will be greater than the maxmin expected utility of holding A or B. If $p \in \left(\underline{p}, \overline{p}\right]$, so that $p > \underline{p}$, then

$$p\left(1-\overline{p}^H\right) > \underline{p}\left(1-\overline{p}^H\right) \tag{17}$$

$$\left(1-p\right)\underline{p}^L \geq \left(1-\overline{p}\right)\underline{p}^L, \tag{18}$$

where the first, strict inequality follows from the fact that $\overline{p}^H < 1$ by assumption. But then

$$
\begin{aligned}
\frac{1}{2}p\left(1-\overline{p}^H\right)+\frac{1}{2}(1-p)\,\underline{p}^L &> \frac{1}{2}\underline{p}\left(1-\overline{p}^H\right)+\frac{1}{2}(1-\overline{p})\,\underline{p}^L && (19)\\
&= \underline{p}\left(1-\overline{p}^H\right) && (20)\\
&= (1-\overline{p})\,\underline{p}^L. && (21)
\end{aligned}
$$

It remains only to consider $p=\underline{p}$. But in this case,

$$
\left(1-\underline{p}\right)\underline{p}^L > (1-\overline{p})\,\underline{p}^L, \tag{22}
$$

since $\underline{p}^L > 0$ by assumption. Thus,

$$
\begin{aligned}
\frac{1}{2}\underline{p}\left(1-\overline{p}^H\right)+\frac{1}{2}\left(1-\underline{p}\right)\underline{p}^L &> \frac{1}{2}\underline{p}\left(1-\overline{p}^H\right)+\frac{1}{2}(1-\overline{p})\,\underline{p}^L && (23)\\
&= \underline{p}\left(1-\overline{p}^H\right) && (24)\\
&= (1-\overline{p})\,\underline{p}^L. && (25)
\end{aligned}
$$

Thus, if the maxmin expected utilities of holding A and B are the same, then the maxmin expected utility of the roulette lottery delivering A with probability $\frac{1}{2}$ and B with probability $\frac{1}{2}$ will be greater than the maxmin expected utility of holding A or B. Therefore, these three maxmin expected utility values cannot all be the same.

A preference for randomization is part of the definition of uncertainty aversion, but the key point is that the model-based multiple-priors investor discussed here does not experience uncertainty about the order of returns, given that there is one high and one low return. This is due to the model-based multiple-priors investor's belief in the i.i.d. model. In contrast, the recursive multiple-priors investor discussed here *does* experience uncertainty aversion about the order of returns, even given the fact that there is one high and one low return.

7 A Simple Two-period Example

In this section, an extremely simple model of portfolio choice is used to illustrate the basic differences between model-based multiple-priors and consequentialist multiple-priors approaches. Specifically, model-based multiple-priors is compared to recursive multiple-priors (Epstein and Schneider (2003)). There are two periods: $t = 0, 1$. Investment decisions are made at the beginning of each period, and the return for each period is realized at the end of that period. This corresponds to the event tree depicted in Figure 2.

Initial wealth is $W_0 > 0$. Utility is of the power form over final wealth:

$$
U(W_2) = \begin{cases} \frac{W_2^{1-\gamma}}{1-\gamma} & \text{if } \gamma \neq 1, \\ \ln(W_2) & \text{if } \gamma = 1 \end{cases}, \tag{26}
$$

where W_2 denotes wealth at the *end* of period 1 (or the beginning of period 2). In the continuous-time models analyzed in Knox (2003), intermediate consumption is

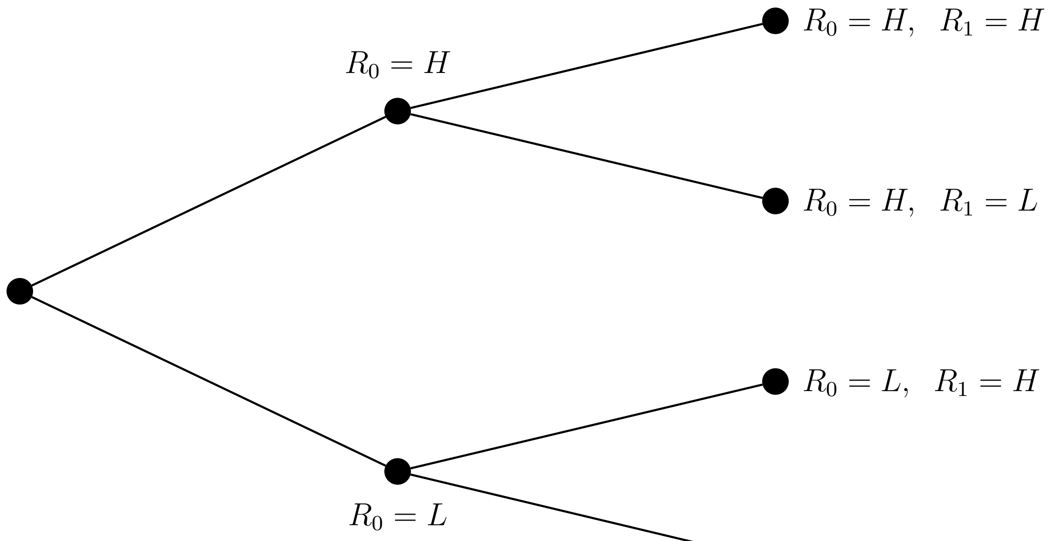

Figure 2: **The Two-period Binomial Model**
This figure depicts an event-tree representation of the two-period model with a binomial risky asset, which is described in Section 7.

considered; it could be included here, but is omitted for the sake of simplicity. There is one riskless asset, and the (gross) riskless rate is denoted $R_f > 1$. There is one risky, and uncertain, asset; in each period, the gross return on this asset takes on one of two possible values. The higher of these two values is denoted H and the lower is denoted L, while the (uncertain) gross return on the risky asset in period t is denoted R_t. In order to avoid arbitrage, $H > R_f > L$ is assumed. At each of the two periods, the investor chooses how much to invest in the risky (and uncertain) asset.

Under the conditions on preferences given by Gilboa and Schmeidler (1989), the investor has a *set* of (subjective) prior probability distributions on the four possible pairs of returns. This set is closed and convex. The investor evaluates any potential portfolio choice by calculating the *minimum* expected utility of that portfolio choice, where the minimum is taken over the set of priors.

In the recursive multiple-priors approach, the set of priors is rectangular:

$$\Pr(R_0 = H) \in \left[\underline{p}, \overline{p}\right] \tag{27}$$

$$\Pr(R_1 = H | R_0 = H) \in \left[\underline{p}^H, \overline{p}^H\right] \tag{28}$$

$$\Pr(R_1 = H | R_0 = L) \in \left[\underline{p}^L, \overline{p}^L\right]. \tag{29}$$

This set of priors is shown in Figure 3.

For a given portfolio choice rule, minimization takes place separately over each of the three intervals of probabilities. This separate minimization over each of the three intervals is a cornerstone of recursive multiple-priors, and must hold whether the interval endpoints are set according to the type of learning advocated by Epstein and Schneider (2002) or according to the "κ-ignorance" specification of Chen and Epstein (2002) (which further specializes the above to sets in which $\underline{p} = \underline{p}^H = \underline{p}^L$ and $\overline{p} = \overline{p}^H = \overline{p}^L$, so that the three returns have the same ranges of uncertainty).

To see the contrast between model-based multiple-priors and a consequentialist approach to uncertainty aversion in dynamic portfolio choice, consider the following investment problem. The set of priors on the probability of a high return is a set of mixtures of two point masses:

$$\Pi \equiv \left\{ \begin{array}{ll} \Pr(R_t = H) = 0.75 & \text{with probability } q \\ \Pr(R_t = H) = 0.25 & \text{with probability } 1-q \end{array} : q \in [0.1, 0.9] \right\}. \tag{30}$$

Note that Π is convex and closed. There is a two-period binomial likelihood for the pair of returns on the risky asset given the probability of a high return, as described above. To make this example fully concrete, suppose that $H = 1.25 = \frac{1}{L}$, so that $L = 0.8$, and that $R_f = \frac{1}{2}H + \frac{1}{2}L = 1.025$ (the values chosen are convenient, but not essential; this is not a knife-edge situation). Further suppose that $\gamma = 2$ (again, this is not essential).

Using the minimax theorem as in Chamberlain (2000) and Knox (2003), the portfolio choice problem of a model-based multiple-priors investor may be solved as follows: solve the standard Bayesian dynamic portfolio choice problem for each prior $\pi \in \Pi$ on p. This maps the set of priors to a set of date-zero utilities. Choose the minimal utility from that set; the prior corresponding to that minimal utility is the least-favorable prior. The model-based multiple-priors investor makes portfolio

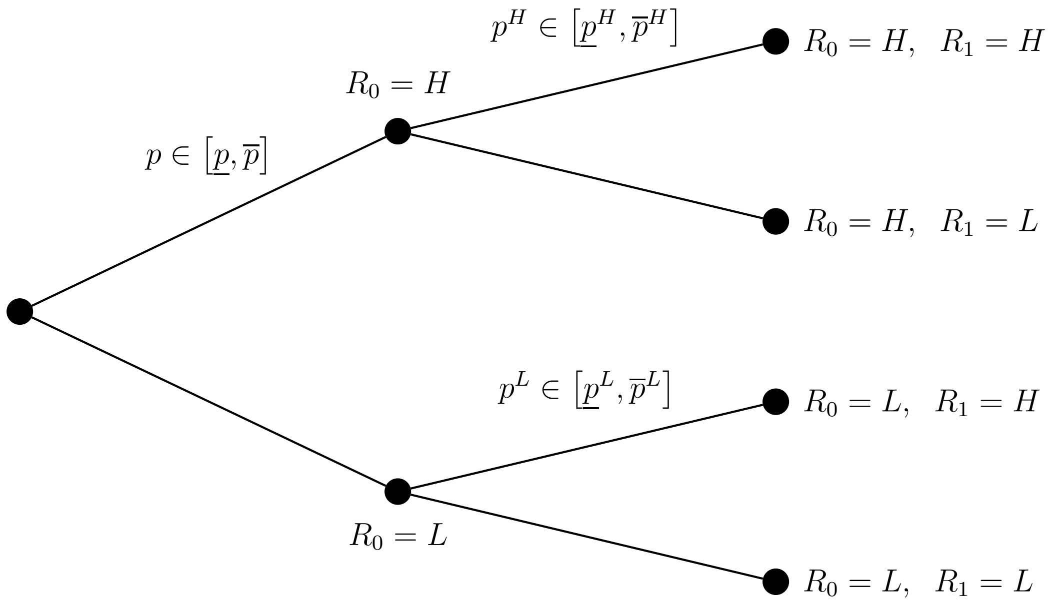

Figure 3: **The Recursive Multiple-Priors Rectangular Set of Priors in the Two-period Binomial Model**
Under the Epstein-Schneider axioms, the investor has a rectangular set of priors, which is shown in the figure above. The quantity p shown in the figure is the probability that the time-0 return on the risky asset is H. Because of uncertainty aversion, this probability is not fixed: the investor is only willing to specify that it is in some interval, denoted $[\underline{p}, \overline{p}]$. Likewise, the probability p^H is the probability that the time-1 return on the risky asset is H, given that the time-0 return on the risky asset was H. As with the probability p, the probability p^H is only specified to be within some interval, which is denoted $[\underline{p}^H, \overline{p}^H]$. Finally, the probability p^L is the probability that the time-1 return on the risky asset is H, given that the time-0 return on the risky asset was L. It is also known only up to some interval, denoted $[\underline{p}^L, \overline{p}^L]$ in the figure above.

