THE CARLO ALBERTO NOTEBOOKS

Fondazione Collegio CarloAlberto

Recursive Smooth Ambiguity Preferences

Peter Klibanoff
Massimo Marinacci
Sujoy Mukerji

No.17, August 2006*
www.collegiocarloalberto.it
*Revised in August 2008

Recursive Smooth Ambiguity Preferences[1]

Peter Klibanoff
MEDS Dept., Kellogg School of Management, Northwestern University
peterk@kellogg.northwestern.edu

Massimo Marinacci
Collegio Carlo Alberto,
Università di Torino
massimo.marinacci@carloalberto.org

Sujoy Mukerji
Dept. of Economics, University of Oxford
sujoy.mukerji@economics.ox.ac.uk

This version: August 2008
First version: July 2006

[1]We thank Nabil Al-Najjar, Paolo Ghirardato, Christian Gollier, Eran Hanany, Ian Jewitt, Mark Machina, Luigi Montrucchio, Igor Prünster, Kevin Sheppard, Rakesh Vohra and Jonathan Weinstein for helpful discussion, Andreas Lange for pointing out an error in a previous version and seminar audiences at the Cowles Foundation workshop "Uncertainty in Economic Theory," the University of Iowa, Harvard/MIT, the University of Texas, Austin, the RUD 2005 conference, the 9^{th} World Congress of the Econometric Society, and the FUR XII conference for their comments.
c 2008 by Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji. Any opinions expressed here are those of the authors and not those of the Collegio Carlo Alberto.

Abstract

This paper axiomatizes an intertemporal version of the Smooth Ambiguity decision model developed in Klibanoff, Marinacci, and Mukerji (2005). A key feature of the model is that it achieves a separation between ambiguity, identified as a characteristic of the decision maker's subjective beliefs, and ambiguity attitude, a characteristic of the decision maker's tastes. In applications one may thus specify/vary these two characteristics independent of each other, thereby facilitating richer comparative statics and modeling flexibility than possible under other models which accomodate ambiguity sensitive preferences. Another key feature is that the preferences are dynamically consistent and have a recursive representation. Therefore techniques of dynamic programming can be applied when using this model.

JEL Classification Numbers: D800, D810.

Keywords: Ambiguity, Uncertainty, Knightian Uncertainty, Ambiguity Aversion, Uncertainty Aversion, Ellsberg Paradox, Dynamic Decision Making, Dynamic Programming under Ambiguity, Smooth Ambiguity.

1 Introduction

This paper axiomatizes and investigates a model of recursive preferences over intertemporal plans, extending the smooth ambiguity model developed in Klibanoff, Marinacci, and Mukerji (2005) (henceforth KMM) to a setting involving dynamic decision making.

In KMM we propose and axiomatize a model of preferences over acts such that the decision maker prefers act f to act g if and only if $\mathbb{E}_\mu \phi\left(\mathbb{E}_\pi u \circ f\right) \geq \mathbb{E}_\mu \phi\left(\mathbb{E}_\pi u \circ g\right)$, where $\mathbb{E}$ is the expectation operator, u is a vN-M utility function, ϕ is an increasing transformation, and μ is a subjective probability over the set Π of probability measures π that the decision maker thinks are relevant given his subjective information. A key feature of our model is that it achieves a separation between ambiguity, identified as a characteristic of the decision maker's subjective beliefs, and ambiguity attitude, a characteristic of the decision maker's tastes. We show that attitudes towards pure risk are characterized by the shape of u, as usual, while attitudes towards ambiguity are characterized by the shape of ϕ. Ambiguity itself is defined behaviorally and is shown to be characterized by properties of the subjective set of measures Π. One advantage of this model is that the well-developed machinery for dealing with risk attitudes can be applied as well to ambiguity attitudes. The model is also distinct from many in the literature on ambiguity in that it allows smooth, rather than kinked, indifference curves. This leads to different behavior and improved tractability, while still sharing the main features (e.g., Ellsberg's Paradox). The maxmin expected utility model (e.g., Gilboa and Schmeidler (1989)) with a given set of measures may be seen as a limiting case of our model with infinite ambiguity aversion.[1]

The functional representation obtained in KMM is particularly useful in economic modeling in answering comparative statics questions involving ambiguity. Take an economic model where agents' beliefs reflect some ambiguity. Next, without perturbing the information structure, it is useful to know how the equilibrium would change if the extent of ambiguity aversion were to decrease; e.g., if we were to replace ambiguity aversion with ambiguity neutrality, holding information and risk attitude fixed. (See, for example, Gollier (2005) for a portfolio choice application.) Another useful comparative statics exercise is to hold ambiguity attitudes fixed and ask how the equilibrium is affected if the perceived ambiguity is varied (see Jewitt and Mukerji (2006) for a definition and characterization of the notion of "more ambiguous"). Working out such comparative statics properly requires a model which allows a conceptual/parametric separation of (possibly) ambiguous beliefs and ambiguity attitude, analogous to the distinction usually made between risk and risk attitude. The model and functional representation in KMM allows that, whereas such a separation is not evident in the pioneering and most popular decision making models that incorporate ambiguity, namely, the multiple priors/maxmin expected utility (MEU) preferences (Gilboa and Schmeidler (1989)) and the Choquet expected utility (CEU) model of Schmeidler (1989).

While the preference model in KMM achieves the task of separating ambiguity and ambiguity attitude, the scope of application of this model is limited by the fact that it is a timeless framework. Yet many economic questions involving uncertain environments, especially in macroeconomics and finance, are more intuitively modeled using intertemporal

[1] For alternative developments of similar models see Ergin and Gul (2004), Nau (2006), Neilson (1993) and Seo (2006). All of these models draw inspiration from Segal (1987), the earliest paper relating ambiguity sensitive behavior to a two-stage functional relaxing reduction.

decision making frameworks. It is of interest to re-examine such questions by adding an ambiguity dimension. Computation and analysis of intertemporal choices is greatly facilitated by applying recursive methods. For these methods to be applicable, preferences have to satisfy a certain dynamic consistency property. A number of recent papers, including Epstein and Schneider (2003b), Wang (2003), Hayashi (2005), and Maccheroni, Marinacci, and Rustichini (2006b), have provided preference foundations for extending other ambiguity models to an intertemporal framework while satisfying this dynamic consistency. All of these, however, share the limitation inherent in the atemporal models they extend, of failing to separate ambiguity from ambiguity attitude without restricting the range of ambiguity attitudes. The present paper avoids this limitation, as does the contemporaneous work of Hanany and Klibanoff (2007a), the only other preference model we know of extending KMM to a dynamic setting. A major difference between the two extensions is that our model is recursive while that in Hanany and Klibanoff (2007a) is not. We discuss these and other papers further in Section 7.

The present paper, then, presents the first recursive model of intertemporal preferences that are ambiguity sensitive and dynamically consistent that *does* allow a separation of ambiguity from ambiguity attitude. To see why a such a preference model can be useful, recall the classic equity premium puzzle. The equity premium puzzle originally refers to the fact that the Lucas (Lucas (1978)) intertemporal general equilibrium model can fit data with steady consumption growth, low risk-free rates and high risk premiums only if the coefficient of (relative) risk aversion is allowed to be high in absolute value. This is a puzzle in the sense that, given the consumption risk identified in the data, the value of risk aversion coefficient required to explain the observed premium is incompatible with behavior under risk observed in other domains, for example in experiments. Various authors (e.g., Chen and Epstein (2002), Epstein and Wang (1994)) have shown theoretically using dynamic extensions of the MEU model that ambiguity aversion has the potential to explain the equity premium puzzle. They argue that ambiguity aversion adds to an agent's aversion to uncertainty over and above risk aversion, and so may add to the risk premium.[2] To evaluate the potential of an approach based on ambiguity, one has to be able to identify the extent of ambiguity aversion needed in conjunction with the ambiguity consistent with the data to explain the observed premium. Then one can check whether that value of ambiguity aversion agrees with levels observed when examining behavior on other domains. To conduct such an exercise, it is absolutely necessary to have a framework that allows one to separate ambiguity from ambiguity aversion and that allows for a variable degree of ambiguity aversion and a quantification of this degree. Only then may we be able to infer from the data a measure of ambiguity and ambiguity aversion. Indeed, recent work by Ju and Miao (2007) successfully applies the model developed in this paper along these lines (see also Hansen (2007)).

In the present paper the basis of the dynamic model is the state space S, the set of all observation paths generated by an event tree, a graph of decision/observation nodes. The root node of the tree, s^0, branches out into a set of (immediate) successor nodes each of which represents a stochastic contingency at time $t = 1$ and is generically denoted by s^1. Each node s^1 further branches out into (immediate) successor nodes s^2 at time $t = 2$ and so on into the infinite future. S is the set of all paths through this event tree; the

[2]However, see Gollier (2005) for caveats to this reasoning.

generic element of S is denoted by s. The decision maker's elements of choice are plans f each of which associates a payoff to each pair (t, s). The decision maker is (subjectively) uncertain about which stochastic process gives the appropriate description of probabilities on the event tree. The domain of this uncertainty is given by a finite parameter space Θ, each element of which (generically denoted by θ) is a vector of parameters exhaustively describing a particular stochastic process π_θ. We denote by $\pi_\theta(B \mid s^t)$ the probability under distribution π_θ that the observation path will belong to the set B, given that we have reached node s^t. Correspondingly, $\pi_\theta(x_{t+1}; s^t)$ is the probability under distribution π_θ that the next observation will be x_{t+1}, given we have reached node s^t. The decision maker's subjective prior belief about the stochastic process, elicited from his preferences, is described by a distribution μ, defined on 2^Θ, i.e., $\mu : 2^\Theta \to [0, 1]$. The decision maker's posterior belief about which stochastic process applies at a node s^t is given by the Bayesian posterior distribution $\mu(\cdot \mid s^t)$.

In the formal analysis that follows we obtain assumptions and conditions on preferences such that preferences on plans f, at a node s^t, are represented by the following recursive functional form:

$$V_{s^t}(f) = u(f(s^t)) + \beta\phi^{-1}\left[\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_\theta(x_{t+1}; s^t)\right) d\mu(\theta \mid s^t)\right],$$

where $V_{s^t}(f)$ is a recursively defined value function, u is a vN-M utility index, β is a discount factor and ϕ a function whose shape characterizes the DM's ambiguity attitude (as in KMM). Note that the represented preferences are recursive and dynamically consistent even though they depart from expected utility (and are also not probabilistically sophisticated (Machina and Schmeidler (1992))). Furthermore, there is a separation between ambiguity and ambiguity attitude along with flexibility in the modelling of each.

The next section sets out the formal structure of the model. Section 3 describes assumptions on preferences that are then shown to deliver the representation in Section 4. That section also includes a discussion of our approach and addresses some important existence and uniqueness questions. Results characterizing Bayesian updating and learning in the model appear in Section 5. Section 6 contains two examples: one comparing with recursive multiple priors, and one on the equity premium. Related literature is discussed in Section 7, followed by a brief concluding section. Proofs are contained in two appendices.

2 Set-Up

2.1 Modeling Information

Denote by $\mathcal{T}$ an infinite time horizon $\{1, ..., t, ...\}$. Consider a sequence of discrete random variables $\{X_t\}_{t\in\mathcal{T}}$ with values in finite observation spaces $\mathcal{X}_t$, endowed with their power sets $\mathcal{A}_t = 2^{\mathcal{X}_t}$. Each $\mathcal{X}_t$ consists of possible observations just before time t, and $x_t \in \mathcal{X}_t$ denotes a realization of the random variable X_t. For convenience, we assume that all X_t are surjective, so that $\mathcal{X}_t$ is the set of all possible realizations of each X_t.

Let $S = \prod_{t\in\mathcal{T}} \mathcal{X}_t$ be the set of all possible observation paths $s = (x_1, ..., x_t, ...)$. Denote by $S^t = \prod_{\tau=1}^t \mathcal{X}_\tau$ the collection of all finite paths $s^t = (x_1, ..., x_t)$. Each finite observation

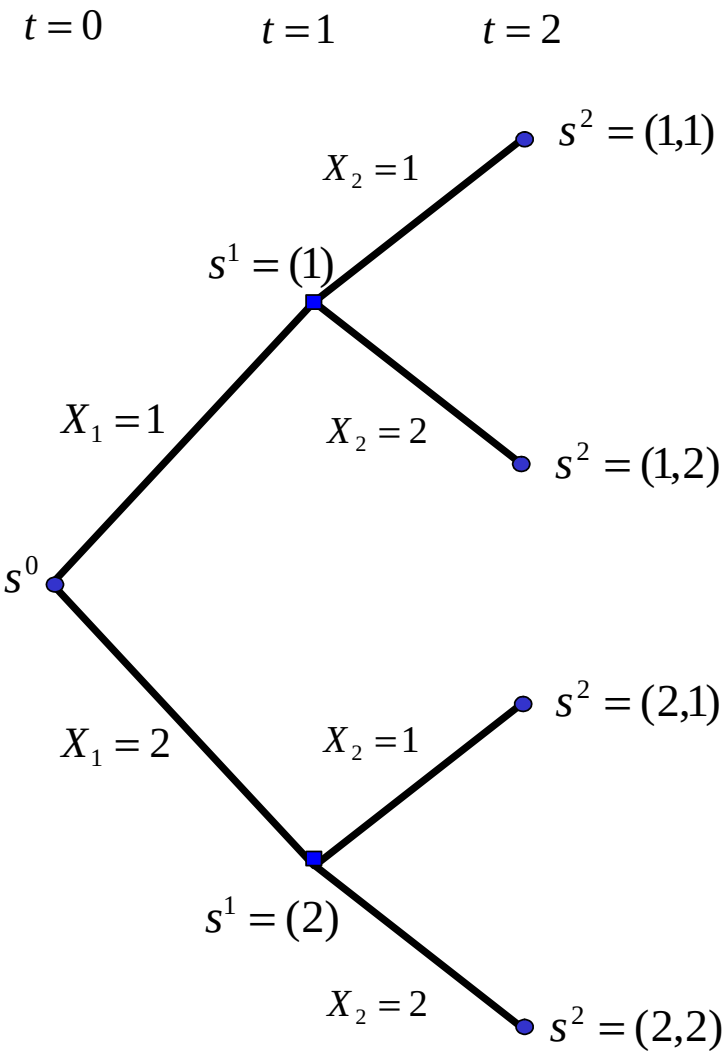

Figure 1: The first few time periods of an event tree with $\mathcal{X}_t = \{1, 2\}$.

path s^t identifies a decision/observation node, which is the history of observations up to time t. For this reason we denote by s^t a generic time t node, while $\mathcal{S}$ denotes the set of all nodes; i.e., $\mathcal{S} = \cup_{t \in \mathcal{T}} S^t$. On $\mathcal{S}$ there is a natural partial order, $\geqslant$, where $s^t \geqslant s^{t'}$ means that node s^t is a successor of node $s^{t'}$. In particular, $S^\tau(s^t)$ with $\tau \geq t$, is the set of time τ successors of the node s^t, i.e., $S^\tau(s^t) = \{s^\tau \in S^\tau : s^\tau \geqslant s^t\}$. Note that any node is considered a successor of itself. We often denote a generic element of $S^{t+1}(s^t)$ by (s^t, x_{t+1}) as it emphasizes the way in which the successor is generated.

On S we consider the product σ-algebra $\Sigma = \ _{t \in \mathcal{T}} \mathcal{A}_t$ generated by all one-dimensional cylinder sets $B_t \times \prod_{t \neq s \in \mathcal{T}} \mathcal{X}_s$, where $B_t \in \mathcal{A}_t$. For brevity, we denote by $\{B_1, ..., B_t\}$ the cylinder set with base $B_1 \times \cdots \times B_t$, namely,

$$\{B_1, ..., B_t\} = B_1 \times \cdots \times B_t \times \mathcal{X}_{t+1} \times \cdots$$

Figure 1 illustrates the first few time periods in our setting.In the figure, $S^1 = \{(1), (2)\}$ and $S^2 = \{(1,1), (1,2), (2,1), (2,2)\}$ are the sets of time 1 and time 2 nodes, respectively. An example of a cylinder set is the set of all observation paths with $X_1 = 1$ and $X_2 = 1$.

We define on Σ a family of probability distributions $\{\pi_\theta\}_{\theta\in\Theta}$, where Θ is a finite parameter space. Throughout we assume that $\pi_\theta(x_1,...,x_t) > 0$ for all "elementary" cylinder sets $\{x_1,...,x_t\}$.[3] For each $B \in \Sigma$,

$$\pi_\theta(B \mid s^t) = \begin{cases} \frac{\pi_\theta(B\cap\{x_1,...,x_t\})}{\pi_\theta(x_1,...,x_t)} & \text{if } t \in \mathcal{T} \\ \pi_\theta(B) & \text{if } t = 0 \end{cases}$$

is the conditional distribution of π_θ given $s^t = (x_1,...,x_t)$. In words, $\pi_\theta(B \mid s^t)$ is the probability under distribution π_θ that the observation path will belong to B, given we have reached node s^t.[4]

For each $t \in \mathcal{T}$, the one-step-ahead probability distribution $\pi_\theta(\cdot; s^t) : \mathcal{A}_{t+1} \to [0,1]$ at node $s^t = (x_1,...,x_t) \in S^t$ is determined by

$$\pi_\theta(x_{t+1}; s^t) = \pi_\theta\left(\{s^t, x_{t+1}\} \mid s^t\right) = \frac{\pi_\theta(x_1,...,x_t,x_{t+1})}{\pi_\theta(x_1,...,x_t)}, \qquad \forall x_{t+1} \in \mathcal{X}_{t+1},$$

where $\{s^t, x_{t+1}\}$ and s^t denote the elementary cylinder sets $\{x_1,...,x_t,x_{t+1}\}$ and $\{x_1,...,x_t\}$, respectively. In words, $\pi_\theta(x_{t+1}; s^t)$ is the probability under distribution π_θ that the next observation will be x_{t+1}, given we have reached node s^t.

The prior distribution μ is defined on 2^Θ, i.e., $\mu : 2^\Theta \to [0,1]$. Given any suitable $B \in \Sigma$, the posterior distribution $\mu(\cdot \mid B) : 2^\Theta \to [0,1]$ is defined as follows:

$$\mu(A \mid B) = \frac{\int_A \pi_\theta(B)\, d\mu}{\int_\Theta \pi_\theta(B)\, d\mu}, \qquad \forall A \in 2^\Theta, \forall B \in \Sigma.$$

In particular, if B is the cylinder $\{B_1,...,B_t\} \in \Sigma$, then

$$\mu(A \mid B_1,...,B_t) = \frac{\int_A \pi_\theta(B_1,...,B_t)\, d\mu}{\int_\Theta \pi_\theta(B_1,...,B_t)\, d\mu}, \qquad \forall A \in 2^\Theta.$$

Example 1 Assume $\mathcal{X}_t = \mathcal{X}$ for all t and suppose that each π_θ makes the sequence $\{X_t\}_{t\in\mathcal{T}}$ i.i.d., with common marginal distribution $q_\theta : \mathcal{A}_t \to [0,1]$. In this case, the probability distributions π_θ are given by the product probabilities $q_\theta^\infty : \Sigma \to [0,1]$, uniquely determined by q_θ as follows:

$$q_\theta^\infty(B_1,...,B_t) = \prod_{i=1}^{t} q_\theta(B_i),$$

[3] Observe the notational difference between the history $(x_1,...,x_t)$, an element of the Cartesian product $\prod_{\tau=1}^t \mathcal{X}_\tau$, and the cylinder $\{x_1,...,x_t\}$, the subset of Σ given by $\{x_1\} \times \cdots \times \{x_t\} \times \mathcal{X}_{t+1} \times \cdots$. For brevity, we write $\pi_\theta(x_1,...,x_t)$ in place of $\pi_\theta(\{x_1,...,x_t\})$.