choices as a Bayesian would, if that Bayesian's prior happened to be the least-favorable prior in the set Π. For a given q, the value function at time zero is:

$$\begin{aligned} & J_q (W_0, 0) \\ = & -W_0^{-1} R_f^{-2} (H-L)^{-2} \begin{pmatrix} \sqrt{\frac{1}{4}+\frac{1}{2}q}\sqrt{\frac{1+8q}{4+8q}} (R_f - L) \\ +\sqrt{\frac{1}{4}+\frac{1}{2}q}\sqrt{\frac{3}{4+8q}}\sqrt{R_f - L}\sqrt{H-R_f} \\ +\sqrt{\frac{3}{4}-\frac{1}{2}q}\sqrt{\frac{3}{12-8q}}\sqrt{R_f - L}\sqrt{H-R_f} \\ +\sqrt{\frac{3}{4}-\frac{1}{2}q}\sqrt{\frac{9-8q}{12-8q}} (H-R_f) \end{pmatrix}^2 \quad (31) \\ = & -W_0^{-1} R_f^{-2} (H-L)^{-2} \left(\frac{1}{4} \begin{pmatrix} \sqrt{1+8q} (R_f - L) \\ +\sqrt{3}\sqrt{R_f - L}\sqrt{H-R_f} \\ +\sqrt{3}\sqrt{R_f - L}\sqrt{H-R_f} \\ +\sqrt{9-8q} (H-R_f) \end{pmatrix} \right)^2 . \quad (32) \end{aligned}$$

The above expression is minimized over $q \in [0.1, 0.9]$ by any $q \in [0.1, 0.9]$ which maximizes

$$\begin{aligned} G(q) \equiv & \sqrt{1+8q} (R_f - L) \quad (33) \\ & +2\sqrt{3}\sqrt{R_f - L}\sqrt{H-R_f} \\ & +\sqrt{9-8q} (H-R_f) . \end{aligned}$$

The unique maximizer of G, and thus the unique q minimizing the value function, is:

$$\begin{aligned} q^{LF} &= \frac{9 (R_f - L)^2 - (H-R_f)^2}{8 \left[(R_f - L)^2 + (H-R_f)^2 \right]} \quad (34) \\ &= \frac{9 \left(\frac{9}{40}\right)^2 - \left(\frac{9}{40}\right)^2}{8 \left[\left(\frac{9}{40}\right)^2 + \left(\frac{9}{40}\right)^2 \right]} \quad (35) \\ &= \frac{1}{2}. \quad (36) \end{aligned}$$

The prior expected probability of a high return in the next period at time zero is thus $E[p] = \frac{1}{2}$, but the posterior expected probability of a high return in the next period after a high return has been observed is $E[p|\, R_0 = H] = \frac{5}{8}$, while the posterior expected probability of a high return in the next period after a low return has been observed is $E[p|\, R_0 = L] = \frac{3}{8}$. At time zero, the investor will not hold or short the risky asset; after a high return, the investor will hold the risky asset at time one; after a low return, the investor will short the risky asset at time one.

The minimized value function, found by substituting $q^{LF} = \frac{1}{2}$ and the given values of H, L, and R_f into the formula above and simplifying, is:

$$J^{LF} (W_0, 0) = -W_0^{-1} R_f^{-2} \left(\frac{6.35 + 3\sqrt{3}}{16} \right) \quad (37)$$

$$\approx \quad -W_0^{-1}R_f^{-2} \times 0.72163453. \tag{38}$$

This will be of interest when the model-based multiple-priors method is contrasted with consequentialist multiple-priors methods.

To examine the behavior of a consequentialist multiple-priors investor, it is convenient (and not overly restrictive) to focus on an investor whose preferences are described by the recursive multiple-priors theory of Epstein and Schneider (2003). In Epstein and Schneider (2002), a method of applying recursive multiple-priors to learn about a parameter of a model is clearly spelled out. Below, the method explicated by Epstein and Schneider (2002) is applied to the current setting. Recursive multiple-priors is, by axiomatic design, consequentialist. Because it is axiomatized without reference to a recursive domain of choice, it is simpler to work with than alternative consequentialist multiple-priors theories such as those formulated by Klibanoff (1995) and Wang (2003).

In order to apply the Epstein and Schneider (2002) method, one must "rectangularize" the set of distributions on returns implied by the likelihood of returns given p and the set of priors on p. This is accomplished by, at each date and in each state of the world (that is, for each possible return history), considering a set of posterior distributions for p obtained by updating the set of time-zero prior distributions on p. At each date and in each state, this set of posteriors implies a set of predictive distributions on returns, and the overall set of distributions on the entire sequence of returns is built up from these sets of one-step-ahead predictive distributions. These operations will be performed explicitly below. The process of "rectangularizing" always produces a set of distributions at least as large as the original set: it constructs the smallest rectangular set of distributions (see Epstein and Schneider (2003)) that contains the original set of distributions.

An important property of all consequentialist multiple-priors theories, including recursive multiple-priors, is that the minimization over the set of distributions, for a fixed horse lottery (*e. g.*, for a fixed portfolio choice strategy), can be performed recursively using dynamic programming. This is a direct consequence of the fact that the set of one-step-ahead predictive distributions at any date and in any state is the same regardless of how minimization might proceed at other dates or in other states. Note that this is the *minimization* portion of the maxmin expected utility problem: in order to solve the full problem, there will need to be both a minimization (over the set of one-step-ahead predictive distributions) and a maximization (over the choice variables) at each node in the event tree. Of course, the set of priors will typically have to be rectangularized, and thus enlarged, in order to employ this approach, so it is not true, in general, that consequentialist methods will be more tractable than alternatives such as model-based multiple-priors. Further, it is not true that only consequentialist methods permit the use of dynamic programming; dynamic programming can be used to maximize expected utility, for a fixed prior, in model-based multiple-priors. Minimization in model-based multiple-priors is then performed over the resulting time-zero value functions.

Applying the Epstein and Schneider (2002) method to the simple example being explored here leads to the consideration of different values of q at time zero, at time one after a high return, and at time one after a low return. Label these values q_0, q_H, and q_L respectively. The recursive multiple-priors investor behaves as though q might change depending on the date and the return history. Under the rectangularized set of distributions $q_0, q_H, q_L \in [0.1, 0.9]$, but it is *not* required that all three of these variables take on the same value. Quite the opposite: minimization occurs separately over $q_0 \in [0.1, 0.9]$, $q_H \in [0.1, 0.9]$, $q_L \in [0.1, 0.9]$. With the potential for differences in q values, the time-zero value function for a Bayesian investor would be:

$$\begin{aligned} & J_{q,q_H,q_L}(W_0, 0) \\ = \; & -W_0^{-1} R_f^{-2} (H-L)^{-2} \left(\begin{array}{c} \sqrt{\frac{1}{4} + \frac{1}{2} q_0} \sqrt{\frac{1+8q_H}{4+8q_H}} (R_f - L) \\ + \sqrt{\frac{1}{4} + \frac{1}{2} q_0} \sqrt{\frac{3}{4+8q_H}} \sqrt{R_f - L} \sqrt{H - R_f} \\ + \sqrt{\frac{3}{4} - \frac{1}{2} q_0} \sqrt{\frac{3}{12-8q_L}} \sqrt{R_f - L} \sqrt{H - R_f} \\ + \sqrt{\frac{3}{4} - \frac{1}{2} q_0} \sqrt{\frac{9-8q_L}{12-8q_L}} (H - R_f) \end{array} \right)^2 . \end{aligned} \quad (39)$$

The expression above shows that the minimization problem in the sort of learning advocated by Epstein and Schneider (2002) will typically be higher-dimensional than the model-based multiple-priors minimization problem, which suggests that model-based multiple-priors investors' problems may be more tractable than those of investors who learn as in Epstein and Schneider (2002). This suggestion is born out by the closed-form solutions to a class of model-based multiple-priors continuous-time consumption and portfolio choice problems given in Knox (2003).

Minimizing the above expression over q_0, q_H, and q_L subject to the constraints laid out above yields the least-favorable values of these variables:

$$q_0^{LF} = \frac{1}{2} \quad (40)$$

$$q_H^{LF} = \frac{1}{4} \quad (41)$$

$$q_L^{LF} = \frac{3}{4}. \quad (42)$$

Under these least-favorable values of q_0, q_H, and q_L, the recursive multiple-priors investor never holds or shorts the risky asset at any time or after any return history. Upon substituting these values into the time-zero value function, the recursive multiple-priors investor's time-zero value function is obtained:

$$J_{RMP}^{LF}(W_0, 0) = -W_0^{-1} R_f^{-2}. \quad (43)$$

This is sensible: wealth grows at the riskless rate for two periods, since the investor never holds or shorts the risky asset. It should be emphasized that, if q were constrained to an interval that was a strict subset of $\left[\frac{1}{4}, \frac{3}{4}\right]$, the recursive multiple-priors investor would hold the risky asset at time one after a high return and would short the risky asset at time one after a low return. However, since rectangularizing the set of distributions used by the model-based multiple-priors investor always yields a set at least as large as the original set, the recursive multiple-priors investor is always at least as uncertainty-averse as the model-based multiple-priors investor. The recursive multiple-priors investor will never hold more of the risky asset after a high return nor short more of the risky asset after a low return than the model-based multiple-priors investor whose set of priors was rectangularized.