[4] The finiteness assumptions on both the parameter space Θ and the observation spaces $\mathcal{X}_t$ along with the full support assumption on π_θ allow us to avoid a number of technical modeling issues as well as sidestep the issue of preferences following zero-probability events. These issues are not central to our paper. We believe that all but the results in Section 5 could be suitably extended to apply to non-finite environments and appropriately non-null events. For this reason we maintain integral notation. See Section 5 for some discussion of the role of finiteness there.

for each cylinder set $\{B_1, ..., B_t\} \in \Sigma$. Hence, for each node $s^t \in S^t$ and observation $x_{t+1} \in X_{t+1}$, we have

$$\pi_\theta\left(x_{t+1} \,\square\, s^t\right) = \frac{\prod_{i=1}^{t+1} q_\theta(x_i)}{\prod_{i=1}^{t} q_\theta(x_i)} = q_\theta(x_{t+1}),$$

while, for each cylinder set $\{B_1, ..., B_t\} \in \Sigma$ and each set $A \subseteq \Theta$,

$$\mu(A \mid B_1, ..., B_t) = \frac{\int_A q_\theta^\infty(B_1, ..., B_t)\, d\mu}{\int_\Theta q_\theta^\infty(B_1, ..., B_t)\, d\mu} = \frac{\int_A \prod_{i=1}^{t} q_\theta(B_i)\, d\mu}{\int_\Theta \prod_{i=1}^{t} q_\theta(B_i)\, d\mu}.$$

▲

Example 2 Assume again $\mathcal{X}_t = \mathcal{X}$ for all t and suppose now that each π_θ makes the sequence $\{X_t\}_{t\in\mathcal{T}}$ a homogeneous Markov chain with transition function $q_\theta : \mathcal{X}_{t\square 1} \times \mathcal{A}_t \to [0,1]$ for $t \geq 2$, where $q_\theta(x, \cdot) : \mathcal{A}_t \to [0,1]$ is a probability measure on $\mathcal{A}_t$ for each $x \in \mathcal{X}_{t\square 1}$, and $q_\theta(\cdot, B_t) : \mathcal{X}_{t\square 1} \to [0,1]$ is an $\mathcal{A}_{t\square 1}$-measurable function for each $B_t \in \mathcal{A}_t$. Given an initial probability distribution $q_\theta(s^0)$ on $\mathcal{A}_1$, the probability distributions π_θ are uniquely determined by q_θ as follows:

$$\pi_\theta(B_1, ..., B_t) = q_\theta\left(\square\, s^0\right)(B_1) \prod_{i=1}^{t\square 1} \int_{B_i} q_\theta(dx_i, B_{i+1})$$

for each cylinder set $\{B_1, ..., B_t\} \in \Sigma$. Hence, for each $s^t = (x_1, ..., x_t) \in S^t$ and $x_{t+1} \in \mathcal{X}_{t+1}$,

$$\pi_\theta\left(x_{t+1} \,\square\, s^t\right) = \frac{q_\theta(s^0)(x_1) \prod_{i=1}^{t} q_\theta(x_i, x_{i+1})}{q_\theta(s^0)(x_1) \prod_{i=1}^{t\square 1} q_\theta(x_i, x_{i+1})} = q_\theta(x_t, x_{t+1}),$$

while, for each $\{B_1, ..., B_t\} \in \Sigma$ and each $A \subseteq \Theta$,

$$\mu(A \mid B_1, ..., B_t) = \frac{\int_A q_\theta(s^0)(B_1) \left(\prod_{i=1}^{t\square 1} \int_{B_i} q_\theta(dx_i, B_{i+1})\right) d\mu}{\int_\Theta q_\theta(s^0)(B_1) \left(\prod_{i=1}^{t\square 1} \int_{B_i} q_\theta(dx_i, B_{i+1})\right) d\mu}.$$

▲

2.2 Plans

At the initial time ($t = 0$) and at each subsequent time t, the decision maker (DM) chooses a "consumption" plan, detailing current and future consumption. The available information at t is given by the realizations of the random variables $X_1, ..., X_t$. At the initial time $t = 0$, the DM has not yet observed any realizations.

Let $\mathcal{C}$ be a payoff space, which is assumed to be a compact interval in $\mathbb{R}$ endowed with the standard topology of the real line. Consider a map $f : \mathcal{T} \cup \{0\} \times S \to \mathcal{C}$ that associates a payoff to each pair (t, s). We can represent f as a collection $\{f_t\}_{t\in\mathcal{T}\cup\{0\}}$, with $f_t : S \to \mathcal{C}$ for each $t \geq 0$.

We say that f is a *plan* if f_0 is $\{\varnothing, S\}$-measurable (i.e., it is a constant) and each f_t is $\sigma(X_1, ..., X_t)$-measurable for each $t \in \mathcal{T}$. In other words, setting $\Sigma_t = \sigma(X_1, ..., X_t)$ for each

$t \in \mathcal{T}$ and $\Sigma_0 = \{\varnothing, S\}$, f is a plan whenever it is a payoff stream adapted to the filtration $\{\Sigma_t\}_{t\in\mathcal{T}\cup\{0\}}$. As $\sigma(X_1, ..., X_t)$ is the σ-algebra generated by the cylinder sets $\{x_1, ..., x_t\}$, each f_t actually depends only on the finite paths $s^t \in S^t$ and $f_t(x_1, ..., x_t)$ reads as follows: "$f_t(x_1, ..., x_t) \in \mathcal{C}$ is the time t payoff the DM receives if he arrives at the node identified by $s^t = (x_1, ..., x_t)$." Here we adopt the convention $s^0 = \{\varnothing\}$.

For this reason we can regard a plan f as a function $f : \mathcal{S} \to \mathcal{C}$ defined over the set of nodes $\mathcal{S}$. For convenience, this is the way we will view plans throughout the paper. Denote the set of all such plans by $\mathcal{F}$.

A subset of plans of special interest are the deterministic plans. Let $d \in \mathcal{F}$ denote a deterministic plan, i.e., a payoff stream such that the payoff obtained at each time t does not depend on the particular node reached at that time. Formally, for each t there is $c \in \mathcal{C}$, denoted by $c(t)$, such that $d(s^t) = c(t)$, $\forall s^t \in S^t$. Finally, given $c \in \mathcal{C}$, with a (standard) slight abuse of notation we will also use c to denote the constant stream that pays off c at each node.

2.3 Continuation Plans

It will be important to have a notion of a continuation plan that specifies payoffs only from a node s^t onward. We define the set of *continuation plans at node* s^t to be the restriction of the set of plans $\mathcal{F}$ to successors of s^t, i.e., to $\cup_{\tau\geq t} S^\tau(s^t)$, and denoted it by $\mathcal{F}_{s^t}$ with generic element f_{s^t}. Given a plan f, the *continuation plan induced by* f *at* s^t, denoted $f_{|s^t}$, is the restriction of f to successors of s^t. Observe that while f_{s^t} is a generic function from $\cup_{\tau\geq t} S^\tau(s^t)$ to $\mathcal{C}$, $f_{|s^t}$ is a specific such function determined by f.

We next introduce the concept of *mixed* continuation plans. An s^t-mixed continuation plan is a probability distribution on $\mathcal{F}_{s^t}$. For example, given a finite collection $\{f_i\}_{i=1}^n$ of continuation plans in $\mathcal{F}_{s^t}$, and a corresponding set of probabilities $\{p_i\}_{i=1}^n$, $\{f_i, p_i\}_{i=1}^n$ denotes the s^t-mixed continuation plan given by the lottery

$$\begin{cases} f_1(s^\tau) & \text{for each } s^\tau \geqslant s^t \text{ with probability } p_1, \\ \quad\vdots & \qquad\qquad\vdots \\ f_n(s^\tau) & \text{for each } s^\tau \geqslant s^t \text{ with probability } p_n. \end{cases}$$

In words, the s^t-mixed continuation plan corresponds to a randomization at node s^t over the continuation plans $\{f_i\}_{i=1}^n$ in $\mathcal{F}_{s^t}$, such that with (an objective) probability p_i the continuation plan f_i is followed on all nodes that succeed s^t.

Though these mixed plans can be studied in full generality, for our purposes it is enough to consider mixing over deterministic plans. Let $\mathcal{D}_{s^t}$ be the subset of $\mathcal{F}_{s^t}$ consisting of the deterministic continuation plans at s^t. As $\mathcal{D}_{s^t} = \prod_{\tau\geq t} \mathcal{C}$, we can endow $\mathcal{D}_{s^t}$ with the product topology. Below, we make use of the set of s^t-mixed deterministic continuation plans given by the set $\mathcal{P}_{s^t}$ of all countably additive Borel probability measures defined on $(\mathcal{D}_{s^t}, \mathcal{B}_{s^t})$, where $\mathcal{B}_{s^t}$ is the Borel σ-algebra induced by the product topology.

2.4 Second Order Acts

Definition 1 *A **second order act** is any function* $\mathfrak{f} : \Theta \to \mathcal{C}$ *that associates an element of* Θ *to a payoff. We denote by* $\mathfrak{F}$ *the set of all second order acts.*

We introduced second-order acts in KMM and used preferences over them to reveal the DM's subjective uncertainty about the probabilities of the observations. They will play a similar role in this paper. There is a question whether preferences over these acts are observable, and thus, the extent to which assumptions on such preferences are behaviorally meaningful. For example, in what sense can one check that the DM prefers a bet paying off if π_{θ_1} is the process governing the observations to a bet paying off if π_{θ_2} is the true process? The only sense in which this could be more difficult than observing choices among usual Savage acts is that one could doubt that the true process could be verified. However, in a dynamic environment as in the current paper, given enough time and sufficient stationarity of the processes, one can discriminate as finely as desired among the finite set of parameters Θ. This follows from recent results in the literature on Bayesian consistency (see Lijoi, Prunster, and Walker (2007)) which imply that under stationarity assumptions the estimator constructed by updating a strictly positive (e.g., uniform) distribution over Θ by Bayes' rule is a consistent estimator of the true parameter.

In this sense, a bet that this estimator will assign sufficiently high weight to θ_1 at a date sufficiently far in the future, is approximately a bet on θ_1 (i.e., a bet that the true process is π_{θ_1}). Further informative discussion of the nature and verifiability of second order acts is contained in KMM, pp. 1854 and 1856.

2.5 Preferences

We will be concerned with preferences over two domains. Of primary interest are the DM's preferences at each node s^t over plans and mixtures over deterministic continuation plans, $\mathcal{F} \cup \mathcal{P}_{s^t}$. At each s^t, this is denoted by the binary relation $\succcurlyeq_{s^t}$. The second order acts, $\mathfrak{F}$, are the other domain of interest. Let $\succcurlyeq^2_{s^t}$ denote the DM's preference ordering over $\mathfrak{F}$ at node s^t.

3 Assumptions on Preferences

Our assumptions on preferences fall naturally into two major categories. First are four assumptions that are quite standard for an intertemporal setting and are not tied in any special way to the smooth ambiguity model. They are followed by four assumptions analogous to the three assumptions KMM used to derive the (atemporal) smooth ambiguity model. It is how these last assumptions are integrated into the intertemporal setting that is key in determining the sense in which the current theory is an extension of the earlier one.

3.1 Four Standard assumptions

The assumptions in this section are imposed on preferences over plans at each time t and each node $s^t \in S^t$.

Assumption 1 (Weak Order) *$\succcurlyeq_{s^t}$ on $\mathcal{F}$ is complete and transitive.*

Assumption 2 (Monotonicity) *Given any $f, g \in \mathcal{F}$, if $f(s) \geq g(s)$ for all $s \geq s^t$, then $f \succcurlyeq_{s^t} g$. If, in addition, f and g are deterministic and $f(\tau) > g(\tau)$ for at least some $\tau \geq t$, then $f \succ_{s^t} g$.*[5]

Assumption 3 (Archimedean) *Given any $f \in \mathcal{F}$ and $c_1, c_2 \in \mathcal{C}$, if $c_1 \succ_{s^t} f \succ_{s^t} c_2$, then there exist $\alpha, \beta \in (0,1)$ such that $\alpha c_1 + (1 \square \alpha)c_2 \succ_{s^t} f \succ_{s^t} \beta c_1 + (1 \square \beta)c_2$.*

Note, in the above assumption, $\alpha c_1 + (1 \square \alpha)c_2$ is simply a constant plan yielding the real number $\alpha c_1 + (1 \square \alpha)c_2$ at each node.

Assumption 4 (Dynamic Consistency) *Given any $f, g \in \mathcal{F}$, if $f(s^t) = g(s^t)$ then*

$$f \succcurlyeq_{(s^t, x_{t+1})} g \text{ for all } x_{t+1} \Longrightarrow f \succcurlyeq_{s^t} g.$$

The first three assumptions are standard and are clearly necessary for any monotonic, real-valued and suitably continuous representation. Observe that monotonicity immediately implies consequentialism, i.e., given any $f, g \in \mathcal{F}$

$$f_{|s^t} = g_{|s^t} \Longrightarrow f \sim_{s^t} g. \tag{1}$$

In words, in evaluating plans at a node only payoffs from that point onward matter.[6] This rules out past realized payoffs or payoffs at unrealized nodes affecting preference. There are a number of circumstances in which it may be plausible for consequentialism to be violated (see e.g., Machina (1989) concerning risk and Pollak (1970) or Abel (1990) from the large literature on habit formation in consumption). However, it is quite useful in practice as it ensures that each problem may be approached and analyzed at the time it is encountered without regard to what occurred or did not occur in the (often unknown and possibly complex) past.[7]

Of the four assumptions above, the only one that imposes conditions relating preferences at different nodes is dynamic consistency. Some condition across preferences at different nodes is needed to pin down dynamic behavior. Dynamic consistency is the standard way of doing this in the economics literature, as some variation of it is inherent in any recursive functional form. Informally, dynamic consistency says that if two plans give the same payoff today, and, no matter what happens in between, one of the plans is always preferred to the other at the next time, then that plan should also be preferred today. It's essence is that today's preference over plans that may differ only from tomorrow onward should agree with "tomorrow's preference" over those plans in cases where "tomorrow's preference" is unanimous (i.e., does not depend on what is observed between today and tomorrow).

[5] In the second part of the axiom, time is the only argument of the acts as they are deterministic.

[6] Note that (1) is, as Hanany and Klibanoff (2007b) emphasize, only part of consequentialism. The assumption that preferences are conditioned only on nodes, and not, for example, on other aspects of the choice problem such as the feasible set of plans available at the initial node, is also part of consequentialism. This aspect of consequentialism is built in to our framework by the fact that the collection of preferences over acts we examine is subscripted only by nodes, s^t.

[7] It would be messy, but not conceptually difficult, to extend our analysis to cover the case where past payoffs may affect preferences. Essentially the functions in our representation would all need to have the past payoff history as an additional argument. See e.g., Kreps and Porteus (1978) and Johnsen and Donaldson (1985) for examples of recursive structures with history-dependence.

3.2 Embedding the (Atemporal) KMM Representation

We now lay out the additional assumptions that, together the standard four above, will deliver recursive smooth ambiguity preferences. To help motivate these additional assumptions and to elucidate the sense in which the recursive theory extends the smooth ambiguity model, it is useful to briefly review the three assumptions used to derive the smooth ambiguity model in KMM. The first KMM assumption is that preferences over a subclass of acts called lottery acts (which may be thought of as essentially lotteries over outcomes) are continuous expected utility preferences. A natural analogue to lottery acts in the setting of the current paper are the s^t-mixed deterministic continuation plans, $\mathcal{P}_{s^t}$, that were defined in section 2.3. Rather than lotteries over single outcomes, these are lotteries over deterministic streams of outcomes, as is appropriate to our intertemporal setting. Our first assumption in this section will be that at each node s^t preferences over $\mathcal{P}_{s^t}$ are continuous expected discounted utility preferences. Our second assumption in this section guarantees that intertemporal and risk trade-offs are stable across nodes. In terms of the representation, this will be equivalent to constancy across nodes of the discount factor and the utility function. These two assumptions together form a natural intertemporal analogue to the first assumption in KMM. The second KMM assumption states that preferences over second order acts are continuous subjective expected utility preferences. In the present context, second order acts are functions from the parameter space, Θ, to the outcome space, $\mathcal{C}$. Therefore, our third assumption will be that, at each node, preferences over second order acts are continuous subjective expected utility preferences. Moreover, we will assume that the utility function is stable across nodes. KMM also proposes a natural association between acts and second order acts that works as follows: given an act, for each parameter value, the associated second order act yields the certainty equivalent of the act under the distribution implied by the parameter value. Finally, the third KMM assumption requires preference over acts to agree with preference over the associated second order acts. In our dynamic setting, our last assumption is also of this form. However, the role of acts is played by continuation plans at s^t that are constant from time $t+1$ onward but where that constant may depend on the node reached at $t+1$ ("one-step-ahead continuation plans") and the role of certainty equivalent of such a continuation plan given a parameter value is played by the outcome that, if received from $t+1$ onward in place of what the continuation plan would give from then on leaves the DM indifferent. This fourth assumption makes clear the sense in which the intertemporal model extends the KMM model: KMM applies at time t when considering plans where all the uncertainty is resolved between t and $t+1$. We now make all of this more formal.

From the above discussion, it is clear that we need to consider preferences over s^t-mixed deterministic continuation plans in addition to preferences over plans. The next assumption says that $\succcurlyeq_{s^t}$ on $\mathcal{P}_{s^t}$ is continuous expected discounted utility and identifies, from the point of view of preference, a degenerate mixtures in $\mathcal{P}_{s^t}$ with a corresponding plan. In reading it recall that a deterministic continuation plan d_{s^t} might well be induced by a plan which is deterministic from s^t onwards but not deterministic overall.

Assumption 5 (Discounting) *For each $d_{s^t} \in \mathcal{D}_{s^t}$, there exists a plan $f \in \mathcal{F}$ with $f_{|s^t} = d_{s^t}$ such that $f \sim_{s^t} \{d_{s^t}, 1\}$. At each node $s^t \in S^t$, $\succcurlyeq_{s^t}$ on $\mathcal{P}_{s^t}$ is represented by the expectation*

of a von Neumann-Morgenstern utility index $U_{s^t} : \mathcal{D}_{s^t} \to \mathbb{R}$ *which has the form*

$$U_{s^t}(d_{s^t}) = \sum_{\tau \geq t} \beta_{s^t}^{\tau \square t} u_{s^t}(d_{s^t}(\tau)) \tag{2}$$

for $\beta_{s^t} \in (0,1)$, u_{s^t} *continuous and not constant on* $\mathcal{C}$. *Moreover,* β_{s^t} *is unique and* u_{s^t} *is unique up to positive linear transformations.*

Note that this assumption is not stated purely in terms of preferences, as the preceding ones are. This should not be bothersome, however, as a purely preference-based foundation for the utility indices, U_{s^t}, can be done exactly as in Theorem 2 of Epstein (1983).

Next is the invariance assumption that will guarantee that β_{s^t} and u_{s^t} in Assumption 5 do not depend on the node s^t. To state it we need some notation: given any $c, c', c'' \in \mathcal{C}$, we denote by $(c', c'', c)_t$ the deterministic continuation plan that pays off c' at t, c'' at $t+1$, and c from $t+2$ onwards.

Assumption 6 (Invariance) *Given any* $t, t' \in \mathcal{T} \cup \{0\}$, *any* $p \in [0,1]$, *and any set of consequences* $c_i, c_i', c_i'', k_i, k_i', k_i''$ *with* $i = 1, 2$, *we have:*

$$\{(c_1', c_1'', c_1)_t, p; (k_1', k_1'', k_1)_t, 1 \square p\} \succcurlyeq_{s^t} \{(c_2', c_2'', c_2)_t, p; (k_2', k_2'', k_2)_t, 1 \square p\}$$

if and only if

$$\{(c_1', c_1'', c_1)_{t'}, p; (k_1', k_1'', k_1)_{t'}, 1 \square p\} \succcurlyeq_{s^{t'}} \{(c_2', c_2'', c_2)_{t'}, p; (k_2', k_2'', k_2)_{t'}, 1 \square p\}.$$

This has the stated consequence:

Lemma 1 *Under Assumption 6, there is* $\beta \in (0,1)$ *and* $u : \mathcal{C} \to \mathbb{R}$ *continuous and not constant such that in Assumption 5,* $\beta_{s^t} = \beta$ *and* $u_{s^t} = u$ *for all* s^t.

Our next assumption states that preferences over second order acts have a continuous subjective expected utility representation.

Assumption 7 (SEU on Second Order Acts) *There exists a unique (additive) probability* $\mu_{s^t} : 2^{\Theta} \to [0,1]$ *and a continuous, strictly increasing* $v : \mathcal{C} \to \mathbb{R}$ *such that, for all* $\mathfrak{f}, \mathfrak{g} \in \mathfrak{F}$,

$$\mathfrak{f} \succcurlyeq_{s^t}^2 \mathfrak{g} \iff \int_{\Theta} v(\mathfrak{f}(\theta))\, d\mu_{s^t} \geq \int_{\Theta} v(\mathfrak{g}(\theta))\, d\mu_{s^t}.$$

Moreover, v *is unique up to positive affine transformations, provided* $0 < \mu_{s^t}(J) < 1$ *for some* $J \subseteq \Theta$.

Remark 1 As is true for Assumption 5 (Discounting), more primitive axioms, such as those used in Theorem V.6.1 of Wakker (1989), could be applied to $\succcurlyeq_{s^t}^2$ to deliver this assumption. Notice, as v does not have the subscript s^t, Assumption 7 implicitly assumes that $\succcurlyeq_{s^t}^2$ restricted to constant elements of $\mathfrak{F}$ does not vary with s^t. Moreover, observe if there is no $J \subseteq \Theta$, such that $0 < \mu_{s^t}(J) < 1$, then v can be unique up to only monotone transformations.

To state our final assumption, we first need to define one-step-ahead continuation plans and one-step-ahead certainty equivalents. Let $\mathcal{F}^*_{s^t}$ be the subset of continuation plans in $\mathcal{F}_{s^t}$ that yield a stream, constant from $t+1$ onwards, that only depends on the immediate successor node of s^t. These are one-step-ahead continuation plans. Formally, $\mathcal{F}^*_{s^t}$ is the set

$$\left\{ f_{s^t} \in \mathcal{F}_{s^t} : f_{s^t}(s^\tau) = f_{s^t}(s^{t+1}) \quad \forall\, s^{t+1} \in S^{t+1} \square s^t \right), \forall\, s^\tau \in S^\tau \square s^{t+1} \right), \forall \tau > t+1 \Big\}.$$

Given any continuation plan $f_{s^t} \in \mathcal{F}_{s^t}$, consider for each $x_{t+1} \in \mathcal{X}_{t+1}$ the deterministic continuation plan

$$d_{f_{s^t}}(x_{t+1}) = \square f_{s^t} \square s^t \right), f_{s^t}(s^t, x_{t+1}), f_{s^t}(s^t, x_{t+1}), ..., f_{s^t}(s^t, x_{t+1}), ... \right).$$

Formally, $d_{f_{s^t}}(x_{t+1}) \in \mathcal{F}^*_{s^t}$ is given by $d_{f_{s^t}}(x_{t+1})(t) = f_{s^t}(s^t)$ and $d_{f_{s^t}}(x_{t+1})(\tau) = f_{s^t}(s^t, x_{t+1})$ for all $\tau \geq t+1$.