The greater uncertainty aversion of the recursive multiple-priors investor has significant implications for welfare. Comparing the least favorable time-zero value function of the model-based multiple-priors investor with that of the recursive multiple-priors investor, it is clear that the maxmin expected utility of the model-based multiple-priors investor is greater (recall that the utilities are negative). In fact, in order to experience maxmin expected utility equal to that of the model-based multiple-priors investor, the recursive multiple-priors investor would need to have initial wealth that was approximately 38.5743 percent greater than the initial wealth of the model-based multiple-priors investor. If the model-based multiple-priors investor began with initial wealth of $500,000$ dollars, the recursive multiple-priors investor would require an initial wealth of approximately $692,872$ dollars to equalize the certain equivalents for the portfolio choice problem.

8 Conclusion

Most asset returns are uncertain, not merely risky: investors do not know the probabilities of different possible future returns. The Ellsberg paradox (Ellsberg (1961)) suggests that investors are averse to uncertainty, as well as to risk. This paper axiomatized the dynamic portfolio and consumption choice behavior of an uncertainty-averse (as well as risk-averse) investor who tries to learn from historical data. The theory developed, *model-based multiple-priors*, relaxes the assumption of consequentialism, which has been imposed in existing axiomatic studies of uncertainty-averse dynamic choice. Examples were given to show that consequentialism, the property that counterfactuals are ignored, can be problematic when combined with uncertainty aversion. A model-based multiple-priors analog of de Finetti's statistical representation theorem was proven; in contrast, consequentialism combines with multiple priors to rule out prior-by-prior exchangeability. A simple dynamic portfolio choice problem illustrated the contrast between a model-based multiple-priors investor and a consequentialist multiple-priors investor. Building on the foundations provided here, a class of continuous-time portfolio and consumption choice problems under learning and uncertainty aversion, including problems in which the investor is uncertain about the accuracy of an asset pricing model, is solved in closed form for a model-based multiple-priors investor in a companion paper, Knox (2003).

A model-based multiple-priors investor has a set of prior distributions on the parameters of some economic model, but a single likelihood (or conditional distribution of the data given the model parameters). Future work will explore frameworks in which there are multiple likelihoods, as well as multiple distributions on the model parameters. The axioms used here, without the specialization of Section 5, can be used to justify such frameworks when the partitions involved are based on the values of model parameters. Research in the future will also focus on evaluating the empirical implications of model-based multiple-priors, particularly for equilibrium asset prices. To accomplish this goal, it will be necessary to characterize general equilibrium when agents' preferences are described by model-based multiple-priors.

Appendix

This Appendix contains proofs of the propositions and theorems stated in the text. To avoid confusion between equations in the text of the paper and equations in this Appendix, the equations in this Appendix are numbered (A.1), (A.2), etc. Throughout the proofs below, $\mathcal{A} = \{A_1, \ldots, A_k\}$ is a partition of S, and given any $A_i \in \mathcal{A}$, we define $(f; g)_i \equiv \begin{cases} f & \text{for } s \in A_i, \\ g & \text{for } s \in A_i^C. \end{cases}$.

Proofs

Lemma 1 *Under Axiom 1 relative to $\mathcal{A}$, and under Axiom 3, $\forall f, g, h \in L_0$, $\forall l \in L_c$, $\forall i \in \{1, \ldots, k\}$, and $\forall \alpha \in (0, 1)$,*

$$(f; h)_i \succsim (g; h)_i \Leftrightarrow (\alpha f + (1 - \alpha) l; h)_i \succsim (\alpha g + (1 - \alpha) l; h)_i .$$

Proof of Lemma 1: By definition, $(f; h)_i$ and $(g; h)_i$ are identical on A_i^C. Consider $m = (l; h)_i$, which by definition is constant on A_i. By Axiom 1 relative to A_i, $\forall \alpha \in (0, 1)$, $(f; h)_i \succ (g; h)_i \Leftrightarrow \alpha (f; h)_i + (1 - \alpha) m \succ \alpha (g; h)_i + (1 - \alpha) m$. But $\forall s \in S$, we have that

$$\begin{aligned} & (\alpha (f; h)_i + (1 - \alpha) m) (s) \\ = \ & (\alpha f + (1 - \alpha) l; \alpha h + (1 - \alpha) h)_i (s) \\ = \ & (\alpha f + (1 - \alpha) l; h)_i (s) . \end{aligned}$$

Thus, by monotonicity, $\alpha (f; h)_i + (1 - \alpha) m \sim (\alpha f + (1 - \alpha) l; h)_i$. Since exactly analogous reasoning can be applied to $\alpha (g; h)_i + (1 - \alpha) m$, we also have that $\alpha (g; h)_i + (1 - \alpha) m \sim (\alpha g + (1 - \alpha) l; h)_i$. Then, by transitivity, we have that $\forall \alpha \in (0, 1)$, $f \succ g \Leftrightarrow (\alpha f + (1 - \alpha) l; h)_i \succ (\alpha g + (1 - \alpha) l; h)_i$. Q.E.D.

Lemma 2 *Under Axiom 1 relative to $\mathcal{A}$, and under Axiom 3, $\forall f, g, h, m \in L_0$, $\forall i \in \{1, \ldots, k\}$, and $\forall \alpha \in (0, 1)$,*

$$(f; h)_i \succ (g; h)_i \Leftrightarrow (f; \alpha h + (1 - \alpha) m)_i \succ (g; \alpha h + (1 - \alpha) m)_i .$$

Proof of Lemma 2: Let $l \in L_c$, so that l is a roulette lottery. We have that $\forall \alpha \in (0, 1)$,

$$\begin{aligned} & (f; h)_i \succ (g; h)_i \\ \Leftrightarrow \ & (\alpha f + (1 - \alpha) l; \alpha h + (1 - \alpha) m)_i \succ (\alpha g + (1 - \alpha) l; \alpha h + (1 - \alpha) m)_i \\ \Leftrightarrow \ & (f; \alpha h + (1 - \alpha) m)_i \succ (g; \alpha h + (1 - \alpha) m)_i . \end{aligned}$$

The first equivalence follows from applying Axiom 1 relative to A_i, where the mixing lottery is $(l; m)_i$ (which is constant on A_i since l is a roulette lottery). The second equivalence follows from Lemma 1 (which we apply with the act on A_i^C being $\alpha h + (1 - \alpha) m$). Q.E.D.

Lemma 3 *Under Axiom 1 relative to $\mathcal{A}$, and under Axiom 3, $\forall i \in \{1, \ldots, k\}$,*

$$\forall f, g, h, m \in L_0, \ \ (f; h)_i \succ (g; h)_i \Leftrightarrow (f; m)_i \succ (g; m)_i$$

Proof of Lemma 3: Suppose that the statement of the lemma does not hold. Then $\exists f, g, h, m \in L_0$ such that $(f; h)_i \succ (g; h)_i$ but $(f; m)_i \precsim (g; m)_i$. Since $(f; h)_i \succ (g; h)_i$, Lemma 2 implies that $\left(f; \frac{1}{2}h + \frac{1}{2}m\right)_i \succ \left(g; \frac{1}{2}h + \frac{1}{2}m\right)_i$. However, since $(f; m)_i \precsim (g; m)_i$, Lemma 2 also implies that $\left(f; \frac{1}{2}h + \frac{1}{2}m\right)_i \precsim \left(g; \frac{1}{2}h + \frac{1}{2}m\right)_i$. This is a contradiction, so the statement of the lemma must hold. Q.E.D.

Proof of Theorem 1: First we prove that Axiom 4 implies Axiom 1 (each being relative to $\mathcal{A}$). Suppose $f, g \in L_0$ are such that $f(s) = g(s) \ \ \forall s \in A_i^C$, and that $f \succ g$. Then A_i is not a null set, since A_i null and $f(s) = g(s) \ \ \forall s \in A_i^C$ would imply $f \sim g$. $\forall j \neq i, \ f(s) = g(s) \ \ \forall s \in A_j$, since the $A_j, \ j \neq i$, partition A_i^C. Thus, by the "focus" portion of Axiom 4, $f \sim_j g \ \ \forall j \neq i$. If $f \precsim_i g$, then $f \precsim_l g \ \ \forall l \in \{1, \ldots, k\}$ (since $f \sim_l g \Rightarrow f \precsim_l g$ by definition), so by the "consistency" portion of Axiom 4 we would have $f \precsim g$. But $f \succ g$, so $f \succ_i g$. Given h that is constant on A_i, $\exists y \in Y$ such that $h(s) = y \ \ \forall s \in A_i$. By the "focus" portion of Axiom 4, and letting $l \in L_c$ be such that $l(s) = y \ \ \forall s \in S$, $l \sim_i h$, since l and h are equal (with value y) at each element of S. By the "multiple priors" portion of Axiom 4, and the "certainty independence" portion of Axiom 3, $\forall \alpha \in (0, 1), \ \alpha f + (1 - \alpha) l \succ_i \alpha g + (1 - \alpha) l$. Since $(\alpha f + (1 - \alpha) l)(s) = (\alpha f + (1 - \alpha) h)(s) \ \ \forall s \in A_i$, the "focus" portion of Axiom 4 implies that $\alpha f + (1 - \alpha) l \sim_i \alpha f + (1 - \alpha) h$. Likewise, since $(\alpha g + (1 - \alpha) l)(s) = (\alpha g + (1 - \alpha) h)(s) \ \ \forall s \in A_i$, the "focus" portion of Axiom 4 implies that $\alpha g + (1 - \alpha) l \sim_i \alpha g + (1 - \alpha) h$. The transitivity of $\succsim_i$ is implied by Axiom 4 and the "weak order" portion of Axiom 3. By this transitivity, then,

$$\begin{aligned} & \alpha f + (1 - \alpha) h \sim_i \alpha f + (1 - \alpha) l \\ \succ_i \ \ & \alpha g + (1 - \alpha) l \sim_i \alpha g + (1 - \alpha) h, \end{aligned}$$

each step of which is proven above, implies that $\forall \alpha \in (0, 1), \ \alpha f + (1 - \alpha) h \succ_i \alpha g + (1 - \alpha) h$.

Now, since $\forall j \neq i, \ f(s) = g(s) \ \ \forall s \in A_j$ as noted above, we have that $\forall j \neq i, \ \forall \alpha \in (0, 1), \ \alpha f(s) + (1 - \alpha) h(s) = \alpha g(s) + (1 - \alpha) h(s) \ \ \forall s \in A_j$. Thus, by the "focus" portion of Axiom 4, $\forall j \neq i, \ \forall \alpha \in (0, 1), \ \alpha f + (1 - \alpha) h \sim_j \alpha g + (1 - \alpha) h$. Now, since A_i was shown to be non-null above, we can invoke the "if, in addition" portion of the "consistency" part of Axiom 4 to conclude that $\forall \alpha \in (0, 1), \ \alpha f + (1 - \alpha) h \succ \alpha g + (1 - \alpha) h$.

From the above, we have that $\forall \alpha \in (0, 1), \ f \succ g \Rightarrow \alpha f + (1 - \alpha) h \succ \alpha g + (1 - \alpha) h$. We must now prove the other part of the assertion made by Axiom 1. We want to show the converse of what we have just proven: $\forall \alpha \in (0, 1), \ \alpha f + (1 - \alpha) h \succ \alpha g + (1 - \alpha) h \Rightarrow f \succ g$.

Given $\alpha \in (0, 1)$, suppose that f, g, h are as described in the previous section of the proof, and that $\alpha f + (1 - \alpha) h \succ \alpha g (1 - \alpha) h$. The steps to follow are quite similar to those above, but we include them for the sake of completeness. A_i cannot be a null set; if it were, then (since $\alpha f + (1 - \alpha) h$ and $\alpha g + (1 - \alpha) h$ are identical on its complement) $\alpha f+(1-\alpha)\,h \sim \alpha g+(1-\alpha)\,h$ would have to hold by the definition of a null set. Now, $\forall j \neq i,\ (\alpha f+(1-\alpha)\,h)\,(s) = (\alpha f+(1-\alpha)\,h)\,(s) \quad \forall s \in A_j$, since the $A_j,\ j \neq i$, partition A_i^C. Thus, by the "focus" portion of Axiom 4, $\alpha f+(1-\alpha)\,h \sim_j \alpha g+(1-\alpha)\,h \ \ \forall j \neq i$. If we had $\alpha f+(1-\alpha)\,h \precsim_i \alpha g+(1-\alpha)\,h$, then we would have $\alpha f+(1-\alpha)\,h \precsim_l \alpha g+(1-\alpha)\,h \ \ \forall l \in \{1,\ldots,k\}$ (since $\alpha f+(1-\alpha)\,h \sim_l \alpha g+(1-\alpha)\,h \Rightarrow \alpha f+(1-\alpha)\,h \precsim_l \alpha g+(1-\alpha)\,h$ by definition), so by the "consistency" portion of Axiom 4 we would have $\alpha f+(1-\alpha)\,h \precsim \alpha g+(1-\alpha)\,h$, which does not hold. Thus, we must have $\alpha f+(1-\alpha)\,h \succ_i \alpha g+(1-\alpha)\,h$. We have by definition that $\exists y \in Y$ such that $h\,(s)=y \ \ \forall s \in A_i$. Letting $l \in L_c$ be such that $l\,(s)=y \ \ \forall s \in S$, we have that $(\alpha f+(1-\alpha)\,l)\,(s) = (\alpha f+(1-\alpha)\,h)\,(s) \ \ \forall s \in A_i$. Thus, the "focus" portion of Axiom 4 implies that $\alpha f+(1-\alpha)\,l \sim_i \alpha f+(1-\alpha)\,h$. Likewise, $(\alpha g+(1-\alpha)\,l)\,(s) = (\alpha g+(1-\alpha)\,h)\,(s) \quad \forall s \in A_i$. Thus, the "focus" portion of Axiom 4 implies that $\alpha g+(1-\alpha)\,l \sim_i \alpha g+(1-\alpha)\,h$. The transitivity of $\succsim_i$ is implied by Axiom 4 and the "weak order" portion of Axiom 3. By this transitivity and the above observations,

$$\begin{aligned} & \alpha f+(1-\alpha)\,l \sim_i \alpha f+(1-\alpha)\,h \\ \succ_i \ & \alpha g+(1-\alpha)\,h \sim_i \alpha g+(1-\alpha)\,l \end{aligned}$$

implies that $\alpha f+(1-\alpha)\,l \succ_i \alpha g+(1-\alpha)\,l$. By the "multiple priors" portion of Axiom 4, and the "certainty independence" portion of Axiom 3, $\alpha f+(1-\alpha)\,l \succ_i \alpha g+(1-\alpha)\,l \Rightarrow f \succ_i g$. We therefore have that $f \succ_i g$.

Now, since $\forall j \neq i,\ f\,(s) = g\,(s) \quad \forall s \in A_j$ as noted above, we have that $\forall j \neq i,\ f \sim_j g$ by the "focus" portion of Axiom 4. Now, since A_i was shown to be non-null above, we can invoke the "if, in addition" portion of the "consistency" part of Axiom 4 to conclude that $f \succ g$.

The above reasoning proves that the restricted independence axiom holds relative to A_i. However, the choice of $i \in \{1,\ldots,k\}$ was completely arbitrary. Thus, we have proven that the restricted independence axiom holds relative to *any* $A_i \in \mathcal{A}$. But then, by definition, Axiom 1 holds relative to $\mathcal{A}$.

We now need to prove that, in the presence of Axiom 3, Axiom 1 (relative to $\mathcal{A}$) implies Axiom 4 (also relative to $\mathcal{A}$). Recall the following notation: given any $A_i \in \mathcal{A}$, let $(f;g)_i = \begin{cases} f & \text{for } s \in A_i, \\ g & \text{for } s \in A_i^C. \end{cases}$.

Given $A_i \in \mathcal{A}$, define the conditional preference relation $\succsim_i$ by:

$$f \succsim_i g \Leftrightarrow \exists h \in L_0 \text{ such that } (f;h)_i \succsim (g;h)_i\,.$$

Lemma 3 shows that this results in $\succsim_i$ being well-defined, since $\forall f,g,h,m \in L_0$, $(f;h)_i \succsim (g;h)_i \Leftrightarrow (f;m)_i \succsim (g;m)_i$.

First we demonstrate that $\succsim_i$ satisfies the "focus" property of Axiom 4. If $f\,(s) = g\,(s)\,\forall s \in A_i$, then $(f;h)_i\,(s) = (g;h)_i\,(s)\,\forall s \in S$ and $\forall h \in L_0$. Thus, $(f;h)_i \sim (g;h)_i$ $\forall h \in L_0$. This implies, by definition, that $f \sim_i g$.

Now we verify that $\succsim_i$ satisfies each of the portions of Axiom 3. These follow because $\succsim$ satisfies Axiom 3 and by the definition of $\succsim_i$. First we show that $\succsim_i$ is a weak order (that is, that $\succsim_i$ is complete and transitive). Suppose that $f \succsim_i g$ and $g \succsim_i h$. Then $\exists m \in L_0$ such that $(f;m)_i \succsim (g;m)_i$ and $\exists n \in L_0$ such that

$(g;n)_i \succsim (h;n)_i$, by the definition of $\succsim_i$. By Lemma 3, $(g;n)_i \succsim (h;n)_i$ implies $(g;m)_i \succsim (h;m)_i$. Thus, by the transitivity of $\succsim$ (which is part of Axiom 3), we have that $(f;m)_i \succsim_i (h;m)_i$. But then, by the definition of $\succsim_i$, we have that $f \succsim_i h$. This proves that $\succsim_i$ is transitive. To see that it is complete, suppose that it is not. Then $\exists f, g \in L_0$ such that *neither* $f \succsim_i g$ *nor* $f \precsim_i g$. Given any $m \in L_0$, we would have (by the definition of $\succsim_i$) that *neither* $(f;m)_i \succsim (g;m)_i$ *nor* $(f;m)_i \precsim (g;m)_i$. However, this contradicts the completeness of $\succsim$, as implied by Axiom 3, so $\succsim_i$ must be complete.

Lemma 1 shows that $\succsim_i$ satisfies the "certainty independence" portion of Axiom 3.

We proceed to verify the "continuity" portion of Axiom 3 for $\succsim_i$. Suppose that $f \succ_i g \succ_i h$. Then $\exists m, n \in L_0$ such that $(f;m)_i \succ (g;m)_i$ and $(g;n)_i \succ (h;n)_i$. By Lemma 3, $(g;m)_i \succ (h;m)_i$. Then we have $(f;m)_i \succ (g;m)_i \succ (h;m)_i$, and since $\succsim$ satisfies Axiom 3 (and its "continuity" portion in particular), $\exists \alpha, \beta \in (0,1)$ such that $\alpha (f;m)_i + (1-\alpha)(h;m)_i \succ (g;m)_i \succ \beta (f;m)_i + (1-\beta)(h;m)_i$. By the definition of $\succsim_i$, this implies that $\alpha f + (1-\alpha) h \succ_i g \succ_i \beta f + (1-\beta) h$, verifying the continuity property.

We now show that $\succsim_i$ satisfies the "monotonicity" portion of Axiom 3. Suppose that $f, g \in L_0$ are such that $f(s) \succsim_i g(s) \, \forall s \in A_i$. Then, for any $h \in L_0$, $(f;h)_i (s) \succsim (g;h)_i (s) \quad \forall s \in S$. Thus, by the monotonicity of $\succsim$ (guaranteed by Axiom 3), we have that $(f;h)_i \succsim (g;h)_i$. By the definition of $\succsim_i$, this implies that $f \succsim_i g$, verifying the monotonicity property.

Consider the "uncertainty aversion" portion of Axiom 3. Suppose that $f, g \in L_0$ satisfy $f \sim_i g$. Then, by definition of $\succsim_i$, $\exists h \in L_0$ such that $(f;h)_i \sim (g;h)_i$. Since $\succsim$ satisfies uncertainty aversion (by Axiom 3), this implies that $\alpha (f;h)_i + (1-\alpha)(g;h)_i \succsim (g;h)_i \quad \forall \alpha \in (0,1)$. By the definition of $\succsim_i$, this, in turn, implies that $\alpha f + (1-\alpha) g \succsim_i g \quad \forall \alpha \in (0,1)$, confirming that $\succsim_i$ satisfies the uncertainty aversion property.

To complete our demonstration that $\succsim_i$ satisfies an appropriately-modified Axiom 3, we need only show that if A_i is not a null set of $\succsim$, then it is "non-degenerate": $\exists f, g \in L_0$ such that $f \succ_i g$. By the definition of a null set, A_i non-null implies that $\exists (f;h)_i, (g;h)_i$ such that either $(f;h)_i \succ (g;h)_i$ or $(f;h)_i \prec (g;h)_i$ (otherwise, there would be indifference between any two acts agreeing on A_i^C; that is, A_i would be null). By the definition of $\succsim_i$, this implies that $\exists f, g \in L_0$ such that $f \succ_i g$ or $g \succ_i f$. Either possibility shows the desired non-degeneracy.