Definition 2 *Given any one-step-ahead continuation plan $f_{s^t} \in \mathcal{F}^*_{s^t}$, the **one-step-ahead certainty equivalent** of f_{s^t} given θ, $c_{f_{s^t}}(\theta)$, is a payoff c such that, for any $g \in \mathcal{F}$, if $g_{|s^t} = (f_{s^t}(s^t), c, c, ..., c, ...)$, then*

$$g \sim_{s^t} \left\{ d_{f_{s^t}}(x_{t+1}), \pi_\theta \square x_{t+1}; s^t \right) \right\}_{x_{t+1} \in \mathcal{X}_{t+1}}.$$

In other words, the one-step-ahead certainty equivalent of the one-step-ahead continuation plan f_{s^t} is the payoff making the DM indifferent at s^t between the continuation plan having that payoff from time $t+1$ onwards and the s^t-mixed plan which pays off $d_{f_{s^t}}(x_{t+1})$ with probability $\pi_\theta(x_{t+1}; s^t)$. Observe that such a mixed plan is indeed an element of $\mathcal{P}_{s^t}$.

Definition 3 *Given $f_{s^t} \in \mathcal{F}^*_{s^t}$, $f^2 \in \mathfrak{F}$ denotes a **second order act associated with** f_{s^t}, defined as follows*

$$f^2(\theta) = c_{f_{s^t}}(\theta) \qquad \text{for all } \theta \in \Theta.$$

Assumption 8 (Consistency with Associated Second Order Acts) *Given $f_{s^t}, g_{s^t} \in \mathcal{F}^*_{s^t}$ and associated $f^2, g^2 \in \mathfrak{F}$, if $f_{s^t}(s^t) = g_{s^t}(s^t)$,*

$$f \succcurlyeq_{s^t} g \iff f^2 \succcurlyeq^2_{s^t} g^2$$

for some $f, g \in \mathcal{F}$ with $f_{|s^t} = f_{s^t}$ and $g_{|s^t} = g_{s^t}$.

4 Main Representation Theorem

In this section, we show that our assumptions yield a recursive smooth ambiguity representation of $\succcurlyeq_{s^t}$ over plans. To do this, it will be helpful to be able to refer directly to induced preferences over continuation plans rather than always referring back to preferences over plans. To this end, consider for each t and s^t, induced preference relations $\succcurlyeq^*_{s^t}$ on $\mathcal{F}_{s^t} \cup \mathcal{P}_{s^t}$, the union of continuation plans at node s^t and s^t-mixed deterministic continuation plans. We call these induced preferences because they are fully determined by $\succcurlyeq_{s^t}$. On $\mathcal{P}_{s^t}$ define $\succcurlyeq^*_{s^t} = \succcurlyeq_{s^t}$. On the set of continuation plans, $\mathcal{F}_{s^t}$, define $\succcurlyeq^*_{s^t}$ as follows:

$$f_{|s^t} \succcurlyeq^*_{s^t} g_{|s^t} \Longleftrightarrow f \succcurlyeq_{s^t} g, \qquad \forall f_{|s^t}, g_{|s^t} \in \mathcal{F}_{s^t}.$$

Observe that each element of $\mathcal{F}_{s^t}$ can be represented as $f_{|s^t}$, where f is a suitable plan in the sense of having that element of $\mathcal{F}_{s^t}$ as continuation at s^t. By consequentialism, as expressed in (1), the particular plan f used for such representation is immaterial, and so $\succcurlyeq^*_{s^t}$ is well defined on $\mathcal{F}_{s^t}$. Finally, when comparing elements of $\mathcal{F}_{s^t}$ and $\mathcal{P}_{s^t}$, complete the specification of $\succcurlyeq^*_{s^t}$ from $\succcurlyeq_{s^t}$ in the obvious manner:

$$f_{|s^t} \succcurlyeq^*_{s^t} p \Longleftrightarrow f \succcurlyeq_{s^t} p, \text{ and } p \succcurlyeq^*_{s^t} f_{|s^t} \Longleftrightarrow p \succcurlyeq_{s^t} f, \qquad \forall f \in \mathcal{F}, p \in \mathcal{P}_{s^t}.$$

We begin by using the assumptions to get a representation of $\succcurlyeq^*_{s^t}$ restricted to plans in $\mathcal{F}^*_{s^t}$ sharing the same payoff at s^t. This will be a key step in deriving the main representation theorem. Denote by $\mathcal{U}$ the range of $(1 \boxdot \beta)^{\boxdot 1} u$.

Proposition 1 *Suppose Assumptions 1-8 hold. Then there exist continuous and strictly increasing functions $u : \mathcal{C} \to \mathbb{R}$ and $\phi : \mathcal{U} \to \mathbb{R}$, and an additive probability $\mu_{s^t} : 2^\Theta \to [0,1]$ such that, given any $f_{s^t}, g_{s^t} \in \mathcal{F}^*_{s^t}$ with $f_{s^t}(s^t) = g_{s^t}(s^t)$, we have $f_{s^t} \succcurlyeq^*_{s^t} g_{s^t}$ if and only if*

$$\begin{aligned} & \int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(f_{s^t}(s^t, x_{t+1}))}{1 \boxdot \beta} d\pi_\theta \overset{\boxdot}{} x_{t+1}; s^t\right) d\mu_{s^t} \\ \geq & \int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(g_{s^t}(s^t, x_{t+1}))}{1 \boxdot \beta} d\pi_\theta \overset{\boxdot}{} x_{t+1}; s^t\right) d\mu_{s^t}. \end{aligned}$$

The probability μ_{s^t} is unique and the function u is unique up to positive affine transformations. Moreover, given u, the function ϕ is unique up to positive affine transformations, provided $0 < \mu_{s^t}(J) < 1$ for some $J \subseteq \Theta$. If $\tilde{u} = \alpha u + \delta$, $\alpha > 0$, then the associated $\tilde{\phi}$ is such that $\tilde{\phi}(\alpha y + \frac{\delta}{1 \boxdot \delta}) = \phi(y)$, where $y \in \mathcal{U}$.

Given $c \in \mathcal{C}$ and $s^t \in S^t$, let $c_{s^t} \in \mathcal{F}_{s^t}$ be the deterministic continuation plan such that $c_{s^t}(\tau) = c$ for all $\tau \geq t$.

Definition 4 *Given $f \in \mathcal{F}$ and $s^t \in S^t$, the **continuation certainty equivalent**, c_{f,s^t}, of f at s^t is a payoff $c \in \mathcal{C}$ such that $c_{s^t} \sim^*_{s^t} f_{|s^t}$.*

Lemma 2 *For each f and s^t, there exists a continuation certainty equivalent.*

Denote by $\vec{c}_{f,s^t}$ the constant continuation plan in $\mathcal{F}_{s^t}$ paying off c_{f,s^t} at s^t and at all nodes that follow it.

Definition 5 *The **continuation value** of f at s^t is given by $V_{s^t}(f) = U(\vec{c}_{f,s^t})$.*

Given our assumptions, the function $V_{s^t} : \mathcal{F} \to \mathbb{R}$ represents $\succcurlyeq_{s^t}$ on plans. In fact, by definition and by Assumption 5, we have:

$$\begin{aligned} f \succcurlyeq_{s^t} g & \Longleftrightarrow f_{|s^t} \succcurlyeq^*_{s^t} g_{|s^t} \Longleftrightarrow \vec{c}_{f,s^t} \succcurlyeq^*_{s^t} \vec{c}_{g,s^t} \\ & \Longleftrightarrow U(\vec{c}_{f,s^t}) \geq U(\vec{c}_{g,s^t}) \Longleftrightarrow V_{s^t}(f) \geq V_{s^t}(g). \end{aligned}$$

The Monotonicity assumption implies that if $f(s) \geq g(s)$ for all $s \geq s^t$, then $V_{s^t}(f) \geq V_{s^t}(g)$. If, in addition, f and g are deterministic and $f(s) > g(s)$ for at least some $s \geq s^t$, then Monotonicity implies $V_{s^t}(f) > V_{s^t}(g)$. We refer to a functional V_{s^t} as *monotonic* if it satisfies these two properties given by Monotonicity.

We can now state our main representation result:

Theorem 1 *For each t and s^t, let $\succcurlyeq_{s^t}$ be a binary relation on $\mathcal{F} \cup \mathcal{P}_{s^t}$ and $\succcurlyeq^2_{s^t}$ be a binary relation on $\mathfrak{F}$. Assumptions 1-8 hold if and only if there exists an additive probability $\mu_{s^t} : 2^{\Theta} \to [0,1]$ for each s^t and continuous and strictly increasing functions $u : \mathcal{C} \to \mathbb{R}$ and $\phi : \mathcal{U} \to \mathbb{R}$, such that*

(i) On $\mathcal{F}$, each preference $\succcurlyeq_{s^t}$ is represented by the monotonic recursive functional $V_{s^t} : \mathcal{F} \to \mathbb{R}$ given by

$$V_{s^t}(f) = u\left(f(s^t)\right) + \beta\phi^{-1}\left[\int_{\Theta}\phi\left(\int_{\mathcal{X}_{t+1}} V_{(s^t,x_{t+1})}(f)\, d\pi_{\theta}(\cdot; s^t)\right) d\mu_{s^t}\right]; \qquad (3)$$

(ii) On $\mathcal{P}_{s^t}$, each preference $\succcurlyeq_{s^t}$ is represented by the expected discounted utility functional $EU_{s^t} : \mathcal{P}_{s^t} \to \mathbb{R}$ given by

$$EU_{s^t}(p) = \int_{\mathcal{D}_{s^t}}\left[\sum_{\tau=t}^{\infty}\beta^{\tau-t}u\left(d(\tau)\right)\right] dp;$$

(iii) On $\mathfrak{F}$, each preference $\succcurlyeq^2_{s^t}$ is represented by the subjective expected utility functional $V^2_{s^t} : \mathfrak{F} \to \mathbb{R}$ given by

$$V^2_{s^t}(\mathfrak{f}) = \int_{\Theta}\phi\left(\frac{u\left(\mathfrak{f}(\theta)\right)}{1-\beta}\right) d\mu_{s^t}.$$

The uniqueness properties of μ_{s^t}, u and ϕ are as stated in Proposition 1.

Equation (3) in (i) is our desired recursive representation of preferences over plans at each node and is the main result of this theorem. The representations in parts (ii) and (iii) over mixed deterministic continuation plans and second order acts, respectively, follow quite directly from the discounting and invariance assumptions and the assumption of SEU on second order acts. Notice that in equation (3), the aggregation across possible time $t+1$ continuation values is done according to the KMM smooth ambiguity functional – this links the dynamic and timeless smooth ambiguity models. The term $\phi^{-1}\left[\int_{\Theta}\phi\left(\int_{\mathcal{X}_{t+1}} V_{(s^t,x_{t+1})}(f)\, d\pi_{\theta}(\cdot; s^t)\right) d\mu_{s^t}\right]$ should be interpreted as the utility equivalent of the possibly ambiguous (in the sense that each θ may induce a different distribution over) continuation values.

Although it may not be immediately apparent, in the case of ambiguity neutrality (i.e., ϕ affine), (3) reduces to subjective expected discounted utility *with Bayesian updating*. We show this formally in the next proposition, where we make explicit the corresponding prior belief. We also show that this belief is relevant to and recoverable from preferences with any ambiguity attitude, as long as ϕ is differentiable.

To state the proposition, it is helpful to define the standard notion of a predictive distribution as in Bayesian statistics (i.e., the μ_{s^t}-average of the π_θ's). Given μ_{s^t}, define the *predictive distribution* $P_{s^t} : \Sigma \to [0,1]$ by:

$$P_{s^t}(B) = \int_\Theta \pi_\theta\left(B \mid s^t\right) d\mu_{s^t}(\theta), \qquad \forall B \in \Sigma.$$

In particular, using μ to denote μ_{s^0}, $P_{s^0}(B) = \int_\Theta \pi_\theta(B)\, d\mu(\theta)$. Observe that $P_{s^t}(B) = P_{s^t}(B \cap s^t)$. We denote by $P_{s^t}(x_{t+1}, ..., x_{t+k})$ the probability of the cylinder set determined by $(s^t, x_{t+1}, ..., x_{t+k})$ (i.e., $P_{s^t}(\{s^t, x_{t+1}, ..., x_{t+k}\})$). In particular, the *one-step-ahead predictive distribution* at s^t is the marginal on $\mathcal{X}_{t+1}$ of P_{s^t}, and the probability it assigns to $X_{t+1} = x_{t+1}$ is denoted $P_{s^t}(x_{t+1})$.

Proposition 2 *Given, for each s^t, a V_{s^t} as in (3), there exists a unique countably additive probability $\eta : \Sigma \to [0,1]$ such that,*

(i) if ϕ is affine, then

$$V_{s^t}(f) = \int_S \left[\sum_{\tau=t}^{\infty} \beta^{\tau \square t} u\left(f\left(s^\tau\right)\right)\right] d\eta_{s^t} \tag{4}$$

where the measure η is related to the π_θ's and the μ_{s^t}'s (via the predictive distributions) through the formula

$$\eta\left(\{x_1, \ldots, x_m\}\right) = P_{s^0}(x_1)\, P_{(x_1)}(x_2) \cdots P_{(x_1, \ldots, x_{m \square 1})}(x_m), \tag{5}$$

for each $m \geq 1$ and $x_1, \ldots, x_m$, and η_{s^t} is the Bayesian update of η.[8]

(ii) if ϕ is differentiable, then

$$\left.\frac{\partial V_{s^t}(f) / \partial\left[\beta^n u(f((s^t, x_{t+1}, \ldots, x_{t+n})))\right]}{\partial V_{s^t}(f) / \partial\left[u(f(s^t))\right]}\right|_{f=d} = \eta_{s^t}\left(\left\{s^t, x_{t+1}, \ldots, x_{t+n}\right\}\right)$$

for each $n \geq 1, t, s^t, x_{t+1}, \ldots, x_{t+n}$ and deterministic plan d.

Part (i) of the above result verifies that when ϕ is affine we get discounted expected utility with Bayesian updating of a unique measure on paths, η. This measure is uniquely determined by the product of one-step-ahead predictive distributions. Equation (5) implies that the one-step-ahead predictive distribution at any node is exactly the one-step-ahead marginal of the Bayesian update of η.

Part (ii) of the proposition shows that the marginal rate of substitution

$$\left.\frac{\partial V_{s^t}(f) / \partial\left[\beta^n u(f((s^t, x_{t+1}, \ldots, x_{t+n})))\right]}{\partial V_{s^t}(f) / \partial\left[u(f(s^t))\right]}\right|_{f=d}$$

at s^t between utility at s^t and the present discounted utility at $(s^t, x_{t+1}, \ldots, x_{t+n})$, when evaluated at any deterministic plan d, is the Bayesian conditional probability $\eta_{s^t}(\{s^t, x_{t+1}, \ldots, x_{t+n}\})$

[8] That is, $\eta_{s^t}(B) \equiv \eta(B \cap s^t) / \eta(s^t)$ for all $s^t \in \mathcal{S}$ and $B \in \Sigma$.

of $\{s^t, x_{t+1}, \ldots, x_{t+n}\}$ given s^t according to η. This shows that η can be meaningfully characterized solely in terms of behavior toward *plans* even when preferences are ambiguity sensitive, as long as ϕ is differentiable (i.e., preferences are smooth). Under expected utility (with monetary payoffs and differentiable utility) subjective probability measures the trade-offs at the margin, evaluated at a riskless position, between sure payoffs and payoffs contingent on a specific event. As the proposition shows, analogously, η measures the trade-offs at the margin between utility at s^t and (discounted) utility at successor nodes, when these trade-offs are evaluated at a deterministic plan.

In addition to η, we have the predictive distributions, P_{s^t}, which are determined from preferences over second order acts. It is natural to ask whether the η_{s^t} agree with the P_{s^t} everywhere, and not simply for the one-step-ahead marginals. As we will show in Section 5, this question is equivalent to the question of whether the μ_{s^t} are the Bayesian updates of μ.

4.1 Discussion of Our Approach

To give some additional insight into the derivation of the model, we present an alternative approach using a different final consistency assumption and show why it runs into difficulty. One might have thought that Assumption 8 could be formulated as follows:

First, some modified definitions. Given any continuation plan $f_{s^t} \in \mathcal{F}_{s^t}$, consider for each $s = (s^t, x_{t+1}, x_{t+2}, \ldots, x_{t+n}, \ldots) \in S \cap \{x_1, ..., x_t\}$ the deterministic continuation plan

$$d_{f_{s^t}}(s) = \left(f_{s^t}\left(s^t\right), f_{s^t}(s^t, x_{t+1}), f_{s^t}(s^t, x_{t+1}, x_{t+2}), ..., f_{s^t}(s^t, x_{t+1}, x_{t+2}, \ldots, x_{t+n}), ...\right).$$

Definition 6 *Given any continuation plan $f_{s^t} \in \mathcal{F}_{s^t}$, the **certainty equivalent** of f_{s^t} given θ, $e_{f_{s^t}}(\theta)$, is a payoff e such that*

$$\left(f_{s^t}\left(s^t\right), e, e, ..., e, ...\right) \sim^*_{s^t} \left\{d_{f_{s^t}}(s), \pi_\theta\left(s \mid s^t\right)\right\}_{s \in S \cap \{x_1, ..., x_t\}}.$$

In other words, the certainty equivalent of the continuation plan f_{s^t} is the payoff making the DM indifferent between the continuation plan having that payoff from time $t+1$ onwards and the s^t-mixed plan which pays off according to $d_{f_{s^t}}(s)$ with probability $\pi_\theta(s \mid s^t)$. Observe that such a mixed plan is an element of $\mathcal{P}_{s^t}$ and so via Assumption 5 it is in the domain of the induced preference $\succcurlyeq^*_{s^t}$.

Definition 7 *Given $f_{s^t} \in \mathcal{F}_{s^t}$, $\hat{f}^2 \in \mathfrak{F}$ denotes a **(hatted) second order act associated with** f_{s^t}, defined as follows*

$$\hat{f}^2(\theta) = e_{f_{s^t}}(\theta) \qquad \text{for all } \theta \in \Theta.$$

Assumption 9 (Consistency with Hatted Second Order Acts) *Given $f_{s^t}, g_{s^t} \in \mathcal{F}_{s^t}$ and $\hat{f}^2, \hat{g}^2 \in \mathfrak{F}$, if $f_{s^t}(s^t) = g_{s^t}(s^t)$,*

$$f \succcurlyeq_{s^t} g \Longleftrightarrow \hat{f}^2 \succcurlyeq^2_{s^t} \hat{g}^2$$

for some $f, g \in \mathcal{F}$ with $f_{|s^t} = f_{s^t}$ and $g_{|s^t} = g_{s^t}$.

Notice the difference between this approach and the one we have taken. In our approach, the consistency assumption applied only to one-step-ahead continuation plans. In this approach it applies to all continuation plans. One can show, along the lines of the proof of Theorem 1, that substituting this new assumption for Assumption 8 and dropping dynamic consistency (Assumption 4) implies that there exists the following "reduced-form" representation for $\succeq_{s^t}$ on plans:

$$\hat{V}_{s^t}(f) = u\left(f\left(s^t\right)\right) + \beta\phi^{-1}\left[\int_{\Theta}\phi\left(\sum_{s\in S\cap\{s^t\}}\pi_\theta\left(s \mid s^t\right)\sum_{\tau\geq t+1}\beta^{\tau-(t+1)}u(f(s)(\tau))\right)d\mu_{s^t}\right] \quad (6)$$

where $\pi_\theta(s \mid s^t)$ is the probability of the path s given history s^t and given that the true parameter is θ and $f(s)(\tau)$ is the time τ payoff according to plan f along path s. A comparison of equation (6) with equation (3) reveals that, though they agree in evaluating one-step-ahead continuation plans, in general they are quite distinct. In particular, (6) will generally violate dynamic consistency, while (3) will generally violate the modified consistency assumption (Assumption 9).

Generally, one can think of two ways of relating preferences in a dynamic model to those in an atemporal setting. The first is by viewing the atemporal model as the special case of the dynamic model with one-period of uncertainty. This is the nature of our recursive extension of the smooth ambiguity model – the atemporal model corresponds to preferences over one-step-ahead continuation plans sharing the same current payoff. The second is by viewing the atemporal model as a reduced-form of the dynamic model, abstracting away from the dynamic structure yet representing the same preferences. The alternative dynamic representation in equation (6) bears this type of relation to the atemporal model for continuation plans sharing the same current payoff.