Since A_i was selected completely arbitrarily in the above argument, our conclusions hold $\forall i \in \{1, \ldots, k\}$. Thus, the conditional preference orderings $\succsim_i$, $i \in \{1, \ldots, k\}$ satisfy the "focus" and "multiple priors" portions of Axiom 4. It remains only to prove that they satisfy the "consistency" portion of Axiom 4. To do so, suppose that $f \succsim_i g \quad \forall i \in \{1, \ldots, k\}$. Then by definition we have that $\forall i \in \{1, \ldots, k\}$, $\exists h_i \in L_0$ such that $(f;h_i)_i \succsim (g;h_i)_i$. In fact, Lemma 3 proves that this is equivalent to: $\forall i \in \{1, \ldots, k\}$ and $\forall h_i \in L_0$, $(f;h_i)_i \succsim (g;h_i)_i$. Since we are thus free to choose the h_i, let

$$h_i(s) = \begin{cases} f(s) & \text{for } s \in A_j \text{ with } j < i, \\ g(s) & \text{for } s \in A_j \text{ with } j \geq i \end{cases}$$

for all $i \in \{1, \ldots, k\}$. Then we have, for $i \in \{2, \ldots, k\}$, $(g;h_i)_i (s) = h_i(s) =$

$(f; h_{i-1})_{i-1}(s)$. This can be seen by considering the values of each of the above expressions on each $A_j \in \mathcal{A}$. Now, since $(f; h_i)_i \succsim (g; h_i)_i$ for each $i \in \{1, \ldots, k\}$, we can use the equality above to conclude that $h_{i+1} = (f; h_i)_i \succsim (g; h_i)_i = h_i$ for $i \in \{1, \ldots, k-1\}$, and thus that $h_{i+1} \succsim h_i$ for $i \in \{1, \ldots, k-1\}$. Applying transitivity repeatedly, this implies that $h_k \succsim h_1$. By definition, $h_1(s) = g(s) \ \ \forall s \in S$. We also have $f(s) = (f; h_k)_k(s) \ \ \forall s \in S$, $(f; h_k)_k \succsim (g; h_k)_k$, and $(g; h_k)_k(s) = h_k(s) \forall s \in S$. Combining these facts, we obtain $f \succsim h_k$. A final application of transitivity yields $f \succsim g$, and thus verifies the main part of the "consistency" portion of Axiom 3.

To confirm that the "if, in addition" part of the "consistency" condition holds, observe that if, in addition to $f \succsim_i g \ \ \forall i \in \{1, \ldots, k\}$, we also have $f \succ_j g$ for some j such that A_j is not a null set, then one of the weak preference relations in the chain of preference that we constructed above is actually a strict preference relation, so that repeated applications of transitivity yield a strict, rather than a weak, preference relation between f and g. Q.E.D.

Proof of Theorem 2: We will prove the theorem by demonstrating that (1) $\Leftrightarrow$ (2) and then that (2) $\Leftrightarrow$ (3)*. We first show that conditions (1) and (2) are equivalent. Assuming condition (1), apply Theorem 1 to obtain a full set of conditional preference relations, $\succsim_i$, $i \in \{1, \ldots, k\}$, for which Axiom 4 holds. It remains to prove that Axiom 2, in the presence of Axiom 1, implies that Axiom 5 holds. However, this is immediate, since $\forall f, g, h \in L_0$ and $\forall i \in \{1, \ldots, k\}$, we have $(f; h)_i \succsim (g; h)_i \Leftrightarrow f \succsim_i g$ (by the definition of the conditional preference relations constructed in the proof of Theorem 1). This proves that (1) $\Rightarrow$ (2).

Now suppose that condition (2) holds. Apply Theorem 1 to prove that Axiom 1 holds. Then Axiom 2, combined with the consequentialism property of the conditional preference relations, implies Axiom 5 directly. This shows that (2) $\Rightarrow$ (1), and combining this with the above yields (1) $\Leftrightarrow$ (2).

We now prove that (2) $\Rightarrow$ (3). Note that since each $A_i \in \mathcal{A}$ is non-null, each conditional preference relation $\succsim_i$, $i \in \{1, \ldots, k\}$ is non-degenerate. This, in addition to the fact that Axiom 4 implies that each conditional preference relation satisfies the other portions of Axiom 3, allows us to apply Theorem 1 of Gilboa and Schmeidler (1989) to each conditional preference relation $\succsim_i$, $i \in \{1, \ldots, k\}$. We can conclude that, $\forall i \in \{1, \ldots, k\}$, $\succsim_i$ is represented by

$$\min_{P_i \in \mathcal{C}_i} \left\{ \int_{s \in A_i} u_i(f(s))\, dP_i(s) \right\},$$

where the closed convex set $\mathcal{C}_i$ of probability distributions is unique and u_i is non-constant, mixture linear, and unique up to a positive affine transformation.

We must verify that we may take $u_i = u$ w.l.o.g. This is implied directly by Axiom 5: since all of the conditional preference relations agree on the roulette lotteries L_c, and since any preference relation on L_c implies a u that is unique up to a positive affine transformation, the u_i differ by at most a positive affine

*I am grateful to Larry Epstein and Martin Schneider for pointing out that a construction using their main result in Epstein and Schneider (2003) could be used to prove that (2) $\Leftrightarrow$ (3). I provide a direct proof because it seems more revealing.

transformation. Since any positive affine transformation of u_i represents preferences over L_c, we can let $u_i = u_1 \;\; \forall i \in \{2, \ldots, k\}$ w.l.o.g.

For any act $f \in L_0$, the minimum in the representation is achieved (as is evident from the construction of the set of distributions in Lemma 3.5 of Gilboa and Schmeidler (1989)). Let $P_i^*(f) \in \mathcal{C}_i$ be a probability distribution achieving the minimum (if there is more than one such probability distribution, choose one arbitrarily). Define $l_i(f) \in L_c$ to be the roulette lottery such that $(l_i(f))(s) = \int_{s\in A_i} f(s)\, d(P_i^*(f))(s)$. Then

$$\begin{aligned}
& \min_{P_i\in\mathcal{C}_i}\left\{\int_{s\in A_i} u(l_i(f))\, dP_i(s)\right\} \\
= \;& u(l_i(f)) \\
= \;& u\left(\int_{s\in A_i} f(s)\, d(P_i^*(f))(s)\right) \\
= \;& \int_{s\in A_i} u(f(s))\, d(P_i^*(f))(s) \\
= \;& \min_{P_i\in\mathcal{C}_i}\left\{\int_{s\in A_i} u(f(s))\, dP_i(s)\right\},
\end{aligned}$$

where the first equality follows from the fact that $l_i(f)$ is a roulette lottery (so the choice of probability distribution from $\mathcal{C}_i$ does not affect the expectation integral), the second equality is by the definition of $l_i(f)$, the third equality is by the mixture linearity of u, and the final equality is by the definition of $P_i^*(f)$.

By the representation result above, this implies that $f \sim_i l_i(f)$.

Now apply Theorem 1 of Gilboa and Schmeidler (1989) to the original preference relation, $\succsim$. This allows us to conclude that $\succsim$ is represented by

$$\min_{P\in\mathcal{P}}\left\{\int_{s\in S} w(f(s))\, dP(s)\right\},$$

where the closed convex set $\mathcal{P}$ of probability distributions is unique and w is non-constant, mixture linear, and unique up to a positive affine transformation.

We must verify that we may take $w = u$ w.l.o.g. We will do so by showing that, for any roulette lotteries $l, q \in L_c$, $u(l) \geq u(q) \Leftrightarrow w(l) \geq w(q)$. This equivalent to showing that, for any two roulette lotteries $l, q \in L_c$, $u(l) \geq u(q) \Rightarrow w(l) \geq w(q)$ and $u(l) < u(q) \Rightarrow w(l) < w(q)$.

Given any two roulette lotteries $l, q \in L_c$, if $u(l) \geq u(q)$ then (by the representation result, and the fact that we have shown that we may take $u_i = u$ for all $i \in \{1, \ldots, k\}$) $l \succsim_i q$ for all $i \in \{1, \ldots, k\}$. The consistency portion of Axiom 4 then implies that $l \succsim q$, so $w(l) \geq w(q)$ since w represents the preference relation $\succsim$ on the set of roulette lotteries L_c. Now suppose that, instead, $u(l) < u(q)$; then (by the representation result, and the fact that we have shown that we may take $u_i = u$ for all $i \in \{1, \ldots, k\}$) $l \prec_i q$ for all $i \in \{1, \ldots, k\}$. The "if, in addition," part of the consistency portion of Axiom 4, along with the non-nullity of each A_i, then implies that $l \prec q$, so $w(l) < w(q)$ since w represents the preference relation $\succsim$ on the set of roulette lotteries L_c. We have thus shown that u and w represent the same preferences over L_c. Since any u, w representing the same preferences over

L_c differ by at most a positive affine transformation, and since any positive affine transformation of u represents the same preferences over L_c that u does, we can set $w = u$ w.l.o.g.

Given any act $f \in L_0$, recall the definition of $l_i(f)$ given above. Define the partitionwise-constant act g by: $\forall i \in \{1, \ldots, k\}, \quad \forall s \in A_i, \quad g(s) = l_i(f)$. Then g is well-defined on all of S, since $\mathcal{A}$ is a partition of S. We then have $f \sim_i g \ \ \forall i \in \{1, \ldots, k\}$, since $f \sim_i l_i(f) \ \ \forall i \in \{1, \ldots, k\}$ as shown above and $l_i(f) \sim_i g \ \ \forall i \in \{1, \ldots, k\}$ by the consequentialism property of conditional preferences. By the consistency property of preferences, this implies that $f \sim g$. Define the set of priors

$$\mathcal{P}_0 = \left\{ \begin{array}{c} P \ : \ \forall B \in \Sigma, P(B) = \sum_{i=1}^{k} P_i(B|\, A_i)\, Q(A_i) \\ \text{for some } P_i \in \mathcal{C}_i, i \in \{1, \ldots, k\} \text{ and } Q \in \mathcal{P} \end{array} \right\}.$$

$\mathcal{P}_0$ is closed and convex because its components are.

$$\begin{aligned} & \min_{P \in \mathcal{P}} \left\{ \int_{s \in S} u(f(s))\, dP(s) \right\} \\ = & \min_{P \in \mathcal{P}} \left\{ \int_{s \in S} u(g(s))\, dP(s) \right\} \\ = & \min_{P \in \mathcal{P}} \left\{ \sum_{i=1}^{k} u(l_i(f))\, P(A_i) \right\} \\ = & \min_{P \in \mathcal{P}} \left\{ \sum_{i=1}^{k} \min_{P_i \in \mathcal{C}_i} \left\{ \int_{s \in A_i} u(f(s))\, dP_i(s) \right\} P(A_i) \right\} \\ = & \min_{P \in \mathcal{P}_0} \left\{ \int_{s \in S} u(f(s))\, dP(s) \right\}, \end{aligned}$$

where the first equality follows from the representation result for $\succsim$ and the fact that $f \sim g$, the second equality follows from the fact that g is constant (at $l_i(f)$) on each $A_i \in \mathcal{A}$, the third equality follows from the results derived for $l_i(f)$ on A_i above, and the final equality follows from the definition of $\mathcal{P}_0$.