It would certainly be an elegant result to have a dynamic model that related to the atemporal model in both ways. We do not generally have this. In this regard, we suggest that the modified consistency assumption used in deriving the reduced form may not be compelling, in that it requires the DM to behave as if all uncertainty were resolving immediately when, in fact, this would be true only for one-step-ahead continuation plans. Why does this matter in our set-up? It matters because our DM is both ambiguity sensitive and respects consequentialism (as noted in Section 3.1, the latter property follows from Monotonicity (Assumption 2)). Ambiguity sensitivity will, in general, result in non-separabilities in the evaluation of payoffs across mutually exclusive future events. Consequentialism requires that, once a given node has been reached, payoffs at unrealized events cannot influence preferences. When all uncertainty resolves immediately, there is no tension between these two. When it resolves gradually, however, there will generally be a strong tension. Consider, for example, a DM who obeys the reduced-form consistency assumption evaluating, at s^t, a trade-off between utility at node (s^t, x_{t+1}, x_{t+2}) and utility at node (s^t, x_{t+1}, x'_{t+2}). Because of ambiguity and the DM's sensitivity to it, this trade-off may well depend on the utility at a third, mutually exclusive, node, say, (s^t, x''_{t+1}). However, if this DM reaches node (s^t, x_{t+1}), then consequentialism demands that the outcome at (s^t, x''_{t+1}) not matter in evaluating the same trade-off. This is why there is a conflict between the reduced-form consistency and dynamic consistency. It seems reasonable that a DM who is aware of the full implications of the dynamic setting will anticipate the decision opportunity at node (s^t, x_{t+1}), and want

to evaluate the consumption possibilities (implied by f_{s^t}) that obtain beyond (s^t, x_{t+1}) using the view of ambiguity that will become available at (s^t, x_{t+1}) rather than the view of ambiguity expressed through μ_{s^t}. Thus it is not obvious that the DM think the evaluation of f_{s^t} is on par with evaluating an associated second order act for which all uncertainty, by definition, resolves immediately upon taking the decision. But this is exactly what the reduced-form approach described in this section assumes. In other words, a difficulty with the reduced-form approach is that ambiguity present at s^t, and reflected in μ_{s^t}, would figure in the (current) evaluation of the associated second order act, whereas this ambiguity is not relevant to evaluating the part of the consumption stream that comes beyond (s^t, x_{t+1}) if that node were to be reached.

Although the recursive and reduced-form approaches conflict in general, there are two important special cases in which they can be reconciled. First, if ϕ is affine, so that preferences are ambiguity neutral, there is no conflict between the two. Second, if ϕ is differentiable and we restrict attention to local behavior at deterministic plans, again the two approaches are compatible. When exploring updating in Section 5.1, these domains of agreement prove quite useful. In both these domains, the tension discussed above is absent because the DM is ambiguity neutral either globally (ϕ affine) or locally around deterministic plans (ϕ differentiable). Therefore the marginal trade-off between two nodes is independent from the outcome at any mutually exclusive third node.

From the point of view of dynamic applications, it is worth noting that even if both a reduced form and a recursive representation were available, it is the recursive form that will be of far more use because of its tractability. The only potential disadvantage of a recursive relative to a reduced form representation, is that existence and uniqueness of the representation becomes a more subtle issue. This is the subject of the next section.

4.2 Existence and Uniqueness

If our preferences obeyed the "reduced-form" representation discussed above, they would be explicitly determined once the elements u, β, ϕ, π and μ are specified. However, the preference functional, V_{s^t}, that by Theorem 1 represents the preference $\succcurlyeq_{s^t}$ on plans, is only implicitly defined by the recursive equation (3). Therefore, an important issue is whether, for a given specification of the elements u, β, ϕ, π and μ, such recursive equation admits a unique monotonic solution. Otherwise, Theorem 1 would be of little use; without uniqueness, it is difficult to understand what it is beyond u, β, ϕ, π and μ that determines preference. Approachability of this solution through iterative methods is also of practical interest.

Our first result in this section shows that there always exists a solution to the recursive equation (3) in the representation. We then provide two sets of sufficient conditions for uniqueness and monotonicity of the solution. Proposition 3 shows that any solution is unique and monotonic when restricted to plans that eventually become deterministic. Theorem 3 shows that uniqueness and monotonicity for all plans holds under conditions on ϕ weak enough to encompass many cases of interest. Under the same conditions, iterative methods for finding V_{s^t} are guaranteed to converge.

The general existence result is the following:

Theorem 2 *For each f there exists a $V_{s^t}(f)$ satisfying the recursive equation (3) in Theorem 1.*

If the time horizon were finite, a simple backward induction argument would be enough to solve equation (3), and the solution would be unique and monotonic. In our infinite horizon set-up this is no longer possible, but we can still prove a very general uniqueness and monotonicity result for eventually deterministic plans, that is, plans that become deterministic after a finite time.

Proposition 3 *When restricted to eventually deterministic plans, the recursive equation (3) has a unique and monotonic solution.*

We now move to explore the case of more general plans. For our uniqueness results, we will need to refer to several classes of functions ϕ: functions with a non-decreasing coefficient of absolute ambiguity aversion, functions with a non-increasing coefficient of relative ambiguity aversion and functions with a strong form of a decreasing coefficient of absolute ambiguity aversion.[9] We will denote these classes as IAAA, IRAA and SDAAA, respectively.

Formally, we will say that a continuous and strictly increasing function $\phi : \mathcal{U} \to \mathbb{R}$ is IAAA if it is twice continuously differentiable and is such that the function

$$\lambda(x) = \square \frac{\phi''(x)}{\phi'(x)}$$

is non-decreasing, while it is IRAA if the function

$$(x) = \square \frac{x\phi''(x)}{\phi'(x)} = x\lambda(x)$$

is non-decreasing and is SDAAA if the function

$$\varsigma(x) = \square \frac{\phi''(x)}{[\phi'(x)]^2} = \frac{\lambda(x)}{\phi'(x)}$$

is non-increasing. Of note for applications, the union of these classes includes, among others, the following classic cases:

1. The class of constant absolute ambiguity aversion (CAAA) functions:

$$\phi(x) = \begin{cases} ax + b & \text{if } \lambda(x) = 0 \text{ for all } x \in \mathcal{U} \\ \square a e^{\square \lambda x} + b & \text{if } \lambda(x) = \lambda > 0 \text{ for all } x \in \mathcal{U} \end{cases}$$

   where $a > 0$ and $b \in \mathbb{R}$. KMM showed that these functions may be thought of as displaying constant ambiguity attitude.

[9] This terminology follows that used in KMM. In particular, KMM relates the Arrow-Pratt coefficient of ϕ to ambiguity attitude.

2. The class of constant relative ambiguity aversion (CRAA) functions:

$$\phi(x) = \begin{cases} a\frac{x^{1-}}{1-} + b & \text{if} \quad (x) = \quad \neq 1 \text{ for all } x \in \mathcal{U} \\ a \ln x + b & \text{if} \quad (x) = 1 \text{ for all } x \in \mathcal{U} \end{cases}$$

where $a > 0$ and $b \in \mathbb{R}$. By analogy, one might view these functions as displaying constant relative ambiguity attitude.

3. Quadratic functions that are increasing on the relevant domain.

We can now state a result showing uniqueness and monotonicity if ϕ is either IAAA, IRAA or a subset of SDAAA.

Theorem 3 *Assume $\mathcal{U} \subseteq \mathbb{R}_+$ and ϕ is twice differentiable on $\mathbb{R}_{++}$. There are unique and monotonic V_{s^t} satisfying the recursive equation (3) in Theorem 1 if at least one of the following holds:*

(i) ϕ is IAAA; or

(ii) $0 \notin \mathcal{U}$ and ϕ is IRAA; or

(iii) ϕ is SDAAA and concave and $\sup_{x \geq 0} \quad (x) < 1$.

It is worth noting that under conditions (i), (ii) or (iii) the unique solution can be found via contraction arguments and this provides uniform convergence of iterative methods of finding a solution. See Appendix B for a formal statement. We emphasize that the above conditions are sufficiency conditions only and we have not been able to construct a counterexample failing uniqueness or monotonicity. We also note that the difficulty in finding complete conditions for uniqueness is not unique to our model. Other, far older and quite popular non-linear recursive models, for example those of Epstein and Zin (1989), are in a similar situation (see Marinacci and Montrucchio (2007)).

5 Bayesian Updating and Learning

5.1 Bayesian Updating of μ

At the end of Section 4 we raised the issue of whether the beliefs η_{s^t} derived in Proposition 2 agree with the predictive distributions P_{s^t}. Recall that Proposition 2 guaranteed that their one-step ahead marginals agree. We remarked that the question of overall agreement was equivalent to the question of whether the μ_{s^t} were the Bayesian updates of μ. In this section, we make precise the conditions under which this is true and discuss the significance of this result.

Consider the DM's marginal trade-off between utility at s^t and (discounted) utility at a successor node $(s^t, x_{t+1}, \ldots, x_{t+n})$, when evaluated at a *deterministic* plan. One could imagine, because this trade-off is being made given only information available at s^t, the trade-off could be recoverable from preferences over second order acts at s^t alone, in particular from

μ_{s^t}. If the DM were expected utility over plans with beliefs given by the μ_{s^t}-average of the π_θ, the trade-off would indeed be identified by the "n-step-ahead" predictive distribution $P_{s^t}(x_{t+1}, \ldots, x_{t+n})$. Clearly, assuming such global ambiguity neutrality is very strong, and would defeat the whole purpose of our modeling exercise. This is where the limitation to evaluating trade-offs at deterministic plans comes in. All that is needed for $P_{s^t}(x_{t+1}, \ldots, x_{t+n})$ to identify these trade-offs is that the DM is ambiguity neutral "locally around determinism" with beliefs given there by the μ_{s^t}-average of the π_θ. If our model is truly smooth (i.e., ϕ differentiable), such local ambiguity neutrality is perfectly compatible with overall sensitivity to ambiguity, just as local risk-neutrality is compatible with global risk aversion or love in the standard expected utility model.

This motivates the following assumption which says exactly that the marginal trade-off between utility at s^t and (discounted) utility at a successor node $(s^t, x_{t+1}, \ldots, x_{t+n})$, when evaluated at a deterministic plan is given by $P_{s^t}(x_{t+1}, \ldots, x_{t+n})$, which, recall, is completely determined by $\succcurlyeq^2_{s^t}$ through μ_{s^t}.

Assumption 10 (MRS) *For each deterministic plan d,*

$$\left.\frac{\partial V_{s^t}(f)/\partial\left[\beta^n u(f((s^t, x_{t+1}, \ldots, x_{t+n})))\right]}{\partial V_{s^t}(f)/\partial\left[u(f(s^t))\right]}\right|_{f=d} = P_{s^t}(x_{t+1}, \ldots, x_{t+n}) \tag{7}$$

for each $n \geq 1, t, s^t, x_{t+1}, \ldots, x_{t+n}$.

We will show that with ϕ differentiable, if μ is updated by Bayes' rule then our model implies Assumption MRS, and furthermore, under mild conditions, Assumption MRS implies μ must be updated by Bayes' rule in our model.

It is informative to note that Assumption MRS is implied by the reduced-form dynamic smooth ambiguity model with discounted utility (as in (6)) when ϕ is differentiable. Why is this true? Any such model is locally ambiguity neutral (i.e., locally expected utility) around any deterministic plan. Under ambiguity neutrality, μ_{s^t}-average probabilities precisely reveal utility trade-offs at the margin. In this sense, Assumption MRS is a very limited version of the closure sometimes assumed for recursive models – here we are not demanding that each $\succcurlyeq_{s^t}$ on plans is represented by a reduced-form dynamic smooth ambiguity model, but simply that it shares the relationship between the predictive distributions and marginal rates of substitution around determinacy with that model.

Observe that closure is (in combination with recursion) what delivers anything that dynamic recursive models have to say about updating – for expected utility it delivers Bayes' rule applied to the overall prior, for recursive multiple priors (Epstein and Schneider (2003b)) it delivers prior–by-prior Bayesian updating applied to the overall rectangular set of priors, for recursive variational preferences (Maccheroni, Marinacci, and Rustichini (2006b)) it delivers a condition on how the ambiguity index must be updated. Our next result shows that adding limited closure in the form of Assumption MRS to our earlier assumptions is equivalent to Bayesian updating of the predictive distributions.

Theorem 4 *Assume each $\succcurlyeq_{s^t}$ satisfies Assumptions 1-8 and that ϕ is differentiable. Then, Assumption MRS is equivalent to the predictive distributions being related by Bayes' rule (i.e., $P_{s^t}(B) = P_{s^0}(B \cap s^t)/P_{s^0}(s^t)$ for each $t, s^t, B \in \Sigma$.)*

To sketch why this holds, recall that under our model, as was shown in Proposition 2, the expression on the left-hand-side of Assumption MRS is given by the product of one-step-ahead predictive distributions. In this way, this trade-off among plans is decomposed into the product of a series of one-step-ahead tradeoffs, each of which is determined (through the $\mu_{s^t}, \mu_{(s^t,x_{t+1})}, \ldots, \mu_{(s^t,x_{t+1},\ldots,x_{t+n})}$ respectively) by preferences over second order acts at different nodes. This stringing together of one-step-ahead trade-offs reflects the dynamic consistency imposed on the model. Assumption MRS says that the same expression is given by $P_{s^t}(x_{t+1},\ldots,x_{t+n})$, the n-step-ahead predictive distribution at s^t, which is determined (through μ_{s^t}) by preferences over second order acts at s^t alone. The consistency between this sequence of one-step-ahead distributions and the single n-step-ahead distribution can be achieved only when the predictive distributions are related by Bayes' rule.

Bayesian updating of the predictive distributions is easily seen to be implied by the μ_{s^t} being derived from μ by Bayes' rule, however the converse is not always true. We now provide a condition under which the two are equivalent. This condition requires a certain diversity among the distributions π_θ. We subsequently show that this condition is commonly satisfied.

Definition 8 *The* full rank condition *holds if, for each node s^t, there exist $m \geq 1$ elementary cylinder sets $B_1^{s^t},\ldots,B_{m\geq 1}^{s^t}$ such that the $m \times m$ matrix*

$$A^{s^t} \equiv \begin{bmatrix} \pi_{\theta_1}\left(B_1^{s^t} \mid s^t\right) & \cdots & \pi_{\theta m}\left(B_1^{s^t} \mid s^t\right) \\ \vdots & \vdots & \vdots \\ \pi_{\theta_1}\left(B_{m\geq 1}^{s^t} \mid s^t\right) & \cdots & \pi_{\theta m}\left(B_{m\geq 1}^{s^t} \mid s^t\right) \end{bmatrix}$$

is of full rank, where $\Theta = \{\theta_1,\ldots,\theta_m\}$.

Corollary 1 *Assume each $\succcurlyeq_{s^t}$ satisfies Assumptions 1-8 and that ϕ is differentiable. If the full rank condition holds, then the MRS condition (7) is equivalent to Bayesian updating of the μ_{s^t} (i.e., $\mu_{s^t}(\theta) = \mu(\theta)\,\pi_\theta(s^t) \,/\, \int_\Theta \pi_\theta(s^t)\,d\mu$ for all t, s^t, θ).*

Observe that for each s^t, since the event tree and observations continue forever, there is an infinite number of ways to select $m \geq 1$ elementary cylinder sets. The full rank condition requires simply that one such selection yields a non-singular matrix. As singularity of a matrix is a non-generic property,[10] one should expect that the full rank condition would generally be satisfied. In any case, our next result shows that when the π_θ correspond to homogenous Markov processes, as is likely to be assumed in many applications, the full rank condition is naturally and easily satisfied.

Proposition 4 *Assume $\mathcal{X}_t = \mathcal{X}$ for all t and suppose that each π_θ makes the sequence $\{X_t\}_{t\in\mathcal{T}}$ a homogeneous Markov chain with transition function $q_\theta : \mathcal{X}\times\mathcal{A} \to [0,1]$. Then, the full rank condition holds provided there is an initial outcome $x^0 \in \mathcal{X}$ so that $q_\theta(s^0)(\cdot) = q_\theta(x^0,\cdot)$ for all θ and there is no redundancy in Θ with respect to the transition functions on the diagonal.*[11]

[10] That is, the set of $m \times m$ singular matrices has zero Lebesgue measure in $\mathbb{R}^{m\times m}$.

[11] That is, if $\theta' \neq \theta''$, then there exist $x \in \mathcal{X}$ such that $q_{\theta'}(x,x) \neq q_{\theta''}(x,x)$.

It is worth noting that in the classic i.i.d. case, a special case of homogeneous Markov, the no redundancy requirement simply says that each θ corresponds to a different i.i.d. process and the initial outcome requirement is irrelevant.

Assumption MRS is the key to justifying Bayesian updating of beliefs in our model. Under mild conditions, we have shown the two are equivalent. Any rationale for Assumption MRS is thus a rationale (in conjunction with the other assumptions underlying our model, especially dynamic consistency) for applying Bayes' rule to μ.

To conclude this section, we summarize the argument for Assumption MRS. Consider a DM whose preferences $\succcurlyeq_{s^t}$ and $\succcurlyeq^2_{s^t}$ satisfy Assumptions 1-8 and whose ϕ is revealed to be affine. This DM is ambiguity neutral and evaluates plans according to discounted expected utility. How does such a DM trade-off, at a deterministic plan given information at node s^t, marginal utility obtainable at s^t with (discounted) marginal utility obtainable at a future node further down the event tree? This trade-off is determined by nothing but the DM's belief (conditional on information at s^t) about the chances of reaching the future node. Given expected utility preferences on plans, it is natural that this belief is the μ_{s^t}-average of the $\pi_\theta(\cdot \mid s^t)$ probabilities of reaching the future node, where μ_{s^t} is inferred from the DM's $\succcurlyeq^2_{s^t}$ preferences. Next we recall from (ii) of Proposition 2, that for a DM with any differentiable ϕ, this trade-off, given that it is measured at determinacy, is the same as that obtained under ambiguity neutrality, so again it <would>SHOULD be described by the μ_{s^t}-average of the $\pi_\theta(\cdot \mid s^t)$. This is precisely Assumption MRS. So why would it make sense to update μ by Bayes' rule? First, because the predictive distributions would then also be updated by Bayes' rule, hence ensuring that Assumption MRS held. Second, under the easily satisfied full rank condition, this is the *only* updating rule that would be consistent with both Assumption MRS and our dynamically consistent model.

5.2 Learning About θ

Bayesian updating of μ generates a theory of learning under ambiguity, enabling us to understand the conditions under which the effect of ambiguity on preferences would (or would not) eventually fade away with progressive accumulation of observations. We explore this issue by giving conditions under which ambiguity may be "learned away" followed by an example of a growth model where those conditions do not hold and ambiguity never completely fades away, despite learning.

Proposition 5 *Assume Bayesian updating of μ. Suppose $\theta_0 \in \operatorname{supp}(\mu)$ is the true parameter and suppose that each π_θ makes the process $\{X_t\}_{t\in\mathcal{T}}$ stationary and ergodic.*[12] *If $\phi^{\square 1}$ is Lipschitz, then*

$$\lim_{t\to\infty}\left|V_{s^t}(f) \square u\left(f\left(s^t\right)\right) \square \beta\int_{\mathcal{X}_{t+1}} V_{(s^t,x_{t+1})}(f)\, d\pi_{\theta_0}\left(x_{t+1}; s^t\right)\right| = 0 \qquad \pi_{\theta_0}\text{-}a.e. \tag{8}$$

for all $f \in \mathcal{F}$, and the convergence is uniform over $\mathcal{F}$.

[12] The terms "stationary" and "ergodic" here follow their standard usage in the probability and statistics literature; see e.g. Durrett (1991), p. 291.

In words, as observations build up, the approximate recursive representation

$$V_{s^t}(f) \approx u\left(f\left(s^t\right)\right) + \beta \int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_{\theta_0}\left(x_{t+1}; s^t\right) \tag{9}$$

of the function $V_{s^t}(f)$ is more and more accurate, provided the process is ergodic and stationary. This is a large class of processes which includes, for instance, i.i.d., exchangeable and ARCH(1) models (see Lijoi, Prunster, and Walker (2007)). In particular, in our setting all stationary Markov processes are in this class.

In Section 2.2 we observed that a plan f can be viewed as a collection $\{f_t\}_{t \in \mathcal{T} \cup \{0\}}$, where each $f_t : S \to \mathcal{C}$ is a $\sigma(X_1, ..., X_t)$-measurable function. From this standpoint, (9) means that the value function $V_{s^t}(f)$ looks more and more similar to

$$\int_S \sum_{\tau \geq t} \beta^{\tau - t} u\left(f_t(s)\right) d\pi_{\theta_0}\left(s \mid s^t\right).$$

Summing up, the DM is more and more choosing as if he knew the true parameter.

It is important to note that this result depends in a strong way on our assumption that the parameter space Θ is finite. In more general environments, such convergence of preference to preference under knowledge of the true parameter need not occur. A simple example of this is given below.

Example 3 Consider a Markov switching model for the growth process for an economy (similar to e.g., Hamilton (1989)). Let G_t denote the GDP of the economy at time t. The growth of the economy is determined by the equation $G_t = G_{t-1} g_t$ where g_t is the gross growth rate of GDP between $t-1$ and t. This growth rate is stochastic and has a distribution determined by a Markov binary state variable z_t. $z_t = 1$ corresponds to a relatively favorable growth rate distribution (a "boom") and $z_t = 0$ corresponds to a relatively unfavorable growth rate distribution (a "bust").

Given the value of z_t, the growth rate of the economy is determined by

$$g_t \mid z_t \sim LN(\mu_{z_t}, \sigma^2_{z_t}) \text{ if } z_t = 0 \tag{10}$$

where $LN(\mu_{z_t}, \sigma^2_{z_t})$ denotes a lognormal distribution with parameters $(\mu_{z_t}, \sigma^2_{z_t})$.

Finally, the state process follows a Markov chain with initial state z_1 and transition matrix

$$\mathrm{P} = \begin{bmatrix} \rho_0 & 1 - \rho_0 \\ 1 - \rho_1 & \rho_1 \end{bmatrix}$$

where $\rho_0 = Pr(z_{t+1} = 0 | z_t = 0)$ and $\rho_1 = Pr(z_{t+1} = 1 | z_t = 1)$. We assume, for simplicity, that $P, \mu_0, \sigma_0^2, \mu_1$, and σ_1^2 are known to the DM.

In the language of our model, let a parameter $\theta = \{z_t\}_{t=1}^{\infty} \in \Theta = \{0, 1\}^{\infty}$. Each period what is observed is GDP, G_t. Thus, the observation space is $X_t = R_+$ for each time t. Note that both the parameter space and the observation space in this example are infinite, in contrast to our assumption above that both of these spaces are finite. The infinite parameter

space is crucial to the learning behavior in the example, but the observation space could just as well be finite (with the growth rate process appropriately discretized) and the message of the example would not change. Given θ, $\pi_\theta(x_{t+1} \mid s^t) = \pi_\theta(G_{t+1} \mid G_t, G_{t\square 1}, \ldots, G_0)$. Given our assumptions, this can be computed as follows: $\pi_\theta(G_{t+1} \mid G_t, G_{t\square 1}, \ldots, G_0) = \pi_\theta(g_{t+1} G_t \mid G_t) = l\left(\frac{G_{t+1}}{G_t}; \mu_{\theta_{t+1}}, \sigma^2_{\theta_{t+1}}\right)$ where l is the lognormal density and $\theta_{t+1} = z_{t+1}$ is the $(t+1)^{st}$ ordinate of θ.

Suppose that the DM perceives ambiguity about the process generating the growth rate. This must be reflected in the initial beliefs, μ_{s^0}, over Θ. On Θ, consider the product σ-algebra $\Lambda = \quad_{t\in\mathcal{T}} 2^{\{0,1\}}$ generated by all one-dimensional cylinder sets $B_t \times \prod_{t\neq s\in\mathcal{T}} \{0,1\}$, where $B_t \in 2^{\{0,1\}}$. Suppose μ_{s^0} on Λ is as follows: for all $t \in T = \{1, 2, \ldots\}$,

$$\mu_{s^0}\quad \{0\} \times \prod_{r\in\mathcal{T}, r\neq t} \{0,1\}\Bigg), \mu_{s^0}\quad \{1\} \times \prod_{r\in\mathcal{T}, r\neq t} \{0,1\}\Bigg)\Bigg) = (p, 1\,\square\, p)\, P^{t\square 1}$$

where $p \in [0,1]$ is a belief that the initial $z_1 = 0$ and P is the transition matrix given earlier.

Because $\pi_\theta(x_{t+1} \mid s^t) = \pi_\theta(x_{t+1} \mid G_t, G_{t\square 1}, \ldots, G_0) = l\left(\frac{G_{t+1}}{G_t}; \mu_{\theta_{t+1}}, \sigma^2_{\theta_{t+1}}\right)$ depends on θ only through θ_{t+1}, the relevant conditionals for μ will be

$$\begin{aligned}
&\mu_{s^t}\quad \{0\} \times \prod_{r\in\mathcal{T}, r\neq t+1} \{0,1\}\Bigg), \mu_{s^t}\quad \{1\} \times \prod_{r\in\mathcal{T}, r\neq t+1} \{0,1\}\Bigg)\Bigg) \\
=\ &\mu_{s^0}\quad \{0\} \times \prod_{r\in\mathcal{T}, r\neq t+1} \{0,1\} \mid G_t, G_{t\square 1}, \ldots, G_0\Bigg), \mu_{s^0}\quad \{1\} \times \prod_{r\in\mathcal{T}, r\neq t+1} \{0,1\} \mid G_t, G_{t\square 1}, \ldots, G_0\Bigg)\Bigg) \\
\equiv\ &\ \square\ p_{t+1}\ \square\ s^t\big), 1\,\square\, p_{t+1}\ \square\ s^t\big)\big)
\end{aligned}$$

The expression for the above may be found recursively by applying Bayes rule. Specifically,

$$\begin{aligned}
p_1\ \square\ s^0\big) &= p, \\
p_1\ \square\ s^1\big) &= \frac{p l\left(\frac{G_1}{G_0}; \mu_0, \sigma_0^2\right)}{p l\left(\frac{G_1}{G_0}; \mu_0, \sigma_0^2\right) + (1\,\square\, p) l\left(\frac{G_1}{G_0}; \mu_1, \sigma_1^2\right)}
\end{aligned}$$

and, for $t = 1, 2, \ldots,$

$$\begin{aligned}
p_{t+1}\ \square\ s^t\big) &= p_t\ \square\ s^t\big)\,\rho_0 + \ \square\ 1\,\square\, p_t\ \square\ s^t\big)\big)\,(1\,\square\,\rho_1), \\
p_{t+1}\ \square\ s^{t+1}\big) &= \frac{p_{t+1}(s^t)\, l\left(\frac{G_{t+1}}{G_t}; \mu_0, \sigma_0^2\right)}{p_{t+1}(s^t)\, l\left(\frac{G_{t+1}}{G_t}; \mu_0, \sigma_0^2\right) + (1\,\square\, p_{t+1}(s^t)) l\left(\frac{G_{t+1}}{G_t}; \mu_1, \sigma_1^2\right)}.
\end{aligned}$$

So,

$$\ \square\ p_{t+1}\ \square\ s^t\big), 1\,\square\, p_{t+1}\ \square\ s^t\big)\big) = \ \square\ p_t\ \square\ s^t\big), 1\,\square\, p_t\ \square\ s^t\big)\big)\, P.$$

The important feature is that, as long as ρ_0 and ρ_1 are strictly between 0 and 1, beliefs about the relevant part of θ never approach degeneracy. This implies that ambiguity remains relevant to preferences even in the limit, in contrast to the result in Proposition 5. What drives the example is the DM views the environment as one in which some new and relevant ambiguity arrives each period. From this perspective, the important feature of the example is that the DM perceives ambiguity about g_{t+1} no matter how many periods have been observed. One special case of our example occurs when $\rho_0 = 1 \square \rho_1$. In this case past observations are completely uninformative about state θ_{t+1} and thus g_{t+1} and so no reduction in relevant ambiguity occurs through learning.
Why isn't an example like this one possible when the parameter space Θ is finite? The intuition is that either knowledge of θ becomes eventually irrelevant to payoffs or, to the extent that it remains relevant, the DM will observe more and more data, and thus will eventually be able to learn the relevant aspects of θ.

The above example showing how ambiguity may persist in our model when the parameter space is rich complements findings on the persistence of ambiguity in earlier models (see e.g., Epstein and Schneider (2003a), Maccheroni and Marinacci (2005)).

6 Examples

We present an example adapted from one in the section on dynamic ambiguity in the excellent survey paper by Backus, Routledge, and Zin (2004) along with an example looking at the implications for asset prices.

6.1 Event Tree Example

This example serves to explicitly illustrate the mechanics of calculating with the model. Moreover, the results of these calculations illustrate an important relationship between our model and the recursive multiple priors model of Epstein and Schneider (2003b). As was the case (see Klibanoff, Marinacci, and Mukerji (2005)) with the timeless versions of the smooth ambiguity model and the multiple priors model, the recursive multiple priors form may be viewed as a limiting case of recursive smooth ambiguity as ambiguity aversion is taken to infinity.

The example (drawn from Backus, Routledge, and Zin (2004), p. 38, an elaboration on an example from Seidenfeld and Wasserman (1993)) is based on the information structure depicted in Figure 1 from Section 2.1.

There is an asset, f, that yields the following payoffs, at time $t = 2$, as a function of the time 2 node realized in the tree:

$$\begin{array}{rcl} f(1,1) & = & 1+\varepsilon \\ f(1,2) & = & \varepsilon \\ f(2,1) & = & 1 \\ f(2,2) & = & 0 \end{array}$$

, where ε is a non-negative constant. Assume that f yields 0 at all other time periods. The realizations of the random variables that determine the paths in the tree are governed by a process with two parameters, δ and ρ. The parameter space is

$\Theta = \{(\delta, \rho) : \delta \in \{\square\bar{\delta}, \bar{\delta}\}, \rho \in \{\square\bar{\rho}, \bar{\rho}\}\}$ where $\bar{\delta} \in [0, 1)$ and $\bar{\rho} \in [0, 1]$ are constants. As a function of the parameters, the path probabilities are given by:

$$\begin{aligned}
\pi_{(\delta,\rho)}(1,1) &= \frac{(1+\delta)(1+\rho)}{4}, \\
\pi_{(\delta,\rho)}(1,2) &= \frac{(1+\delta)(1 \square \rho)}{4}, \\
\pi_{(\delta,\rho)}(2,1) &= \frac{(1 \square \delta)(1 \square \rho)}{4}, \\
\pi_{(\delta,\rho)}(2,2) &= \frac{(1 \square \delta)(1+\rho)}{4}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\pi_{(\delta,\rho)}(1) &= \frac{(1+\delta)}{2}, \\
\pi_{(\delta,\rho)}(2) &= \frac{(1 \square \delta)}{2}, \\
\pi_{(\delta,\rho)}(1; s^1 &= (1)) = \frac{(1+\rho)}{2}, \\
\pi_{(\delta,\rho)}(2; s^1 &= (1)) = \frac{(1 \square \rho)}{2}, \\
\pi_{(\delta,\rho)}(1; s^1 &= (2)) = \frac{(1 \square \rho)}{2}, \\
\pi_{(\delta,\rho)}(2; s^1 &= (2)) = \frac{(1+\rho)}{2}.
\end{aligned}$$

For simplicity and ease of comparison with Backus, Routledge, and Zin (2004), assume u is the identity and there is no discounting (i.e., $\beta = 1$).[13] Normalize ϕ so that $\phi(1) = 1$ and $\phi(0) = 0$.

How is f evaluated? Given that the payoff is zero from time 3 onward, we can solve backwards starting from time 2. $V_{(1,1)}(f) = u(1+\varepsilon) = 1+\varepsilon$, $V_{(1,2)}(f) = \varepsilon$, $V_{(2,1)}(f) = 1$ and $V_{(2,2)}(f) = 0$. Now we write the valuations at the time 1 nodes:

$$\begin{aligned}
V_{(1)}(f) &= \phi^{\square 1}\left[E_{\mu(\cdot|s^1=(1))}\phi\left[\frac{(1+\rho)}{2}(1+\varepsilon) + \frac{(1 \square \rho)}{2}\varepsilon\right]\right] \\
&= \phi^{\square 1}\left[E_{\mu(\cdot|s^1=1)}\phi\left[\frac{(1+\rho)}{2} + \varepsilon\right]\right]
\end{aligned}$$

and

$$\begin{aligned}
V_{(2)}(f) &= \phi^{\square 1}\left[E_{\mu(\cdot|s^1=(2))}\phi\left[\frac{(1 \square \rho)}{2}(1) + \frac{(1+\rho)}{2}(0)\right]\right] \\
&= \phi^{\square 1}\left[E_{\mu(\cdot|s^1=(2))}\phi\left[\frac{(1 \square \rho)}{2}\right]\right]
\end{aligned}$$

[13] This does not strictly satisfy our assumptions, as Assumption 5 requires $\beta \in (0, 1)$.

. Finally,

$$V_{s^0}(f) = \phi^{\square 1}\left[E_\mu \phi\left[\frac{(1+\delta)}{2}V_{(1)}(f) + \frac{(1\ \square\ \delta)}{2}V_{(2)}(f)\right]\right]$$

.

Now take μ to be uniform across the four parameter configurations and assume Bayesian updating. For tractability let

$$\phi(x) = \begin{cases} \frac{1\square e^{\square \alpha x}}{1\square e^{\square \alpha}} & if \quad \alpha > 0 \\ x & if \quad \alpha = 0. \end{cases}$$

We set the same values for $\bar{\delta}, \bar{\rho}$ and ε as Backus, Routledge, and Zin (2004): $\bar{\delta} = \bar{\rho} = \frac{1}{2}$ and $\varepsilon = 1$. This yields $V_{s^0}(f) = \frac{1}{\alpha}\ln\left(\frac{4e^{\frac{3}{2}\alpha}}{\left(1+e^{\frac{\alpha}{2}}\right)^2}\right) = \frac{\ln 4}{\alpha} + \frac{3}{2}\ \square\ \frac{2}{\alpha}\ln\overset{\square}{1} + e^{\frac{\alpha}{2}}\big)$ when $\alpha > 0$ (ambiguity aversion) and $V_{s^0}(f) = 1$ when ambiguity neutrality ($\alpha = 0$) prevails. Backus, Routledge, and Zin (2004) calculate the valuation using the recursive multiple priors model of Epstein and Schneider (2003b) to be $\frac{1}{2}$. This is the valuation obtained in our model in the infinitely ambiguity averse limit. Specifically, $\lim_{\alpha\to\infty} V_{s^0}(f) = \frac{1}{2}$. For more moderate ambiguity aversion, $V_{s^0}(f)$ will lie between $\frac{1}{2}$ and 1 as displayed in Figure 2.

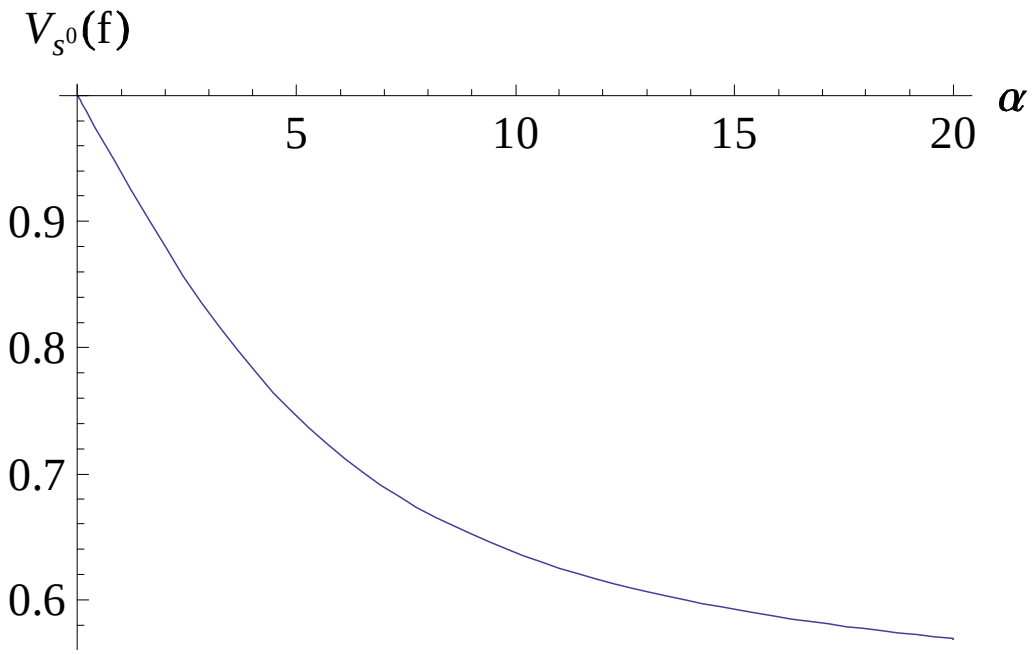

Figure 2: Plot of $V_{s^0}(f)$ as the coefficient of ambiguity aversion, α, increases from 0 to 20.

6.2 Asset pricing example

The point of this example is to *illustrate* the empirical scope of the recursive smooth ambiguity model in an asset pricing context, such as the case of the equity premium puzzle, discussed in the introductory section. The example draws heavily on the recent work of Ju

and Miao (2007) who apply our model. The setting is a Lucas-type intertemporal general equilibrium model of an exchange economy with a single representative agent. The agent trades a stock with uncertain returns with unit supply (a Lucas tree) and a risk-free bond with zero supply. The stock pays dividends D_t in period $t = 0, 1, 2, ...$ The dividend process is

$$\ln\left(\frac{D_{t+1}}{D_t}\right) = \theta_{\bar{g}} + \theta_\sigma \varepsilon_{t+1} \tag{11}$$

where ε_t is i.i.d. standard normal. However, the agent is uncertain about $\theta \equiv (\theta_{\bar{g}}, \theta_\sigma)$, and has a prior μ_0 over Θ, the set of possible θ's. The agent has recursive smooth ambiguity preferences over contingent consumption streams, C, represented as:

$$V_{s^t}(C) = u(c_t) + \beta\phi^{\square 1}\left[\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(C)\, d\pi_\theta(x_{t+1}|s^t)\right) d\mu(\theta|s^t)\right].$$

where $x_t = D_t$ and π_θ is a dividend process as described in (11) and $\mu(\theta|s^t)$ is the Bayesian update of μ_0 given observations of dividends up to time t. Suppose $u(c_t) = \ln c_t$ and $\phi(x) = \square \exp(\square \alpha x)$.

Let $R_{e,t+1}$ and $R_{f,t+1}$ denote the gross returns on the stock and bond, respectively, between periods t and $t+1$. Let W_t denote the period t financial wealth and let $\ _t$ be the proportion of wealth after consumption invested in the uncertain asset. The agent's budget constraint is given by

$$W_{t+1} = (W_t \square c_t) R_{m,t+1}, \tag{12}$$

where the market return $R_{m,t+1}$ is given by

$$R_{m,t+1} = \ _t R_{e,t+1} + (1 \square \ _t) R_{f,t+1}$$

One may formulate the agent's problem using a standard dynamic programming argument, showing that it satisfies the following Bellman equation :

$$J(W_t, \mu_t) = max_{c_t,\ _t} \ln(c_t) + \beta \ln[E_{\mu_t}(\exp(\square \alpha E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1}))))]$$

subject to the budget constraint (12).

Denoting the bond price as p_t^f and the stock price p_t, the FOC (for the case where $u(c_t) = \log c_t$ and $\phi(x) = \square \exp(\square \alpha x)$) is:

$$\begin{aligned} \beta E_{\mu_t^\star}\left[E_{\pi_\theta}\left(u'(c_{t+1})\right)\right] &= p_t^f u'(c_t) \\ \beta E_{\mu_t^\star}\left[E_{\pi_\theta}\left((p_{t+1} + D_{t+1})u'(c_{t+1})\right)\right] &= p_t u'(c_t), \end{aligned}$$

where $\mu_t^\star$ is a measure absolutely continuous w.r.t. μ_t having Radon-Nikodym derivative $\frac{d\mu_t^\star}{d\mu_t}(\theta) = \xi_t(\theta) \equiv \frac{\phi'(E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1})))}{E_{\mu_t}\left[\phi'(E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1})))\right]}$. Using these conditions, it may be shown that the pricing kernel, $M_{t+1,\theta}$, is given by,

$$M_{t+1,\theta} = \beta\left(\frac{c_t}{c_{t+1}}\right)\xi_t(\theta) = \beta\left(\frac{c_t}{c_{t+1}}\right)\frac{\exp(\square \alpha E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1})))}{E_{\mu_t}\left[\exp(\square \alpha E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1})))\right]}$$

and the risk free rate by,

$$\frac{1}{R_{f,t+1}} = E_{\mu_t^\star}\left[M_{t+1,\theta}\right] = \beta E_{\mu_t^\star}\left[E_{\pi_\theta}\left(\frac{D_t}{D_{t+1}}\right)\right].$$

ξ_t is key to understanding why ambiguity aversion may raise the predicted equity premium; it effectively changes the posterior from μ_t to $\mu_t^\star$, with the latter overweighting θ's associated with lower continuation values. It is as if there were more weight on dividend processes with high variances and low means. This may lower the risk free rate and increase the equity premium. The extent of overweighting increases as the coefficient of ambiguity aversion, α, rises. We may assess a plausible range for α by, for instance, looking at the experimental data on ambiguity premiums in Ellsberg-like experiments. Camerer (1999) suggests an ambiguity premium on the order of 10-20% of the expected value of a bet. For the log-exponential specification with a uniform μ, for example, a 10% premium corresponds to an $\alpha \approx 61$.

How much effect ambiguity aversion has depends, of course, on the ambiguity reflected in μ_t. One approach to identifying this is what Hansen (2007) calls "statistical ambiguity." Starting with a uniform prior and updating it on the basis of the data (D_t) would yield posteriors that are largely data driven. There is more ambiguity the "fatter" the posterior, i.e., the less informative the data is about the statistical process driving the future outcomes. While θ in this example must eventually be statistically learned, a specification which includes "hidden states" will allow ambiguity to persist, as explored in Hansen (2007) and suggested in our Example 3. Ju and Miao (2007) pursue this suggestion and show that ambiguity aversion with a power-power specification for ϕ and u may explain a significant part of the equity premium along with related empirical phenomena.

Thus the recursive smooth ambiguity model provides a framework which one may use to assess whether ambiguity may plausibly and significantly affect the equity premium, and more generally, asset market equilibria.

7 Related literature

A number of recent papers have proposed dynamic preference models allowing for ambiguity. The most obvious difference between our model and all others is that only ours is a recursive extension of the smooth ambiguity model of KMM. The only other work we know of proposing any dynamic extension of the smooth ambiguity model is the recent model of updating smooth ambiguity preferences developed in Hanany and Klibanoff (2007a). Their approach is fundamentally different, in that the KMM model is taken to determine preferences over plans and conditional preferences are found by updating beliefs, μ. This method of extension is shown to be incompatible with the dynamic consistency assumed in this paper, thus leading to a non-recursive model. Imposing a dynamic consistency property that weakens consequentialism, they derive an update rule. Since their dynamic consistency is enough so that plans chosen according to the KMM model are carried out, the Hanany and Klibanoff (2007a) model allows for different behavior than our recursive approach, including, for example, plausible behavior in some natural dynamic extensions of the Ellsberg Paradox. On the other hand, their update rule is non-consequentialist in that it generally requires

updating to depend on past choices, and the lack of recursion may make solving problems more difficult.