Since the above equality holds $\forall \ f \in L_0$, we conclude that we can replace $\mathcal{P}$ with $\mathcal{P}_0$ in the utility representation of $\succsim$. Part of the representation result, however, is that $\mathcal{P}$ is the only closed, convex set of probability distributions for which the utility representation holds. Thus, we must have $\mathcal{P} = \mathcal{P}_0$, which is prismatic if we can show that $\forall i \in \{1, \ldots, k\}$ and $\forall P \in \mathcal{P}, \ P(A_i) > 0$.

Suppose not; then $\exists j \in \{1, \ldots, k\}$ and $P \in \mathcal{P}$ such that $P(A_j) = 0$. By the non-degeneracy condition, there exist two roulette lotteries $l, q \in L_c$ such that $l \succ_j q$. Since we can select any positive affine transformation of u in the representation result, and since $u(l) > u(q)$, we can w.l.o.g. choose u such that $u(l) > 0$ and $u(q) = 0$. We do so. Obviously, $q \sim_i q$ for all $i \in \{1, \ldots, k\}$. Consider the act $f = (l; q)_j$. Using our selection of u to evaluate its utility, we have:

$$\begin{aligned} & \min_{P \in \mathcal{P}} \left\{ \int_{s \in S} u(f(s))\, dP(s) \right\} \\ = & \min_{P \in \mathcal{P}} \left\{ \sum_{i=1, i \neq j}^{k} u(q)\, P(A_i) + u(l)\, P(A_j) \right\} \end{aligned}$$

$$
\begin{aligned}
&= \min_{P\in\mathcal{P}} \{u(l)\, P(A_j)\} \\
&= 0,
\end{aligned}
$$

where the first equality follows from the construction of $f = (l;q)_j$, the second equality follows from the fact that we have (as explained above) set $u(q) = 0$ w.l.o.g., and the third equality follows from the facts that, by assumption, $\exists P \in \mathcal{P}$ such that $P(A_j) = 0$ and that (again w.l.o.g., as explained above) we have $u(l) > 0$. However, we also have that

$$
\begin{aligned}
&\min_{P\in\mathcal{P}} \left\{ \int_{s\in S} u(q(s))\, dP(s) \right\} \\
&= u(q) \\
&= 0,
\end{aligned}
$$

where the first equality holds because $q \in L_c$ is a roulette lottery and the second holds because $u(q) = 0$ by our selection (made w.l.o.g) of u. But, by the representation result, we have $f \sim q$. This contradicts the "if, in addition" portion of the consistency part of Axiom 4, which (along with the non-nullity of A_j) implies that $f \succ q$. Our assumption that $\exists j \in \{1,\ldots,k\}$ and $P \in \mathcal{P}$ such that $P(A_j) = 0$ must, then, have been false. As a consequence, $\forall i \in \{1,\ldots,k\}$ and $\forall P \in \mathcal{P}$, $P(A_i) > 0$ must hold, and $\mathcal{P}$ is prismatic by definition.

We have now shown that (1) $\Leftrightarrow$ (2) and that (2) $\Rightarrow$ (3). It remains to show that (3) $\Rightarrow$ (2). Theorem 1 of Gilboa and Schmeidler (1989) shows that (3) implies Axiom 3. Thus, we only need to verify that (3) implies Axioms 4 and 2 relative to the partition $\mathcal{A}$. Using the representation for conditional preferences given by (3) and again applying Theorem 1 of Gilboa and Schmeidler (1989), we can conclude that conditional preferences satisfy the slightly modified version of Axiom 3 that Axiom 4 states they must. Also, since each $\mathcal{C}_i$ contains only P_i such that $P_i(A_i) = 1$, the consequentialism property of conditional preferences is clear. If $f \succsim_i g \;\; \forall i \in \{1,\ldots,k\}$, then we have that

$$
\begin{aligned}
\forall i \in \{1,\ldots,k\}, \qquad &\min_{P_i\in\mathcal{C}_i} \left\{ \int_{s\in A_i} u(f(s))\, dP_i(s) \right\} \\
&\geq \min_{P_i\in\mathcal{C}_i} \left\{ \int_{s\in A_i} u(g(s))\, dP_i(s) \right\}.
\end{aligned}
$$

By the prismatic structure of $\mathcal{P}$,

$$
\begin{aligned}
&\min_{P\in\mathcal{P}} \left\{ \int_{s\in S} u(f(s))\, dP(s) \right\} \\
&= \min_{P\in\mathcal{P}} \left\{ \sum_{i=1}^{k} \min_{P_i\in\mathcal{C}_i} \left\{ \int_{s\in A_i} u(f(s))\, dP_i(s) \right\} P(A_i) \right\} \\
&\geq \min_{P\in\mathcal{P}} \left\{ \sum_{i=1}^{k} \min_{P_i\in\mathcal{C}_i} \left\{ \int_{s\in A_i} u(g(s))\, dP_i(s) \right\} P(A_i) \right\} \\
&= \min_{P\in\mathcal{P}} \left\{ \int_{s\in S} u(g(s))\, dP(s) \right\},
\end{aligned}
$$

so that $f \succsim g$, confirming that the first portion of the consistency property of conditional preferences holds. We must still show that the second, "if, in addition," portion of the consistency property of conditional preferences holds. Since all $A_i \in \mathcal{A}$ are non-null, suppose that $f \succsim_i g \quad \forall i \in \{1, \ldots, k\}$ and that $\exists j \in \{1, \ldots, k\}$ such that $f \succ_j g$. Then

$$\forall i \in \{1, \ldots, k\}, \qquad \min_{P_i \in \mathcal{C}_i} \left\{ \int_{s \in A_i} u(f(s)) \, dP_i(s) \right\} \geq \min_{P_i \in \mathcal{C}_i} \left\{ \int_{s \in A_i} u(g(s)) \, dP_i(s) \right\}.$$

Also,

$$\exists j \in \{1, \ldots, k\}, \qquad \text{such that}$$
$$\min_{P_j \in \mathcal{C}_i} \left\{ \int_{s \in A_j} u(f(s)) \, dP_j(s) \right\} > \min_{P_j \in \mathcal{C}_j} \left\{ \int_{s \in A_i} u(g(s)) \, dP_j(s) \right\}.$$

Now, $P(A_j) > 0 \ \ \forall P \in \mathcal{P}$ since $\mathcal{P}$ is prismatic. Thus, (recalling that $\mathcal{P}$ is closed, so that the strict inequality is preserved even on its boundary)

$$\begin{aligned}
& \min_{P \in \mathcal{P}} \left\{ \int_{s \in S} u(f(s)) \, dP(s) \right\} \\
= \ & \min_{P \in \mathcal{P}} \left\{ \sum_{i=1}^{k} \min_{P_i \in \mathcal{C}_i} \left\{ \int_{s \in A_i} u(f(s)) \, dP_i(s) \right\} P(A_i) \right\} \\
> \ & \min_{P \in \mathcal{P}} \left\{ \sum_{i=1}^{k} \min_{P_i \in \mathcal{C}_i} \left\{ \int_{s \in A_i} u(g(s)) \, dP_i(s) \right\} P(A_i) \right\} \\
= \ & \min_{P \in \mathcal{P}} \left\{ \int_{s \in S} u(g(s)) \, dP(s) \right\},
\end{aligned}$$

so $f \succ g$, showing that the "if, in addition" portion of the consistency property holds.

Finally, we must show that Axiom 2 holds. This is a direct consequence of the fact that the same function u appears in the representation of each $\succsim_i$, $i \in \{1, \ldots, k\}$. Since the probability measure is irrelevant to computing utility for a roulette lottery (because a roulette lottery, by definition, does not depend on the state s), we have that $\forall l, q \in L_c$ and $\forall i, j \in \{1, \ldots, k\}$, $\ l \succsim_i q \Leftrightarrow u(l) \geq u(q) \Leftrightarrow l \succsim_j q$. We have thus verified that Axiom 2 holds, and therefore that all of condition (2) holds. Having shown that (2) $\Leftrightarrow$ (3), we have completed the proof. Q.E.D.

Proof of Proposition 1: Invoke Lemma 3.1 of Gilboa and Schmeidler (1989), or Chapter 8 of Fishburn (1979) (which is cited by Gilboa and Schmeidler (1989)) to prove the representation result and the uniqueness of v up to a positive affine transformation. In order to prove that v is additively time-separable, note that (by construction), only the set of time-t marginal distributions of consumption, for each $t \in \{0, \ldots, T\}$, matter in ranking lotteries. This is due to our restriction of the domain of preferences; all of the roulette lotteries we consider have each individual roulette lottery over consumption at time t being independent. By Fishburn (1979), Theorem 11.1 (on page 149), the function v is additively time-separable (note that the condition of the theorem is satisified *a fortiori*). Q.E.D.

Proof of Theorem 3: First observe that Axiom 3 is assumed in both of Theorems 1 and 2. Thus we may apply Proposition 1 to conclude that there is a von Neumann-Morgenstern utility function v that represents $\succsim$ in comparing roulette lotteries. v evidently maps adapted acts to functions from S to Y^U, since if $f \in \mathcal{H}$, then $v(f(s))$ is in $Y^U = X^U$ for every $s \in S$. If we can show that v is an isomorphism *on indifference classes* of $f \in \mathcal{H}$, then we can define a new preference relation $\succsim^V$ by $\forall f, g \in \mathcal{H},\ f \succsim g \Leftrightarrow v(f) \succsim^V v(g)$ and, since v is an isomorphism on indifference classes, the new preference relation $\succsim^V$ will be well-defined. By the definition of $Y^U = X^U$ (and the mixture linearity of v), we can see that v is onto. To show that it is one-to-one *as a mapping of indifference classes*, note that $v(f) = v(g) \quad \forall s \in S$ implies that $f(s) \sim g(s) \quad \forall s \in S$, since v represents $\succsim$ on roulette lotteries (and the constant act with value $f(s)$ is a roulette lottery). By monotonicity, then, $f \sim g$. This implies that v is one-to-one as a mapping of indifference classes.

The identical reasoning may be applied to any conditional preference relation. Further, any axiom satisfied by $\succsim$ is also satisfied by $\succsim^V$, due to the mixture linearity of v. (We might call v a "mixture isomorphism," since $v(\alpha f + (1-\alpha) g) = \alpha v(f) + (1-\alpha) v(g)$.)