Many of the existing dynamic ambiguity models extend the MEU and/or CEU models. We briefly discuss some of these models here. We focus on those models that display at least some form of dynamic consistency. The tension between non-expected utility and dynamic consistency (e.g., Machina (1989), Karni and Schmeidler (1991), and Ghirardato (2002)) and, more specifically, non-probabilistically sophisticated behavior (such as sensitivity to ambiguity) and dynamic consistency (e.g., Epstein and LeBreton (1993)) is well known. For example, the best known proposals for extending MEU or CEU by applying updating procedures (e.g., Dempster-Shafer updating or Full Bayesian updating or Maximum Likelihood updating) to the sets of measures or non-additive measures appearing in those theories all fail to generally satisfy dynamic consistency. As mentioned in the Introduction, Epstein and Schneider (2003b), Wang (2003), Hayashi (2005) have provided preference foundations for extending the MEU model to an intertemporal framework while preserving the dynamic consistency needed for recursion. Epstein and Schneider (2003b) characterize the recursive subclass of the MEU model. It builds on the insight of Sarin and Wakker (1998) that limiting attention to particular filtrations or decision trees may relax the force of dynamic consistency enough to allow the consistent extension of some MEU preferences. The Wang (2003) model satisfies a stronger dynamic consistency condition and maintains recursivity while allowing for attitudes towards the timing of the resolution of uncertainty (a la Kreps and Porteus (1978), Epstein and Zin (1989)) and departing from the MEU class of preferences. Hayashi (2005) provides a generalization of Epstein and Schneider (2003b) also in the direction of allowing for attitudes towards the timing of the resolution of uncertainty, also maintaining recursivity and, except where it coincides with Epstein and Schneider (2003b), leaving the MEU class. The essential lack of reduction to MEU over plans in these two papers is analogous to the lack of reduction of our model to the KMM smooth ambiguity model over plans. In all three cases it is this lack of reduction that is key to accommodating the dynamic consistency satisfied by recursive models. Hansen and Sargent (2001) describes a robust control approach to model uncertainty. One formulation of their model may be viewed as a dynamic extension of a subset of MEU. As described by Epstein and Schneider (2003b), this extension satisfies a version of dynamic consistency that requires consistency only relative to a given optimal plan and updates sets of measures in a way that is non-consequentialist. Hanany and Klibanoff (2007b) is not recursive in general, but provides a way of extending the *entire* class of MEU preferences by requiring dynamic consistency but, as mentioned above when discussing Hanany and Klibanoff (2007a), allowing updating of the set of measures to depend on non-consequentialist aspects of the choice problem. It extends Epstein and Schneider's recursive multiple priors model in the sense that for the subset of MEU preferences defined by that model, both approaches are equivalent to Full Bayesian updating (i.e., applying Bayes' rule to each measure in the set of measures). All of these approaches, however, share the limitation inherent to the MEU model of failing to separate ambiguity from ambiguity attitude without restricting the range of ambiguity attitudes. A similar limitation also applies to some recent extensions of CEU proposed by Eichberger, Grant, and Kelsey (2005) and Nishimura and Ozaki (2003). Recently, Maccheroni, Marinacci, and Rustichini (2006a) developed a class of preferences called variational preferences, which include both MEU preferences and a version of Hansen and Sargent (2001) preferences known as multi-

plier preferences as special cases. In Maccheroni, Marinacci, and Rustichini (2006b), they extend variational preferences to an intertemporal setting and also characterize when this extension is dynamically consistent, thus generalizing the characterization of Epstein and Schneider (2003b) for MEU. However, as with MEU and CEU, variational preferences are also subject to the limitation mentioned above concerning the separation of ambiguity from ambiguity attitude.

One view of the model in the present paper is as representing the preferences of a DM concerned with robustness to model/parameter uncertainty. As we discussed in Section 5, the model allows for learning through updating beliefs concerning this uncertainty. The combination of robustness concerns with learning has been mentioned as an important challenge for the literature on model uncertainty by Hansen and Sargent (2006) who also investigate these issues. As mentioned in the application section, Ju and Miao (2007) uses log-exponential and power-power specifications of our model to examine asset prices under ambiguity or concern for robustness and learning, while Hansen (2007) looks at these concerns using a continuous-time version of a log-exponential specification.

Siniscalchi (2004) takes a different approach to dynamic behavior than the rest of the literature we have discussed. Rather than focusing on dynamic consistency of preferences over acts, following the literature starting from Strotz (1955-6) he accepts the fact that preferences over acts may be dynamically inconsistent and investigates behavior that is *sophisticated* in the sense that when the DM is considering a decision tree that involves future choices, it is assumed that the DM can correctly anticipate those choices (which will be governed by conditional preferences). Then the unconditional preferences are applied in comparing the acts that result from replacing the future choice nodes with the future choices. His framework is not tied to a specific model. In principle, this could form the basis of an alternative strategy for extending smooth ambiguity preferences to dynamic settings.

8 Conclusion

We have proposed, axiomatized and investigated a model of recursive preferences over intertemporal plans, extending the smooth ambiguity model developed in Klibanoff, Marinacci, and Mukerji (2005) to a setting involving dynamic decision making. The model has the desirable properties of allowing sensitivity to ambiguity, of separating ambiguity attitude from ambiguity perception, of dynamic consistency and consequentialism, of a well-founded theory of updating beliefs via Bayes' rule, thereby generating a theory of learning under ambiguity and, finally, of nesting the standard discounted expected utility model as a special case. One notable benefit of the separation provided is the ability to do comparative statics in ambiguity attitude for dynamic problems while holding information and ambiguity perception unchanged. A full range of ambiguity attitudes, including ambiguity neutrality, is available for any given beliefs.

9 Appendix A: Proofs

Proof of Lemma 1. Let $t, t' \in \mathcal{T} \cup \{0\}$. Set $c'_1 = c''_1 = c_1 \equiv c'$, $k'_1 = k''_1 = k_1 \equiv c''$, $c'_2 = c''_2 = c_2 = k'_2 = k''_2 = k_2 \equiv c$. Given $c \in \mathcal{C}$ and $s^t \in S^t$, let $c_{s^t} \in \mathcal{F}_{s^t}$ be the deterministic

continuation plan such that $c_{s^t}(\tau) = c$ for all $\tau \geq t$. Then, Assumption 6 means that, for all $p \in [0,1]$ and all $c, c', c'' \in \mathcal{C}$, we have:

$$\{c'_{s^t}, p; c''_{s^t}, 1 \square p\} \sim_{s^t} c_{s^t} \iff \left\{c'_{s^{t'}}, p; c''_{s^{t'}}, 1 \square p\right\} \sim_{s^{t'}} c_{s^{t'}}. \tag{13}$$

Then, Assumption 5 implies that, for all $p \in [0,1]$ and all $c, c', c'' \in \mathcal{C}$,

$$\begin{aligned} &\frac{u_{s^t}(c')}{1 \square \beta_{s^t}} p + \frac{u_{s^t}(c'')}{1 \square \beta_{s^t}} (1 \square p) = \frac{u_{s^t}(c)}{1 \square \beta_{s^t}} \\ &\iff \frac{u_{s^{t'}}(c')}{1 \square \beta_{s^{t'}}} p + \frac{u_{s^{t'}}(c'')}{1 \square \beta_{s^{t'}}} (1 \square p) = \frac{u_{s^{t'}}(c)}{1 \square \beta_{s^{t'}}}, \end{aligned} \tag{14}$$

and so, for all $p \in [0,1]$ and all $c, c', c'' \in \mathcal{C}$,

$$\begin{aligned} &u_{s^t}(c')\, p + u_{s^t}(c'')\,(1 \square p) = u_{s^t}(c) \\ &\iff u_{s^{t'}}(c')\, p + u_{s^{t'}}(c'')\,(1 \square p) = u_{s^{t'}}(c). \end{aligned} \tag{15}$$

Normalize u_{s^t} and $u_{s^{t'}}$ so that

$$u_{s^t}(c^*) = u_{s^{t'}}(c^*) = 1 \quad \text{and} \quad u_{s^t}(c^{**}) = u_{s^{t'}}(c^{**}) = 0 \tag{16}$$

for some $c^*, c^{**} \in \mathcal{C}$. Taking $p = 1$, there exists a strictly increasing such that

$$u_{s^t}(c) = (u_{s^{t'}}(c)), \qquad \forall c \in \mathcal{C}. \tag{17}$$

In particular, $(1) = 1$ because $1 = u_{s^t}(c^*) = (u_{s^{t'}}(c^*)) = (1)$. By (15) and (17), for all $p \in [0,1]$ and all $c', c'' \in \mathcal{C}$ we have:

$$(u_{s^{t'}}(c'))\, p + (u_{s^{t'}}(c''))\,(1 \square p) = (u_{s^{t'}}(c')\, p + u_{s^{t'}}(c'')\,(1 \square p)).$$

Hence, $(1)\, p = (p)$ for all $p \in [0,1]$, and so (17) implies $u_{s^t}(c) = u_{s^{t'}}(c)$ for all $c \in \mathcal{C}$.

Set $u_{s^t} = u_{s^{t'}} \equiv u$. When $p = 1$ and $c_1 = c_2 \equiv c$, Assumption 6 means that, for all $c, c'_i, c''_i \in \mathcal{C}$ with $i = 1, 2$,

$$\{(c'_1, c''_1, c)_t, 1\} \succcurlyeq_{s^t} \{(c'_2, c''_2, c)_t, 1\} \iff \{(c'_1, c''_1, c)_{t'}, 1\} \succcurlyeq_{s^{t'}} \{(c'_2, c''_2, c)_{t'}, 1\}.$$

By Assumption 5,

$$\begin{aligned} &u(c'_1) + \beta_{s^t} u(c''_1) \geq u(c'_2) + \beta_{s^t} u(c''_2) \\ &\iff u(c'_1) + \beta_{s^{t'}} u(c''_1) \geq u(c'_2) + \beta_{s^{t'}} u(c''_2). \end{aligned} \tag{18}$$

Since $\mathcal{C}$ is an interval of $\mathbb{R}$ and u is continuous, the normalization (16) we chose is such that $[0,1] \subseteq u(\mathcal{C})$. Hence, we can choose $c'_1, c''_1, c'_2, c''_2 \in \mathcal{C}$ so that $u(c''_2) \geq u(c''_1)$, and

$$\beta_{s^t} = \frac{u(c'_1) \square u(c'_2)}{u(c''_2) \square u(c''_1)}.$$

Thus, (18) implies $\beta_{s^t} \geq \beta_{s^{t'}}$. A similar argument proves the converse inequality, and so $\beta_{s^t} = \beta_{s^{t'}}$. ■

Proof of Lemma 2. Consider the ordering $\succcurlyeq^*_{s^t}$ restricted to $\mathcal{F}_{s^t}$. W.l.o.g., set $\mathcal{C} = [m, M]$. By Assumption 2, $M \succcurlyeq^*_{s^t} f_{|s^t} \succcurlyeq^*_{s^t} m$. Suppose $M \succ^*_{s^t} f_{|s^t} \succ^*_{s^t} m$, otherwise the result is trivially true. Set

$$\alpha^* = \sup\left\{\alpha \in [0,1] : f_{|s^t} \succcurlyeq^*_{s^t} \alpha M + (1 \square \alpha)\, m\right\}.$$

Suppose $\alpha^* = 0$, i.e., $\alpha M + (1 \square \alpha)\, m \succ^*_{s^t} f_{|s^t}$ for all $\alpha \in (0,1]$. As $M \succ^*_{s^t} f_{|s^t} \succ^*_{s^t} m$, by Assumption 3, there is $\alpha \in (0,1)$ such that $f_{|s^t} \succ^*_{s^t} \alpha M + (1 \square \alpha)\, m$, a contradiction. Next, suppose $\alpha^* = 1$. This means that $f_{|s^t} \succcurlyeq^*_{s^t} \alpha M + (1 \square \alpha)\, m$ for all $\alpha \in [0,1)$. As $M \succ^*_{s^t} f_{|s^t} \succ^*_{s^t} m$, by Assumption 3, there is $\alpha \in (0,1)$ such that $\alpha M + (1 \square \alpha)\, m \succ^*_{s^t} f_{|s^t}$, a contradiction.

Finally, suppose $\alpha^* \in (0,1)$. If $\alpha \in (\alpha^*, 1]$, by the definition of α^* we have $\alpha M + (1 \square \alpha)\, m \succ^*_{s^t} f_{|s^t}$. If $\alpha \in [0, \alpha^*)$, Assumption 2 implies $\alpha^* M + (1 \square \alpha^*)\, m \succ^*_{s^t} \alpha M + (1 \square \alpha)\, m$. Hence, $f_{|s^t} \succcurlyeq^*_{s^t} \alpha M + (1 \square \alpha)\, m$. For, suppose *per contra* that $\alpha M + (1 \square \alpha)\, m \succ^*_{s^t} f_{|s^t}$. Then, Assumption 2 implies $\beta M + (1 \square \beta)\, m \succ^*_{s^t} f_{|s^t}$ for all $\beta \in [\alpha, 1]$, which contradicts the definition of α^*.

It remains to prove that $\alpha^* M + (1 \square \alpha^*)\, m \sim^*_{s^t} f_{|s^t}$. Suppose $f_{|s^t} \succ^*_{s^t} \alpha^* M + (1 \square \alpha^*)\, m$. By Assumption 3 and consequentialism, there is $\beta \in (0,1)$ such that

$$f_{|s^t} \succ^*_{s^t} \beta\left(\alpha^* M + (1 \square \alpha^*)\, m\right) + (1 \square \beta)\, M.$$

That is, $f_{|s^t} \succ^*_{s^t} \left(\alpha^* \beta + (1 \square \beta)\right) M + \beta\,(1 \square \alpha^*)\, m$. This implies $f_{|s^t} \succ^*_{s^t} \alpha M + (1 \square \alpha)\, m$ for all $\alpha \in [0, \alpha^* \beta + (1 \square \beta)]$, which contradicts the definition of α^* since $\alpha^* \beta + (1 \square \beta) > \alpha^*$.

Suppose $\alpha^* M + (1 \square \alpha^*)\, m \succ^*_{s^t} f_{|s^t}$. By Assumptions 3 and consequentialism, there is $\beta \in (0,1)$ such that

$$\beta\left(\alpha^* M + (1 \square \alpha^*)\, m\right) + (1 \square \beta)\, m \succ^*_{s^t} f_{|s^t}.$$

That is, $\beta\alpha^* M + \left(\beta\,(1 \square \alpha^*) + (1 \square \beta)\right) m \succ^*_{s^t} f_{|s^t}$. This implies $\alpha M + (1 \square \alpha)\, m \succ^*_{s^t} f_{|s^t}$ for all $\alpha \in [\beta\alpha^*, 1]$, which contradicts the definition of α^* since $\beta\alpha^* < \alpha^*$. ■

Proof of Proposition 1. We begin by showing that the function u is strictly increasing. Let c_1, $c_2 \in \mathcal{F}$ be two constant plans, with $c_1(t) = c_1$ and $c_2(t) = c_2$ $\forall t \geq 0$. Then, by Assumptions 2, 5, Lemma 1, and by the definition of $\succcurlyeq^*_{s^t}$, we have:

$$\begin{aligned} c_1 \;>\; c_2 &\Rightarrow c_1 \succ_{s^t} c_2 \Rightarrow c_{1|s^t} \succ^*_{s^t} c_{2|s^t} \\ &\Rightarrow U\left(c_{1|s^t}\right) > U\left(c_{2|s^t}\right) \\ &\Rightarrow \frac{u(c_1)}{1 \square \beta} > \frac{u(c_2)}{1 \square \beta} \Rightarrow u(c_1) > u(c_2). \end{aligned}$$

Next consider $f_{s^t} \in \mathcal{F}^*_{s^t}$ and the associated $c_{f_{s^t}}(\theta)$. By definition,

$$\left(f_{s^t}, s^t\right), c_{f_{s^t}}(\theta), c_{f_{s^t}}(\theta), ..., c_{f_{s^t}}(\theta), ...\right) \sim^*_{s^t} \left\{d_{f_{s^t}}(x_{t+1}), \pi_\theta\left(x_{t+1}; s^t\right)\right\}_{x_{t+1} \in \mathcal{X}_{t+1}}$$

By Assumption 5,

$$\begin{aligned} & u(f_{s^t}(s^t)) + \frac{\beta u(c_{f_{s^t}}(\theta)))}{1 \square \beta} \\ = & \sum_{x_{t+1} \in \mathcal{X}_{t+1}} \pi_\theta\left(x_{t+1}; s^t\right) \left[u(f_{s^t}(s^t)) + \frac{\beta u(f_{s^t}(s^t, x_{t+1}))}{1 \square \beta}\right]. \end{aligned}$$

Hence,

$$u(c_{f_{s^t}}(\theta)) = \sum_{x_{t+1}\in\mathcal{X}_{t+1}} \pi_\theta \,\square\, x_{t+1}; s^t)\, u(f_{s^t}(s^t, x_{t+1})). \tag{19}$$

By Assumption 8, $f_{s^t} \succcurlyeq^*_{s^t} g_{s^t} \iff f^2 \succcurlyeq^2_{s^t} g^2$. By Assumption 7, $f^2 \succcurlyeq^2_{s^t} g^2 \Leftrightarrow \int_\Theta v \,\square\, c_{f_{s^t}}(\theta))\, d\mu_{s^t} \geq \int_\Theta v \,\square\, c_{g_{s^t}}(\theta))\, d\mu_{s^t}$. Hence,

$$f_{s^t} \succcurlyeq^*_{s^t} g_{s^t} \iff \int_\Theta v \,\square\, c_{f_{s^t}}(\theta))\, d\mu_{s^t} \geq \int_\Theta v \,\square\, c_{g_{s^t}}(\theta))\, d\mu_{s^t}. \tag{20}$$

Since v and u are strictly increasing, $v \,\square\, c_{f_{s^t}}(\theta)) = \phi\left(\frac{u\left(c_{f_{s^t}}(\theta)\right)}{1\square\beta}\right)$ for some strictly increasing ϕ. In particular, $\phi = v \circ u^{\square 1} \circ (1 \square \beta)$. Since v and u are continuous, so is ϕ. Substituting for $v \,\square\, c_{f_{s^t}}(\theta))$ in (20) and using (19), we get

$$\begin{aligned}
& f_{s^t} \succcurlyeq^*_{s^t} g_{s^t} \\
\iff & \int \phi\left(\frac{u \,\square\, c_{f_{s^t}}(\theta))}{1 \square \beta}\right) d\mu_{s^t} \geq \int \phi\left(\frac{u \,\square\, c_{g_{s^t}}(\theta))}{1 \square \beta}\right) d\mu_{s^t} \\
\iff & \int \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(f_{s^t}(s^t, x_{t+1}))}{1 \square \beta} d\pi_\theta \,\square\, x_{t+1}; s^t)\right) d\mu_{s^t} \\
\geq & \int \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(g_{s^t}(s^t, x_{t+1}))}{1 \square \beta} d\pi_\theta \,\square\, x_{t+1}; s^t)\right) d\mu_{s^t}
\end{aligned}$$

which proves the representation claim in the Proposition. The uniqueness claims follow straightforwardly from the uniqueness in Assumptions 5 and 7. ■

Proof of Theorem 1. We begin by showing that the assumptions imply the representations. Using Lemma 1, Assumption 5 and Assumption 6 imply the representation EU_{s^t}. Assumption 7 and setting $\phi = v \circ u^{\square 1} \circ (1 \square \beta)$ delivers the representation $V^2_{s^t}$ over second order acts. The argument for the representation V_{s^t} of preferences over plans is more involved. By Definition 5, $f_{|s^t, x_{t+1}} \sim^*_{(s^t, x_{t+1})} \vec{c}_{f,(s^t,x_{t+1})}$ for any $x_{t+1} \in \mathcal{X}_{t+1}$ and any plan $f \in \mathcal{F}$. Define a continuation plan $\tilde{f}_{s^t} \in \mathcal{F}_{s^t}$ as follows:

$$\tilde{f}_{s^t}\left(s^{t'}\right) = \begin{cases} f(s^t) & \text{if} \quad s^{t'} = s^t \\ \vec{c}_{f,(s^t,x_{t+1})}(s^{t'}) & \text{if} \quad s^{t'} \geqslant (s^t, x_{t+1}) \end{cases}$$

By Assumption 4, given any $f, g \in \mathcal{F}$,

$$f_{|s^t,x_{t+1}} \succcurlyeq^*_{(s^t,x_{t+1})} g_{|s^t,x_{t+1}} \text{ for all } x_{t+1} \Longrightarrow f_{|s^t} \succcurlyeq^*_{s^t} g_{|s^t} \tag{21}$$

whenever $f(s^t) = g(s^t)$. By (21), $f_{|s^t} \sim^*_{s^t} \tilde{f}_{s^t}$. Moreover, the continuation plan $\tilde{f}_{s^t}$ belongs to $\mathcal{F}^*_{s^t}$.

The function $V_{s^t} : \mathcal{F} \to \mathbb{R}$ defined by $V_{s^t}(f) = U(\vec{c}_{f,s^t})$ represents $\succcurlyeq_{s^t}$ on plans. Moreover, by Assumption 5 and Assumption 6,

$$V_{s^t}(f) = U(\vec{c}_{f,s^t}) = \sum_{\tau \geq t} \beta^{\tau \square t} u(c_{f,s^t}) = \frac{u(c_{f,s^t})}{1 \square \beta}. \tag{22}$$

Assumption 2 ensures that V_{s^t} is monotonic in the sense mentioned in the text before the statement of Theorem 1.

Suppose $f, g \in \mathcal{F}$ are such that $f(s^t) = g(s^t)$. By definition,

$$\vec{c}_{f,(s^t,x_{t+1})}(t') = c_{f,(s^t,x_{t+1})}$$

and

$$\vec{c}_{g,(s^t,x_{t+1})}(t') = c_{g,(s^t,x_{t+1})} \text{ for all } t' \geq t+1.$$

By (22),

$$V_{(s^t,x_{t+1})}(f)(1 \square \beta) = u(c_{f,(s^t,x_{t+1})})$$

and

$$V_{(s^t,x_{t+1})}(g)(1 \square \beta) = u(c_{g,(s^t,x_{t+1})})$$

for each (s^t, x_{t+1}). By Proposition 1,

$$\begin{aligned}
& f \succcurlyeq_{s^t} g \\
\iff & f_{|s^t} \succcurlyeq^*_{s^t} g_{|s^t} \iff \tilde{f}_{s^t} \succcurlyeq^*_{s^t} \tilde{g}_{s^t} \\
\iff & \int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(c_{f,(s^t,x_{t+1})})}{1 \square \beta} d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t} \\
\geq & \int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(c_{g,(s^t,x_{t+1})})}{1 \square \beta} d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t} \\
\iff & \int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \frac{(1 \square \beta) V_{(s^t,x_{t+1})}(f)}{1 \square \beta} d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t} \\
\geq & \int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \frac{(1 \square \beta) V_{(s^t,x_{t+1})}(g)}{1 \square \beta} d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t} \\
\iff & \phi^{\square 1}\left[\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} V_{(s^t,x_{t+1})}(f) d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t}\right] \\
\geq & \phi^{\square 1}\left[\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} V_{(s^t,x_{t+1})}(g) d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t}\right].
\end{aligned}$$

All this implies that $u(f(s^t)) + \beta\phi^{\square 1}\left[\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} V_{(s^t,x_{t+1})}(f) d\pi_\theta(x_{t+1}; s^t)\right) d\mu_{s^t}\right]$ represents $\succcurlyeq_{s^t}$ over plans sharing the same payoff at s^t.

To extend the representation to all plans, define $f^*_{s^t} \in \mathcal{F}$ as follows:[14]

$$f^*_{s^t}\left(s^{t'}\right) = \begin{cases} f \overset{\square}{} s^{t'}) & \text{if } s^{t'} \not\succcurlyeq s^t \\ u^{\square 1} \circ (1 \square \beta)\phi^{\square 1} \int\limits_\Theta \phi \int\limits_{\mathcal{X}_{t+1}} V_{(s^t,x_{t+1})}(f) d\pi_\theta(\cdot; s^t)\Big) d\mu_{s^t} & \text{if } s^{t'} \succcurlyeq s^t \\ & \text{and } t' \geq t+1 \end{cases}$$

Note that $f^*_{s^t}$ shares the same payoff with f at s^t and that the outer inverse in the bottom term exists because ϕ is defined over the range of $\frac{u}{1 \square \beta}$. Now we show that $f \sim_{s^t} f^*_{s^t}$. Observe

[14] Note $f^*_{s^t}$ is a plan, not a continuation plan; to this extent there is here some abuse of notation.

that,

$$\begin{aligned}
&\int_{\Theta}\phi\left(\int_{\mathcal{X}_{t+1}}V_{(s^t,x_{t+1})}\left(f^*_{s^t}\right)d\pi_\theta \overset{\square}{} x_{t+1};s^t\right)\right)d\mu_{s^t}\\
=&\int_{\Theta}\phi\left(\int_{\mathcal{X}_{t+1}}\frac{u \quad u^{\square 1}\circ\left(1\,\square\,\beta\right)\phi^{\square 1} \quad \int_{\Theta}\phi \quad \int_{\mathcal{X}_{t+1}}V_{(s^t,x_{t+1})}\left(f\right)d\pi_\theta\left(x_{t+1};s^t\right)\right)d\mu_{s^t}\Big)\Big)}{1\,\square\,\beta}d\pi_\theta \overset{\square}{} x_{t+1};s^t\right)\right)d\mu_{s^t}\\
=&\int_{\Theta}\phi\left(\int_{\mathcal{X}_{t+1}}V_{(s^t,x_{t+1})}\left(f\right)d\pi_\theta \overset{\square}{} x_{t+1};s^t\right)\right)d\mu_{s^t}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&u \overset{\square}{} f^*_{s^t} \overset{\square}{} s^t\right)\right)+\beta\phi^{\square 1}\left[\int_{\Theta}\phi\left(\int_{\mathcal{X}_{t+1}}V_{(s^t,x_{t+1})}\left(f^*_{s^t}\right)d\pi_\theta \overset{\square}{} x_{t+1};s^t\right)\right)d\mu_{s^t}\right]\\
=&\ u \overset{\square}{} f \overset{\square}{} s^t\right)\right)+\beta\phi^{\square 1}\left[\int_{\Theta}\phi\left(\int_{\mathcal{X}_{t+1}}V_{(s^t,x_{t+1})}\left(f\right)d\pi_\theta \overset{\square}{} x_{t+1};s^t\right)\right)d\mu_{s^t}\right].
\end{aligned}$$

So, by what we proved before, $f\sim_{s^t} f^*_{s^t}$. Note that any $f^*_{s^t}$ has a deterministic continuation plan $f^*_{s^t|s^t}$ at s^t. Hence, by Assumption 5,

$$\begin{aligned}
&V_{s^t}\left(f\right)=V_{s^t}\left(f^*_{s^t}\right)=U \overset{\square}{} f^*_{s^t|s^t}\right)\\
=&\ u \overset{\square}{} f \overset{\square}{} s^t\right)\right)+\frac{\beta}{1\,\square\,\beta}u\circ u^{\square 1}\circ\left(1\,\square\,\beta\right)\phi^{\square 1}\left(\int_{\Theta}\phi\left(\int_{\mathcal{X}_{t+1}}V_{(s^t,x_{t+1})}\left(f\right)d\pi_\theta \overset{\square}{} x_{t+1};s^t\right)\right)d\mu_{s^t})\right)\\
=&\ u \overset{\square}{} f \overset{\square}{} s^t\right)\right)+\beta\phi^{\square 1}\left(\int_{\Theta}\phi\left(\int_{\mathcal{X}_{t+1}}V_{(s^t,x_{t+1})}\left(f\right)d\pi_\theta \overset{\square}{} x_{t+1};s^t\right)\right)d\mu_{s^t}\right),
\end{aligned}$$

as desired.

Now we show that the representations imply the assumptions. Assumptions 1 and 3 follow from the real-valued nature of the representations V_{s^t}. Assumption 2 follows from the stated monotonicity of the V_{s^t}. Assumption 4 follows from the recursivity of V_{s^t}. That for each $d_{s^t}\in\mathcal{D}_{s^t}$, there exists a plan $f\in\mathcal{F}$ with $f_{|s^t}=d_{s^t}$ such that $f\sim_{s^t}\{d_{s^t},1\}$ follows from observing that V_{s^t} collapses to discounted expected utility as in EU_{s^t} for plans deterministic from s^t onward (that this is the unique solution for V_{s^t} in such a case follows from Proposition 3). The remainder of Assumption 5 and Assumption 6 may be verified from the representation EU_{s^t}. Substituting $v=\phi\left(\frac{u}{1\square\beta}\right)$ in the representation $V^2_{s^t}$ yields Assumption 7. To verify

Assumption 8, observe that for $f_{s^t}, g_{s^t} \in \mathcal{F}^*_{s^t}$ and associated $f^2, g^2 \in \mathfrak{F}$, if $f_{s^t}(s^t) = g_{s^t}(s^t)$,

$$\begin{aligned}
& f^2 \succcurlyeq^2_{s^t} g^2 \\
\iff & \int_\Theta \phi\left(\frac{u(c_{f_{s^t}}(\theta))}{1 \square \beta}\right) d\mu_{s^t} \\
\geq & \int_\Theta \phi\left(\frac{u(c_{g_{s^t}}(\theta))}{1 \square \beta}\right) d\mu_{s^t} \\
\iff & \int \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(f_{s^t}(s^t, x_{t+1}))}{1 \square \beta} d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t} \\
\geq & \int \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(g_{s^t}(s^t, x_{t+1}))}{1 \square \beta} d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t} \\
\iff & u \overset{\square}{} f_{s^t}(s^t)\big) + \beta\phi^{\square 1}\left(\int \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(f_{s^t}(s^t, x_{t+1}))}{1 \square \beta} d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t}\right) \\
\geq & u \overset{\square}{} g_{s^t}(s^t)\big) + \beta\phi^{\square 1}\left(\int \phi\left(\int_{\mathcal{X}_{t+1}} \frac{u(g_{s^t}(s^t, x_{t+1}))}{1 \square \beta} d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t}\right) \\
\iff & V_{s^t}(f) \geq V_{s^t}(g) \text{ for } f, g \text{ with } f_{|s^t} = f_{s^t} \text{ and } g_{|s^t} = g_{s^t}.
\end{aligned}$$

■

The next lemma is a standard result (see e.g., page 401 of Durrett (1991).)

Lemma 3 *Suppose $v^1, v^2 : \Sigma \to [0,1]$ are countably additive probabilities. If $v^1(B) = v^2(B) > 0$ for all **elementary cylinder sets** B, then $v^1 = v^2$.*

Proof of Proposition 2. Define η on elementary cylinder sets by (5). Lemma 3 says η is the unique countably additive extension to Σ. We next show that (ii) holds by calculating $\left.\frac{\partial V_{s^t}(f)/\partial\left[u(f(\left(s^t, x_{t+1}, \ldots, x_{t+n}\right)))\right]}{\partial V_{s^t}(f)/\partial[u(f(s^t))]}\right|_{f=d}$.

By Theorem 1, $V_{s^t}(f) = u(f(s^t)) + \beta\phi^{\square 1}\left[\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_\theta(\cdot; s^t)\right) d\mu_{s^t}\right]$. Differentiating $V_{s^t}(f)$ with respect to $u(f(s^t))$ yields $\partial V_{s^t}(f)/\partial[u(f(s^t))] = 1$.

Straightforward but tedious calculations, omitted for brevity, may be used to show that, for general n, when evaluated at any deterministic f,

$$\begin{aligned}
& \left.\frac{\partial V_{s^t}(f)/\partial\left[u(f(\overset{\square}{} s^t, x^*_{t+1}, \ldots, x^*_{t+n}))\right]}{\partial V_{s^t}(f)/\partial[u(f(s^t))]}\right|_{f=d} \\
= & \beta^n P_{s^t} \overset{\square}{} x^*_{t+1}\big) P_{s^t, x^*_{t+1}} \overset{\square}{} x^*_{t+2}\big) * \ldots * P_{s^t, x^*_{t+1}, \ldots, x^*_{t+n \square 1}} \overset{\square}{} x^*_{t+n}\big) \qquad (24) \\
= & \beta^n \frac{\eta \overset{\square}{} s^t, x^*_{t+1}, \ldots, x^*_{t+n}\big)}{\eta(s^t)} = \beta^n \eta_{s^t} \overset{\square}{} x^*_{t+1}, \ldots, x^*_{t+n}\big).
\end{aligned}$$

This proves (ii). To show (i), suppose ϕ is affine, so that $\phi(x) \equiv ax + b$ for some $a > 0, b \in \mathbb{R}$. Substituting into (3) yields,

$$\begin{aligned}
V_{s^t}(f) &= u \overset{\square}{} f \overset{\square}{} s^t)\big) + \beta\left(\frac{1}{a}\left[a\int_\Theta \left(\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_\theta \overset{\square}{} x_{t+1}; s^t\right)\right) d\mu_{s^t}(\theta) + b\right] \square \frac{b}{a}\right) \\
&= u \overset{\square}{} f \overset{\square}{} s^t)\big) + \beta\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, dP_{s^t}(x_{t+1}).
\end{aligned}$$

By Theorem 3, such a recursive functional has a unique and monotonic solution. We now check that (4) is that solution. Substituting for $V_{(s^t,x_{t+1})}(f)$ yields

$$\begin{aligned}
V_{s^t}(f) &= u\left(f\left(s^t\right)\right)+\beta\int_{\mathcal{X}_{t+1}}\int_S\left[\sum_{\tau=t+1}^{\infty}\beta^{\tau\square(t+1)}u\left(f\left(s^\tau\right)\right)\right]d\eta_{(s^t,x_{t+1})}\left(s^\tau\right)\Bigg)dP_{s^t}\left(x_{t+1}\right)\\
&= u\left(f\left(s^t\right)\right)+\beta\int_S\left[\sum_{\tau=t+1}^{\infty}\beta^{\tau\square(t+1)}u\left(f\left(s^\tau\right)\right)\right]d\left(\int_{\mathcal{X}_{t+1}}\eta_{(s^t,x_{t+1})}\left(s^\tau\right)dP_{s^t}\left(x_{t+1}\right)\right)\\
&= u\left(f\left(s^t\right)\right)+\beta\int_S\left[\sum_{\tau=t+1}^{\infty}\beta^{\tau\square(t+1)}u\left(f\left(s^\tau\right)\right)\right]d\left(\int_{\mathcal{X}_{t+1}}\frac{\eta\left(s^\tau\cap\left(s^t,x_{t+1}\right)\right)}{\eta\left(s^t,x_{t+1}\right)}dP_{s^t}\left(x_{t+1}\right)\right)\\
&= u\left(f\left(s^t\right)\right)+\beta\int_S\left[\sum_{\tau=t+1}^{\infty}\beta^{\tau\square(t+1)}u\left(f\left(s^\tau\right)\right)\right]d\left(\frac{\eta\left(s^\tau\cap s^t\right)}{\eta\left(s^t\right)}\right)\\
&= \int_S\left[\sum_{\tau=t}^{\infty}\beta^{\tau\square t}u\left(f\left(s^\tau\right)\right)\right]d\eta_{s^t}\left(s^\tau\right).
\end{aligned}$$

This proves (i) and completes the proof.. ■

Proof of Theorem 4. Assume the predictive distributions are related by Bayes' rule. From Proposition 2,

$$\begin{aligned}
&\left.\frac{\partial V_{s^t}(f)/\partial\left[u(f((s^t,x_{t+1},\ldots,x_{t+n})))\right]}{\partial V_{s^t}(f)/\partial\left[u(f(s^t))\right]}\right|_{f=d}=\beta^n\eta_{s^t}\left(\left\{s^t,x_{t+1},\ldots,x_{t+n}\right\}\right)\\
&= \beta^n P_{s^t}\left(x_{t+1}\right)*\ldots*P_{s^t,x_1,\ldots,x_{t+n\square1}}\left(x_{t+n}\right)\\
&= \beta^n\frac{P_{s^0}\left(s^t,x_{t+1}\right)}{P_{s^0}\left(s^t\right)}\frac{P_{s^0}\left(s^t,x_{t+1},x_{t+2}\right)}{P_{s^0}\left(s^t,x_{t+1}\right)}\cdots\frac{P_{s^0}\left(s^t,x_{t+1},\ldots,x_{t+n}\right)}{P_{s^0}\left(s^t,x_{t+1},\ldots,x_{t+n\square1}\right)}\\
&= \beta^n\frac{P_{s^0}\left(s^t,x_{t+1},\ldots,x_{t+n}\right)}{P_{s^0}\left(s^t\right)}\\
&= \beta^n P_{s^t}\left(x_{t+1},\ldots,x_{t+n}\right).
\end{aligned}$$

This proves the direction of the proposition going from Bayesian updating of the predictive distributions to the MRS condition.

Now, for the other direction, assume the MRS condition (7) holds. Combining the MRS condition and Proposition 2 yields

$$P_{s^t}\left(x_{t+1},\ldots,x_{t+n}\right)=\eta_{s^t}\left(\left\{s^t,x_{t+1},\ldots,x_{t+n}\right\}\right)$$

for each $n\geq1,t,s^t,x_{t+1},\ldots,x_{t+n}$ where $\eta_{s^t}(B)\equiv\frac{\eta\left(B\cap s^t\right)}{\eta(s^t)}$, $\quad\forall s^t\in\mathcal{S},\forall B\in\Sigma$. Therefore, by Lemma 3,

$$P_{s^t}(B)=\eta_{s^t}(B),\qquad\forall s^t\in\mathcal{S},\forall B\in\Sigma.$$

Observe that

$$P_{s^t}(B)=\eta_{s^t}(B)=\frac{\eta\left(B\cap s^t\right)}{\eta\left(s^t\right)}=\frac{P_{s^0}\left(B\cap s^t\right)}{P_{s^0}\left(s^t\right)}.$$

This proves Bayes' rule holds for the predictive distributions.. ■

Proof of Corollary 1. Assume the existence of the sets $B_1^{s^t}, \ldots, B_{m-1}^{s^t}$ in the statement of the full rank condition for each s^t. From Theorem 4, the MRS condition (7) is equivalent to Bayesian updating for predictive distributions. Therefore, for each s^t,

$$A^{s^t} \begin{bmatrix} \mu_{s^t}(\theta_1) \\ \mu_{s^t}(\theta_2) \\ \vdots \\ \mu_{s^t}(\theta_m) \end{bmatrix} = \begin{bmatrix} 1 \\ P_{s^t}(B_1^{s^t}) \\ \vdots \\ P_{s^t}(B_{m-1}^{s^t}) \end{bmatrix}. \tag{25}$$

Observe that Equation (25) has a unique solution μ_{s^t} since A^{s^t} is of full rank. Calculation shows that $\mu_{s^t}(\theta_n) = \frac{\mu(\theta_n)\pi_{\theta_n}(s^t)}{P_{s^0}(s^t)} = \frac{\mu(\theta_n)\pi_{\theta_n}(s^t)}{\int_\Theta \pi_\theta(s^t)d\mu}$ for $n = 1, \ldots, m$ is a solution, and thus the unique solution. This proves the MRS condition implies Bayesian updating of the μ_{s^t}.

For the other direction, assume $\mu_{s^t}(\theta_n) = \frac{\mu(\theta_n)\pi_{\theta_n}(s^t)}{\int_\Theta \pi_\theta(s^t)d\mu}$ for $n = 1, \ldots, m$. For any $B \in \Sigma$,

$$\begin{aligned} P_{s^t}(B) &= \int_\Theta \pi_\theta(B \mid s^t)\, d\mu_{s^t} = \sum_{n=1}^m \frac{\pi_{\theta_n}(B \cap s^t)}{\pi_{\theta_n}(s^t)} \mu_{s^t}(\theta_n) \\ &= \sum_{n=1}^m \frac{\pi_{\theta_n}(B \cap s^t)}{\pi_{\theta_n}(s^t)} \frac{\mu(\theta_n)\pi_{\theta_n}(s^t)}{\sum_{l=1}^m \pi_{\theta_l}(s^t)\mu(\theta_l)} \\ &= \frac{\sum_{n=1}^m \pi_{\theta_n}(B \cap s^t)\mu(\theta_n)}{\sum_{n=1}^m \pi_{\theta_n}(s^t)\mu(\theta_n)} \\ &= \frac{\int_\Theta \pi_\theta(B \cap s^t)\, d\mu}{\int_\Theta \pi_\theta(s^t)\, d\mu} = \frac{P_{s^0}(B \cap s^t)}{P_{s^0}(s^t)}. \end{aligned}$$

Thus, the predictive distributions are updated by Bayes' rule on $B \in \Sigma$, which, by Theorem 4, implies the MRS condition (7).. ■

Proof of Proposition 4. Fix any s^t. Since $\mathcal{X}$ and Θ are finite, let $|\mathcal{X}| = n$ and $|\Theta| = m$ and index their elements as $y_1, \ldots y_n$ and $\theta_1, \ldots, \theta_m$, respectively with $y_1 = x_t$ (if $s^t = s^0$ adopt the convention that $x_t = x^0$). We use q_{ijk} to denote $q_{\theta_k}(y_i, y_j)$. Our maintained assumption that each elementary cylinder set is given positive weight by each π_{θ_k} implies $0 < q_{ijk} < 1$ for all i, j, k.