It only remains to show that the function u of Theorem 2 is, in fact, v (at least, up to a positive affine transformation). But if this were not so, then u could not represent $\succsim$ on roulette lotteries, which would contradict the representation result of Theorem 2. Thus, u is at most a positive affine transformation of v. Since v is additively time-separable, u must be as well. Q.E.D.

Proof of Theorem 4: First, suppose that we are given a closed, convex set $\mathcal{P}$ of distributions on $[0,1]$. Let $H : \mathcal{P} \to \mathcal{E}$ (where $\mathcal{E}$ is a set of distributions on $\{0,1\}^\infty$) be defined such that $H(Q)$, for any $Q \in \mathcal{P}$ and for any $N < \infty$ and for any distinct $i_1, i_2, \ldots, i_N \in \mathbb{N}$, satisfies

$$\Pr_{H(Q)}(x_{i_1}, x_{i_2}, \ldots, x_{i_N}) \equiv \int_0^1 p^{\sum_{j=1}^N x_{i_j}} (1-p)^{N - \sum_{j=1}^N x_{i_j}} \, dQ(p). \qquad \text{(A.1)}$$

We will show that H is continuous (under the topology of weak convergence), invertible, and linear for convex combinations, and that H^{-1} is linear for convex combinations. By demonstrating that H^{-1} is also continuous (under the topology of weak convergence), we will complete the proof of both portions of the theorem: the inverse image of a closed set under a continuous function is itself closed, and the image of a convex set under a function that is linear under convex combinations is itself convex.

First, we prove that H is continuous. We do so by showing that, if $Q_i \in \mathcal{P} \ \ \forall i$ and $Q_i \Rightarrow Q^*$ (weak convergence), then $H(Q_i) \Rightarrow H(Q^*)$ (again, weak convergence). Note that the space of all sequences of zeros and ones is countably infinite. Let $f(x|\, p)$ be the infinite binomial distribution with fixed probability p of $x_i = 1$; that is, for any $N < \infty$ and for any distinct $i_1, i_2, \ldots, i_N \in \mathbb{N}$,

$$f\left(x_{i_1}, x_{i_2}, \ldots, x_{i_N} \middle| p\right) = p^{\sum_{j=1}^{N} x_{i_j}} (1-p)^{N - \sum_{j=1}^{N} x_{i_j}}. \tag{A.2}$$

For any bounded, continuous function g on $\{0,1\}^{\infty}$ (where continuity is with respect to the metric given in the text), we have

$$\begin{aligned} E_{H(Q)}\left[g(X)\right] &= \sum_y g(y) H(Q)(y) && \text{(A.3)} \\ &= \sum_y g(y) \int_0^1 f(y|p)\, dQ(p) && \text{(A.4)} \\ &= \int_0^1 \sum_y g(y) f(y|p)\, dQ(p) && \text{(A.5)} \\ &= \int_0^1 w(p;g)\, dQ(p), && \text{(A.6)} \end{aligned}$$

where the first equality is by the definition of expectation (recalling that $\{0,1\}^{\infty}$ is countably infinite, so the expectation involves a sum), the second equality is by the definition of $H(Q)$ (and the definition of f), the third equality is by the linearity of integration (allowing us to place g inside the integral over p) and then Fubini's theorem, and the fourth equality follows by defining

$$\begin{aligned} w(p;g) &\equiv \sum_y g(y) f(y|p) && \text{(A.7)} \\ &= E_f\left[g(Y)|p\right]. && \text{(A.8)} \end{aligned}$$

We wish to show that $g(\cdot)$ bounded and uniformly continuous over $\{0,1\}^{\infty}$ (under the given metric) implies that $w(\cdot;g)$ is bounded and continuous over $[0,1]$. It is obvious that w is bounded if g is:

$$\begin{aligned} \sup_{p\in[0,1]} |w(p;g)| &= \sup_{p\in[0,1]} \left| \sum_y g(y) f(y|p) \right| && \text{(A.9)} \\ &\leq \sup_{p\in[0,1]} \left| \sup_{y\in\{0,1\}^{\infty}} |g(y)| \right| && \text{(A.10)} \\ &= \sup_{y\in\{0,1\}^{\infty}} |g(y)|, && \text{(A.11)} \end{aligned}$$

where the inequality is clear from the interpretation of w as an expectation of g under f (conditional on some p).

We now prove that g bounded and uniformly continuous implies w continuous. Given any $\epsilon > 0$, the uniform continuity of g guarantees that there exists some $\delta > 0$ such that

$$\forall x, y \in \{0,1\}^{\infty}, \sum_{i=1}^{\infty} \frac{1}{2^i} |x_i - y_i| < \delta \implies |g(x) - g(y)| < \frac{\epsilon}{3}. \tag{A.12}$$

Choose M sufficiently large that $\sum_{i=1}^{\infty} \frac{1}{2^i} < \delta$. Then

$$\begin{aligned}
& |w(p;g) - w(q;g)| \\
= & \left| \sum_y g(y) f(y|p) - \sum_y g(y) f(y|q) \right| && \text{(A.13)} \\
= & \left| \begin{array}{c} \sum_y [g(y) - g(y_1, y_2, \ldots, y_M, 0, 0, \ldots)] f(y|p) \\ + \sum_y g(y_1, y_2, \ldots, y_M, 0, 0, \ldots) [f(y|p) - f(y|q)] \\ + \sum_y [g(y_1, y_2, \ldots, y_M, 0, 0, \ldots) - g(y)] f(y|q) \end{array} \right| && \text{(A.14)} \\
\leq & \begin{array}{c} \left| \sum_y [g(y) - g(y_1, y_2, \ldots, y_M, 0, 0, \ldots)] f(y|p) \right| \\ + \left| \sum_y g(y_1, y_2, \ldots, y_M, 0, 0, \ldots) [f(y|p) - f(y|q)] \right| \\ + \left| \sum_y [g(y_1, y_2, \ldots, y_M, 0, 0, \ldots) - g(y)] f(y|q) \right| \end{array} && \text{(A.15)} \\
\equiv & T_1 + T_2 + T_3, && \text{(A.16)}
\end{aligned}$$

where the first equality is by definition, the second equality is by adding and subtracting both $g(y_1, y_2, \ldots, y_M, 0, 0, \ldots) f(y|p)$ and $g(y_1, y_2, \ldots, y_M, 0, 0, \ldots) f(y|q)$, the inequality is by the triangle inequality, and the last line defines the terms T_1, T_2, and T_3 as the first, second, and third terms in the previous expression.

We first analyze T_1:

$$\begin{aligned}
T_1 & \leq \sum_y |g(y) - g(y_1, y_2, \ldots, y_M, 0, 0, \ldots)| f(y|p) && \text{(A.17)} \\
& < \sum_y \frac{\epsilon}{3} f(y|p) && \text{(A.18)} \\
& = \frac{\epsilon}{3} \sum_y f(y|p) && \text{(A.19)} \\
& = \frac{\epsilon}{3}, && \text{(A.20)}
\end{aligned}$$

where the first inequality is by the triangle inequality and by the nonnegativity of f, the second inequality is by the uniform continuity of g and the fact that, by the construction of M, $d(y, (y_1, y_2, \ldots, y_M, 0, 0, \ldots)) < \delta$, the first equality is by the linearity of summation, and the final equality is by the fact that f is a probability distribution over $\{0,1\}^{\infty}$.

Exactly the same logic shows that $T_3 < \frac{\epsilon}{3}$. All that remains is to bound T_2. To do so, notice that the function $g(y_1, y_2, \ldots, y_M, 0, 0, \ldots)$ does not depend on any coordinate of the sequence y beyond the M^{th}. Thus, its expectation can be evaluated using the restriction of f to the first M coordinates of y:

$$\begin{aligned}
& \sum_y g(y_1, y_2, \ldots, y_M, 0, 0, \ldots) f(y|p) \\
= & \sum_{y_1=0}^{1} \sum_{y_2=0}^{1} \cdots \sum_{y_M=0}^{1} g(y_1, y_2, \ldots, y_M, 0, 0, \ldots) f(y_1, y_2, \ldots, y_M|p). && \text{(A.21)}
\end{aligned}$$

But $f(y_1, y_2, \ldots, y_M|p) = p^{\sum_{i=1}^{M} x_i} (1-p)^{M - \sum_{i=1}^{M} y_i}$ is continuous in p for any fixed vector $(y_1, y_2, \ldots, y_M)$ of zeros and ones, and g is bounded, so the expression to the right of the equality in the display above is a finite linear combination of continuous functions of p, and is thus a continuous function of p. But this means that, given $\frac{\epsilon}{3} > 0$, there exists some $\delta_2 > 0$ such that $T_2 < \frac{\epsilon}{3}$.

We have shown that for any bounded and uniformly continuous g, and for any $p \in [0,1]$, given $\epsilon > 0$, there exists $\delta_2 > 0$ such that $|p - q| < \delta_2$ implies that $|w(p; g) - w(q; g)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$. But this shows that g bounded and uniformly continuous implies that $w(\cdot; g)$ is continuous.

We now return to the problem of demonstrating that H is continuous. By Billingsley (1999), Theorem 2.1, part (i), a sequence of distributions $H(Q_i)$ converges weakly to $H(Q^*)$ if and only if $E_{H(Q_i)}[g(Y)] \to E_{H(Q^*)}[g(Y)]$ for every bounded, uniformly continuous function g. Suppose that $Q_i \Rightarrow Q^*$. Given any bounded, uniformly continuous g, we have by the definitions above that

$$\forall i, \; E_{H(Q_i)}[g(Y)] = E_{Q_i}[w(P; g)] \tag{A.22}$$

$$E_{H(Q^*)}[g(Y)] = E_{Q^*}[w(P; g)], \tag{A.23}$$

where $w(\cdot; g)$ is bounded and continuous. Since $Q_i \Rightarrow Q^*$, we have that

$$E_{Q_i}[w(P; g)] \quad \to \quad E_{Q^*}[w(P; g)]$$

by the definition of weak convergence (see Billingsley (1999), page 7). But, by the equalities above, this implies that $E_{H(Q_i)}[g(Y)] \to E_{H(Q^*)}[g(Y)]$. Since g was an arbitrary bounded, uniformly continuous function, we have that $H(Q_i) \Rightarrow H(Q^*)$. This shows that $Q_i \Rightarrow Q^*$ implies that $H(Q_i) \Rightarrow H(Q^*)$, which proves that H is continuous.

Next, we show that H is invertible. The classical de Finetti theorem, found on pages 228 and 229 of Feller (1971), shows that H is surjective as a map of all distributions on $[0,1]$ to all exchangeable distributions on the set of all infinite sequences of zeros and ones: given any exchangeable distribution on the set of all infinite sequences of zeros and ones, we can find a distribution on $[0,1]$ that generates it in the sense given above. It remains to show that H is injective. By the proof of de Finetti's theorem found on pages 228 and 229 of Feller (1971), any exchangeable distribution on the space of infinite sequences of zeros and ones uniquely determines a completely monotone sequence $\{c_i\}_{i=0}^{\infty}$ such that $c_0 = 1$, and this completely monotone sequence is the sequence of moments of the distribution on $[0,1]$ that generates the exchangeable distribution in the sense used above. But any such completely monotone moment sequence uniquely determines a probability distribution on $[0,1]$ by Feller (1971), Theorem 1 on pages 225 to 227. Thus, an exchangeable distribution on the set of all infinite sequences of zeros and ones uniquely determines the distribution on $[0,1]$ that generates it in the sense used above. That is, $H(Q)$ uniquely determines Q. This implies that H is injective; since we also showed that it is surjective, H is thus invertible. In fact, we have shown not only that H is invertible as a function from $\mathcal{P}$ to $\mathcal{E}$, but also that it is invertible as a function from the set of all distributions on $[0,1]$ to the set of all exchangeable distributions on the set of all infinite sequences of zeros and ones.

To see that H is linear under convex combinations, simply examine the definition of H in (A.1). The linearity of integration in the integrating measure implies the result. The same approach shows that H^{-1} is also linear under convex combinations.

It remains to demonstrate that H^{-1} is continuous under the topology of weak convergence. Since the set of all measures on $[0,1]$ is tight (for any $\epsilon > 0$, $[0,1]$ itself is a compact set such that any distribution on $[0,1]$ has $\Pr([0,1]) > 1-\epsilon$; see the definition of tightness on page 59 of Billingsley (1999)), the set of all measures on $[0,1]$ is relatively (or sequentially) compact under the topology of weak convergence by Prohorov's theorem (see Billingsley (1999), Theorem 5.1).

Suppose that H^{-1} is not continuous. Then $\exists \{Q_n\}_{n=1}^{\infty} \subset \mathcal{P}$ and $Q^* \in \mathcal{P}$ such that $H(Q_n) \Rightarrow H(Q^*)$ (recalling that these probability measures are in the domain of H^{-1}) but $Q_n \not\Rightarrow Q^*$. If this is so, then (by the definition of weak convergence) there exists some continuous, bounded function m on $[0,1]$ such that $\int_0^1 m dQ_n \nrightarrow \int_0^1 m dQ^*$. Since m is bounded, there is some a such that $\int_0^1 m dQ_n \in [-a,a]$ $\forall n$. By the definition of nonconvergence, $\exists \epsilon > 0$ such that $\forall N$, $\exists n \geq N$ such that $\left|\int_0^1 m dQ_n - \int_0^1 m dQ^*\right| \geq \epsilon$. Consider the compact set $A \equiv [-a,-\epsilon] \cup [\epsilon,a]$. The logic above shows that we can take an infinite subsequence $\left\{Q_{n(k)}\right\}$ of $\{Q_n\}$ such that $\int_0^1 m dQ_{n(k)} \in A$ $\forall k$. But, since A is compact, we can extract a further subsubsequence $\left\{Q_{n(k,j)}\right\}$ of $\left\{Q_{n(k)}\right\}$ such that $\int_0^1 m dQ_{n(k,j)} \to c_0 \in A$; obviously, $c_0 \in A$ implies that $c_0 \neq \int_0^1 m dQ^*$. Now, recall that we showed in the previous paragraph that the set of all measures on $[0,1]$ is sequentially compact under the topology of weak convergence. Thus, we may extract a further subsubsubsequence $\left\{Q_{n(k,j,i)}\right\}$ such that $Q_{n(k,j,i)} \Rightarrow Q^{**}$. But $Q^{**} \neq Q^*$, since $\int_0^1 m dQ^{**} = c_0$ (by the construction of $\left\{Q_{n(k,j)}\right\}$ and the fact that $\left\{Q_{n(k,j,i)}\right\}$ is a subsequence of $\left\{Q_{n(k,j)}\right\}$) and $c_0 \neq \int_0^1 m dQ^*$. By the continuity of H, $H\left(Q_{n(k,j,i)}\right) \Rightarrow H(Q^{**})$. But $H(Q_n) \Rightarrow H(Q^*)$, so $H\left(Q_{n(k,j,i)}\right) \Rightarrow H(Q^*)$ because a subsequence of a convergent sequence must converge to the same limit as the parent sequence. Since H is injective and $Q^* \neq Q^{**}$ as shown above, $H(Q^*) \neq H(Q^{**})$. Thus, there exists a bounded, continuous function r on $\{0,1\}^{\infty}$ such that $\int r dH(Q^*) \neq \int r dH(Q^{**})$. But then $\int r dH\left(Q_{n(k,j,i)}\right) \to \int r dH(Q^*)$ and $\int r dH\left(Q_{n(k,j,i)}\right) \to \int r dH(Q^{**})$, where $\int r dH(Q^*) \neq \int r dH(Q^{**})$, which is a contradiction since the limit of a sequence of real numbers, if it exists, is unique. Thus, H^{-1} must be continuous. Q. E. D.

References

Anderson, E., L. P. Hansen, and T. J. Sargent (2000). Robustness, detection and the price of risk. Unpublished manuscript, University of Chicago.

Anscombe, F. and R. J. Aumann (1963). A definition of subjective probability. *Annals of Mathematical Statistics 34*, 199–205.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance 55*, 225–264.

Billingsley, P. (1999). *Convergence of Probability Measures* (Second ed.). New York: John Wiley and Sons.

Brennan, M. J. (1998). The role of learning in dynamic portfolio decisions. *European Finance Review 1*, 295–306.

Brennan, M. J. and Y. Xia (2001). Assessing asset pricing anomalies. *Review of Financial Studies 14*, 905–942.

Cagetti, M., L. P. Hansen, T. J. Sargent, and N. Williams (2002). Robustness and pricing with uncertain growth. *The Review of Financial Studies 15*, 363–404.

Casadesus-Masanell, R., P. Klibanoff, and E. Ozdenoren (2000). Maxmin expected utility over Savage acts with a set of priors. *Journal of Economic Theory 92*, 35–65.

Chamberlain, G. (2000). Econometric applications of maxmin expected utility. *Journal of Applied Econometrics 15*, 625–644.

Chamberlain, G. (2001). Minimax estimation and forecasting in a stationary autoregression model. *American Economic Review Papers and Proceedings 91*, 55–59.

Chen, Z. and L. Epstein (2002). Ambiguity, risk and asset returns in continuous time. *Econometrica 70*, 1403–1443.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics 75*, 643–669.

Epstein, L. and M. Schneider (2002). Learning under ambiguity. Unpublished manuscript, University of Rochester.

Epstein, L. and M. Schneider (2003). Recursive multiple-priors. *Journal of Economic Theory*, forthcoming.

Epstein, L. and T. Wang (1994). Intertemporal asset pricing under Knightian uncertainty. *Econometrica 62*, 283–322.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (Second ed.), Volume 2. New York: John Wiley and Sons.

Fishburn, P. C. (1979). *Utility Theory for Decision Making* (Second ed.). Huntington, New York: Robert E. Krieger Publishing Co., Inc.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*, 141–153.

Gollier, C. (2002). Optimal dynamic portfolio risk with first-order and second-order predictability. Unpublished manuscript, University of Toulouse.

Hammond, P. J. (1988). Consequentialist foundations for expected utility. *Theory and Decision 25*, 25–78.

Hansen, L. P. and T. J. Sargent (1995). Discounted linear exponential quadratic Gaussian control. *IEEE Transactions on Automatic Control 40*, 968–971.

Hansen, L. P. and T. J. Sargent (2001). Time inconsistency of robust control? Unpublished manuscript, University of Chicago.

Hansen, L. P., T. J. Sargent, and T. D. Tallarini (1999). Robust permanent income and pricing. *Review of Economic Studies 66*, 873–907.

Hansen, L. P., T. J. Sargent, G. A. Turmuhambetova, and N. Williams (2001). Robustness and uncertainty aversion. Unpublished manuscript, University of Chicago.

Hansen, L. P., T. J. Sargent, and N. E. Wang (2002). Robust permanent income and pricing with filtering. *Macroeconomic Dynamics 6*, 40–84.

Ingersoll, Jr., J. E. (1987). *Theory of Financial Decision Making*. Savage, Maryland: Rowman and Littefield Publishers, Inc.

Kandel, S. and R. F. Stambaugh (1996). On the predictability of stock returns: an asset allocation perspective. *Journal of Finance 51*, 385–424.

Klibanoff, P. (1995). Dynamic choice with uncertainty aversion. Unpublished manuscript, Kellogg School of Management, Northwestern University.

Klibanoff, P., M. Marinacci, and S. Mukerji (2003). A smooth model of decision making under ambiguity. Unpublished manuscript, Kellogg School of Management, Northwestern University.

Knox, T. A. (2003). Analytical methods for learning how to invest when returns are uncertain. Unpublished manuscript, University of Chicago Graduate School of Business.

Kreps, D. M. (1988). *Notes on the Theory of Choice*. Boulder, Colorado: Westview Press.

Liu, J., J. Pan, and T. Wang (2003). An equilibrium model of rare event premia. Unpublished manuscript, MIT.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature 27*, 1622–1668.

Maenhout, P. J. (2001). Robust portfolio rules, hedging and asset pricing. Unpublished manuscript, INSEAD.

Miao, J. (2001). Ambiguity, risk and portfolio choice under incomplete information. Unpublished manuscript, University of Rochester.

Pástor, L. (2000). Portfolio selection and asset pricing models. *Journal of Finance 55*, 179–223.

Pástor, L. and R. F. Stambaugh (2000). Comparing asset pricing models: An investment perspective. *Journal of Financial Economics 56*, 335–381.

Routledge, B. R. and S. E. Zin (2001). Model uncertainty and liquidity. NBER working paper number w8683.

Savage, L. J. (1954). *The Foundations of Statistics*. New York, New York: John Wiley and Sons.

Siniscalchi, M. (2001). Bayesian updating for general maximin expected utility preferences. Unpublished manuscript, Princeton University.

Skiadas, C. (1997a). Conditioning and aggregation of preferences. *Econometrica 65*, 347–367.

Skiadas, C. (1997b). Subjective probability under additive aggregation of conditional preferences. *Journal of Economic Theory 76*, 242–271.

Wang, T. (2003). Conditional preferences and updating. *Journal of Economic Theory 108*, 286–321.

Xia, Y. (2001). Learning about predictability: the effects of parameter uncertainty on dynamic asset allocation. *Journal of Finance 56*, 205–246.