Claim. Given any s^t, there is an elementary cylinder $\{s^t, x_{t+1}, \ldots, x_{t+\tau}\}$, with $x_t = x_{t+1} = x_{t+\tau}$, such that

$$\pi_{\theta'}\left(\{s^t, x_{t+1}, \ldots, x_{t+\tau}\} \mid s^t\right) \neq \pi_{\theta''}\left(\{s^t, x_{t+1}, \ldots, x_{t+\tau}\} \mid s^t\right), \qquad \forall \theta', \theta'' \in \Theta. \tag{26}$$

Proof of the Claim. Observe that any $\{x_{t+1}, \ldots, x_{t+\tau}\}$ can be identified with a unique $n \times n$ matrix $W = (w_{ij})$ of non-negative integers, where w_{ij} is the number of times the pair (y_i, y_j) occurs in the sequence $\{x_{t+1}, \ldots, x_{t+\tau}\}$. In particular, for each $k \in \{1, \ldots m\}$,

$$\pi_{\theta_k}\left(\{s^t, x_{t+1}, \ldots, x_{t+\tau}\} \mid s^t\right) = q_{\theta_k}(x_t, x_{t+1}) \prod_{i=1}^\tau \prod_{j=1}^\tau q_{ijk}^{w_{ij}}.$$

Though it is not true that every such matrix W has an associated sequence $\{x_{t+1}, \ldots, x_{t+\tau}\}$, the following restrictions are sufficient to ensure it: W is symmetric, written $W \in Sym(n, n)$, and its all entries are strictly positive (i.e., $w_{ij} \geq 1$). This follows from a well known property of strongly connected directed multigraphs: there is a path traversing each edge exactly once and starting and ending at the same vertex (known as an Eulerian circuit) if and only if each vertex has the same number of outgoing edges as incoming edges (e.g., Theorem 1.4.24 in West (2001)). To relate this to the matrix, think of the directed multigraph with n vertices, corresponding to the n elements $y_1, \ldots y_n$, and with w_{ij} edges going from vertex i to vertex j for each i, j. Since $w_{ij} \geq 1$ this multigraph is strongly connected. Symmetry implies that each vertex has the same number of outgoing edges as incoming edges. Given an Eulerian circuit, one may start at any vertex and, by following the circuit, traverse every edge exactly once and end at that same vertex. If we start from the vertex corresponding to $x_t (= y_1)$, the $x_{t+1}, \ldots, x_{t+\tau}$ is simply an ordered list of the vertices visited along the path corresponding to the Eulerian circuit. Observe that without loss of generality we can set $x_t = x_{t+1} = x_{t+\tau}$.

We now show that there exists $\{s^t, x_{t+1}, \ldots, x_{t+\tau}\}$ with $x_t = x_{t+1} = x_{t+\tau}$ such that, for each k', k'' with $k' \neq k''$,

$$q_{\theta_{k'}}(x_t, x_{t+1}) \prod_{i=1}^{\tau} \prod_{j=1}^{\tau} q_{ijk'}^{w_{ij}} \neq q_{\theta_{k''}}(x_t, x_{t+1}) \prod_{i=1}^{\tau} \prod_{j=1}^{\tau} q_{ijk''}^{w_{ij}}. \tag{27}$$

Since $0 < q_{ijk} < 1$, (27) is equivalent to

$$\ln \overset{\square}{q}_{\theta_{k'}}(x_t, x_t)\big) + \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} w_{ij} \ln(q_{ijk'}) \neq \ln \overset{\square}{q}_{\theta_{k''}}(x_t, x_t)\big) + \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} w_{ij} \ln(q_{ijk''}).$$

Any $W \in Sym(n, n)$ that does not satisfy (27) must have, for some $k' \neq k''$,

$$\ln \overset{\square}{q}_{\theta_{k'}}(x_t, x_t)\big) \square \ln \overset{\square}{q}_{\theta_{k''}}(x_t, x_t)\big) + \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} w_{ij} \left(\ln(q_{ijk'}) \square \ln(q_{ijk''})\right) = 0. \tag{28}$$

Denote by $\square$ the, possibly empty, set of all $W \in Sym(n, n)$ that satisfy (28). It is convex and closed. Moreover, (28) can be written as,

$$w_{i^*i^*} = \frac{\ln \overset{\square}{q}_{\theta_{k''}}(x_t, x_t)\big) \square \ln \overset{\square}{q}_{\theta_{k'}}(x_t, x_t)\big)}{\ln(q_{i^*i^*k'}) \square \ln(q_{i^*i^*k''})} \square \frac{\sum_{i \neq i^*} \sum_{j=1}^{\tau} w_{ij} \left(\ln(q_{ijk'}) \square \ln(q_{ijk''})\right) + \sum_{j \neq i^*} w_{i^*j} \left(\ln(q_{i^*jk'}) \square \ln(q_{i^*jk''})\right)}{\ln(q_{i^*i^*k'}) \square \ln(q_{i^*i^*k''})}$$

where i^* is chosen so that $q_{i^*i^*k'} \neq q_{i^*i^*k''}$. No redundancy in Θ with respect to the transition functions on the diagonal implies such an i^* exists (notice that diagonal elements are not restricted by the symmetry requirement). Therefore, $\square$ has dimension at most $2^{\square 1} n(n+1) \square 1$, where $\dim(Sym(n, n)) = 2^{\square 1} n(n+1)$. We conclude that $\square$ is a nowhere dense subset of $\mathbb{R}^{\frac{n(n+1)}{2}}$.

Since there are $\binom{m}{2}$ distinct pairs k', k'', the set of all $W \in Sym(n,n)$ that solve (28) for at least one pair is the union of $\binom{m}{2}$ nowhere dense subsets of $\mathbb{R}^{\frac{n(n+1)}{2}}$. Hence, it is still a nowhere dense subset of $\mathbb{R}^{\frac{n(n+1)}{2}}$, and so there exists a symmetric W with strictly positive rational entries such that (27) is satisfied for all distinct pairs k', k''. Multiplying this matrix by the product of the denominators, M, and adding $M - 1$ to w_{11} yields an integer valued $W \in Sym(n,n)$ also satisfying (27) since both sides of (27) are taken to the power M. Any cylinder $\{s^t, x_{t+1}, \ldots, x_{t+\tau}\}$ with $x_t = x_{t+1} = x_{t+\tau}$ associated with such a matrix is assigned distinct $\pi_{\theta_k}(\{s^t, x_{t+1}, \ldots, x_{t+\tau}\} \mid s^t)$, as desired. □

Let $\{s^t, \hat{x}_{t+1}, \ldots, \hat{x}_{t+\tau}\}$ be the elementary cylinder set that satisfies (26), with $x_t = \hat{x}_{t+1} = \hat{x}_{t+\tau}$. Setting $B_1^{s^t} = \{s^t, \hat{x}_{t+1}, \ldots, \hat{x}_{t+\tau}\}$, $B_2^{s^t} = \{s^t, \hat{x}_{t+1}, \ldots, \hat{x}_{t+\tau}, \hat{x}_{t+1}, \ldots, \hat{x}_{t+\tau}\}$, and so on, delivers A^{s^t} of the form

$$\begin{bmatrix} 1 & \ldots & 1 \\ c_1 & \ldots & c_m \\ c_1^2 & \ldots & c_m^2 \\ \vdots & \vdots & \vdots \\ c_1^{m-1} & \ldots & c_m^{m-1} \end{bmatrix}$$

where $c_k = q_{\theta_k}(x_t, x_t) \prod_{i=1}^{\tau} \prod_{j=1}^{\tau} q_{ijk}^{w_{ij}} \in (0,1)$. Given an n-tuple $\{\alpha_i\}_{i=1}^m \subseteq \mathbb{R}^m$, suppose $\sum_{i=0}^{m-1} \alpha_i c_j^i = 0$ for all $j = 1, ..., m$. If some of the α_i are not zero, then the equation $\sum_{i=0}^{m-1} \alpha_i x_j^i = 0$ has degree at least 1 and at most $m - 1$, and so it has at most $m - 1$ solutions. But, $\{c_j\}_{j=1}^m$ is a set of m distinct solutions of the this equation, a contradiction. We conclude that all α_i must be zero. Hence, the transpose of A^{s^t} (and so A^{s^t} itself) has full rank. ■

The following Bayesian consistency result is a special case of that shown by Lijoi, Prunster, and Walker (2007).

Lemma 4 *Suppose that each π_θ makes the process $\{X_t\}_{t \in \mathcal{T}}$ stationary and ergodic. Given any $\theta_0 \in \operatorname{supp}(\mu)$, we have*

$$\lim_{t \to \infty} \mu(\theta \mid x_1, ..., x_t) = \delta_{\theta_0}(\theta), \qquad \pi_{\theta_0} \quad a.e.,$$

for all $\theta \in \Theta$.

Proof of Proposition 5. By Lemma 4, π_{θ_0} - *a.e.* we have

$$\lim_{t \to \infty} |\mu(\theta \mid x_1, ..., x_t) - \delta_{\theta_0}(\theta)| = 0 \qquad \forall \theta \in \Theta. \tag{29}$$

Let S^* be the subset of S, with $\pi_{\theta_0}(S^*) = 1$, for which (29) holds. Fix a path $(x_1, ..., x_t, ...) \in S^*$. Now s^t will denote the time t node reached on the fixed path.

Since Θ is finite, given any $\varepsilon > 0$ there is $t_\varepsilon \geq 1$ such that for all $t \geq t_\varepsilon$ we have:

$$\max_{\theta \in \Theta} |\mu(\theta \mid x_1, ..., x_t) - \delta_{\theta_0}(\theta)| \leq \varepsilon.$$

As ϕ^{-1} is Lipschitz, there is $k > 0$ such that

$$\left|\phi^{-1}(x) - \phi^{-1}(y)\right| \leq k |x - y| \qquad \forall x, y \in \phi(\mathcal{U}).$$

Hence, for all $t \geq t_\varepsilon$ we have:

$$\begin{aligned}
&\left| V_{s^t}(f) \square \left[u \overset{\square}{f} s^t) \right) + \beta \int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_{\theta_0} \overset{\square}{x}_{t+1}; s^t) \right] \right| \\
= \;& \beta \left| \phi^{\square 1} \sum_{\theta \in \Theta} \phi \left(\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_\theta \overset{\square}{x}_{t+1}; s^t) \right) \mu \overset{\square}{\theta} \mid s^t) \right) \square \int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_{\theta_0} \overset{\square}{x}_{t+1}; s^t) \right| \\
\leq \;& \beta k \left| \sum_{\theta \in \Theta} \phi \left(\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_\theta \overset{\square}{x}_{t+1}; s^t) \right) \mu \overset{\square}{\theta} \mid s^t) \square \phi \left(\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_{\theta_0} \overset{\square}{x}_{t+1}; s^t) \right) \right| \\
= \;& \beta k \left| \sum_{\theta \in \Theta} \phi \left(\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_\theta \overset{\square}{x}_{t+1}; s^t) \right) \overset{\square}{\mu} \overset{\square}{\theta} \mid s^t) \square \delta_{\theta_0}(\theta) \right) \right| \\
\leq \;& \beta k \sum_{\theta \in \Theta} \left| \phi \left(\int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_\theta \overset{\square}{x}_{t+1}; s^t) \right) \right| \left| \mu \overset{\square}{\theta} \mid s^t) \square \delta_{\theta_0}(\theta) \right| \\
\leq \;& \beta k \sum_{\theta \in \Theta} \left(\max \left\{ \left| \phi(u(M)) \right|, \left| \phi(u(m)) \right| \right\} \right) \varepsilon = \beta k \left(\max \left\{ \left| \phi(u(M)) \right|, \left| \phi(u(m)) \right| \right\} \right) |\Theta| \, \varepsilon.
\end{aligned}$$

As ε was arbitrary, we conclude that for the fixed path $(x_1, ..., x_t, ...) \in S^*$ it holds

$$\lim_{t \to \infty} \left| V_{s^t}(f) \square u \overset{\square}{f} s^t) \right) \square \beta \int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_{\theta_0} \overset{\square}{x}_{t+1}; s^t) \right| = 0.$$

Since the path $(x_1, ..., x_t, ...) \in S^*$ was arbitrary and $\pi_0(S^*) = 1$, in turn this implies the result.

Observe that t_ε is only a function of μ and θ_0, and in particular is not a function of f. Moreover, the bound on the difference

$$\left| V_{s^t}(f) \square \left[u \overset{\square}{f} s^t) \right) + \beta \int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_{\theta_0} \overset{\square}{x}_{t+1}; s^t) \right] \right|$$

derived above is also independent of f. Therefore this bound holds for all f simultaneously. Thus, the difference converges to zero uniformly over f. Specifically, for any $\varepsilon > 0$, there exists $t_\varepsilon \geq 1$ such that for all $t \geq t_\varepsilon$, and all plans f,

$$\begin{aligned}
&\left| V_{s^t}(f) \square \left[u \overset{\square}{f} s^t) \right) + \beta \int_{\mathcal{X}_{t+1}} V_{(s^t, x_{t+1})}(f)\, d\pi_{\theta_0} \overset{\square}{x}_{t+1}; s^t) \right] \right| \\
\leq \;& \beta k \left(\max \left\{ \left| \phi(u(M)) \right|, \left| \phi(u(m)) \right| \right\} \right) |\Theta| \, \varepsilon.
\end{aligned}$$

■

10 Appendix B: Recursive Equations and Proofs for Section 4.2

In this Appendix we will present some results on recursive equations that will imply the results of Section 4.2, as shown at the end of this Appendix.

10.1 Functions

Consider the set $B(\mathcal{S})$ of all bounded real-valued functions on $\mathcal{S}$, endowed with the supnorm $\|\cdot\|_\infty$. The pair $(B(\mathcal{S}), \|\cdot\|_\infty)$ is a Banach space.

Given a subset I of $\mathbb{R}$, set $B(\mathcal{S}, I) = \{\varphi \in B(\mathcal{S}) : \varphi(s) \in I \text{ for all } s \in \mathcal{S}\}$. When $I = [0, \infty)$, we just write $B_+(\mathcal{S})$. If I is closed, the pair $(B(\mathcal{S}, I), d_\infty)$ is a complete metric space, where d_∞ is the supnorm metric.

10.2 Existence

Given $\in B(\mathcal{S})$ and $\beta \in (0,1)$, set

$$I = \left[\frac{\inf_{s\in\mathcal{S}} (s)}{1 \square \beta}, \frac{\sup_{s\in\mathcal{S}} (s)}{1 \square \beta}\right].$$

Observe that $I \subseteq \mathbb{R}_+$ if and only if $\in B_+(\mathcal{S})$.

Consider the recursive equation:

$$\varphi \overset{\square}{s^t}) = \overset{\square}{s^t}) + \beta\phi^{\square 1}\left(\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \varphi \overset{\square}{s^t}, x_{t+1})\, d\pi_\theta \overset{\square}{x_{t+1}}; s^t)\right) d\mu_{s^t}\right), \tag{30}$$

where $\varphi \in B(\mathcal{S}, I)$, $\phi : I \to \mathbb{R}_+$ is a strictly monotone function, and $\mu_{s^t} : 2^\Theta \to [0,1]$ is any probability measure on the power set of Θ at node s^t.

Our purpose is to show that (30) has a solution, possibly unique. In fact, if we set $= u \circ f$, then any solution of (30) would be a solution for (3).

In order to solve (30), we need to consider the operator T on $B(\mathcal{S}, I)$ given by

$$T(\varphi) \overset{\square}{s^t}) = \overset{\square}{s^t}) + \beta\phi^{\square 1}\left(\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \varphi \overset{\square}{s^t}, x_{t+1})\, d\pi_\theta \overset{\square}{x_{t+1}}; s^t)\right) d\mu_{s^t}\right) \tag{31}$$

Lemma 5 *We have $T(\varphi) \in B(\mathcal{S}, I)$ whenever $\varphi \in B(\mathcal{S}, I)$.*

Proof. Since ϕ is strictly monotone, T is non-decreasing, i.e., $T(\varphi_1) \leq T(\varphi_2)$ if $\varphi_1 \leq \varphi_2$. Consider the functions $\varphi_*, \varphi^* \in B(\mathcal{S}, I)$ respectively given by $\varphi_*(s) = \frac{\inf_{s\in\mathcal{S}} (s)}{1\square\beta}$ and $\varphi^*(s) = \frac{\sup_{s\in\mathcal{S}} (s)}{1\square\beta}$ for all $s \in \mathcal{S}$. By the monotonicity of T, we have

$$\begin{aligned}
\frac{\inf_{s\in\mathcal{S}} (s)}{1 \square \beta} &= \inf_{s\in\mathcal{S}} (s) + \beta\frac{\inf_{s\in\mathcal{S}} (s)}{1 \square \beta} \leq T(\varphi_*) \leq T(\varphi) \\
&\leq T(\varphi^*) \leq \sup_{s\in\mathcal{S}} (s) + \beta\frac{\sup_{s\in\mathcal{S}} (s)}{1 \square \beta} = \frac{\sup_{s\in\mathcal{S}} (s)}{1 \square \beta}
\end{aligned}$$

for all $\varphi \in B(\mathcal{S}, I)$. Hence, $T(\varphi) \in B(\mathcal{S}, I)$. ■

By Lemma 5, we can write $T : B(\mathcal{S}, I) \to B(\mathcal{S}, I)$. Moreover, it is easy to see that the recursive equation (30) has a solution if and only if T has a fixed point.

We begin with a general existence result for fixed points.

Proposition 6 *The operator $T : B(\mathcal{S}, I) \to B(\mathcal{S}, I)$ given by (31) has a fixed point provided $\in B(\mathcal{S})$ and $\phi : I \to \mathbb{R}$ is strictly monotone.*

Proof. Observe that $B(\mathcal{S}, I)$ is a complete lattice with respect to the pointwise order $\leq$. In fact, given any subset E of $B(\mathcal{S}, I)$, define $\varphi^*, \varphi_* \in B(\mathcal{S}, I)$ by $\varphi^*(s) = \sup_{\varphi \in E} \varphi(s)$ and $\varphi_*(s) = \inf_{\varphi \in E} \varphi(s)$ for each $s \in S$, respectively. It is easy to check that φ^* and φ_* are, respectively, the least upper bound and the greatest lower bound of E under the order $\leq$.

Since the operator $T : B(\mathcal{S}, I) \to B(\mathcal{S}, I)$ is monotone, by the Tarski Fixed Point Theorem (see Tarski (1955, Thm 1)), we conclude that T has a fixed point. ■

10.3 Uniqueness and Monotonicity

Proposition 7 *The operator $T : B(\mathcal{S}, I) \to B(\mathcal{S}, I)$ given by (31) has a unique fixed point provided is eventually deterministic, i.e., for each $s^{t'} \in \mathcal{S}$ there exist a node $s^{t''} \in \mathcal{S}$ with $s^{t''} \geq s^{t'}$ and a function $\eta_{s^{t''}} : \mathbb{N} \to \mathbb{R}$ such that $(s^t) = \eta_{s^{t''}}(t)$ for all $s^t \in \mathcal{S}$ with $s^t \geq s^{t''}$. In this case, such fixed point is such that $\varphi(s^t) = \sum_{\tau \geq 0} \beta^\tau \eta_{s^{t''}}(t + \tau)$ for all $s^t \in \mathcal{S}$ with $s^t \geq s^{t''}$.*

Proof. Let $s^{t''} \in \mathcal{S}$ and $\eta_{s^{t''}} : \mathbb{N} \to \mathbb{R}$ such that $(s^t) = \eta_{s^{t''}}(t)$ for all $s^t \in \mathcal{S}$ with $s^t \geq s^{t''}$. By hypothesis,

$$T(\varphi)\left(s^t\right) = \eta_{s^{t''}}(t) + \beta \phi^{\square 1}\left(\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \varphi\left(s^t, x_{t+1}\right) d\pi_\theta\left(x_{t+1}; s^t\right)\right) d\mu_{s^t}\right),$$

for all $s^t \in \mathcal{S}$ with $s^t \geq s^{t''}$, and it is easy to check that there exists a fixed point $\widehat{\varphi}$ such that $\widehat{\varphi}(s^t) = \sum_{\tau \geq 0} \beta^\tau \eta_{s^{t''}}(t + \tau)$ for all $s^t \in \mathcal{S}$ with $s^t \geq s^{t''}$.[15]

Suppose there exists a fixed point $\varphi \in B(\mathcal{S}, I)$ with $\varphi \neq \widehat{\varphi}$ on all $s^t \in \mathcal{S}$ with $s^t \geq s^{t''}$. Then, there exists $s^t \geq s^{t''}$ such that $\varphi(s^t) \neq \sum_{\tau \geq 0} \beta^\tau \eta_{s^{t''}}(t + \tau)$. As φ is a fixed point, we have

$$\varphi\left(s^t\right) = \eta_{s^{t''}}(t) + \beta \phi^{\square 1}\left(\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \varphi\left(s^t, x_{t+1}\right) d\pi_\theta\left(x_{t+1}; s^t\right)\right) d\mu_{s^t}\right),$$

and so

$$\phi\left(\frac{\varphi(s^t) \square \eta_{s^{t''}}(t)}{\beta}\right) = \int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \varphi\left(s^t, x_{t+1}\right) d\pi_\theta\left(x_{t+1}; s^t\right)\right) d\mu_{s^t}.$$

Hence, there exist $x'_{t+1}, x''_{t+1} \in \mathcal{X}_{t+1}$ such that

$$\varphi\left(s^t, x'_{t+1}\right) \geq \frac{\varphi(s^t) \square \eta_{s^{t''}}(t)}{\beta} \geq \varphi\left(s^t, x''_{t+1}\right).$$

On the other hand, for each $x_{t+1} \in \mathcal{X}_{t+1}$ we have

$$\begin{aligned} & \varphi\left(s^t, x_{t+1}\right) \\ = \; & \eta_{s^{t''}}(t+1) + \beta \phi^{\square 1}\left(\int_\Theta \phi\left(\int_{\mathcal{X}_{t+2}} \varphi\left(s^t, x_{t+1}, x_{t+2}\right) d\pi_\theta\left(x_{t+2}; s^t, x_{t+1}\right)\right) d\mu_{(s^t, x_{t+1})}\right), \end{aligned}$$

[15] Since is bounded, also the function $\eta_{s^{t''}}$ is bounded. Hence, the series $\sum_{\tau \geq 0} \beta^{t+\tau} \eta_{s^{t''}}(t + \tau)$ is absolutely convergent.

and so a similar argument shows that there exist $x'_{t+2}, x''_{t+2} \in \mathcal{X}_{t+2}$ such that

$$\varphi\left(s^t, x_{t+1}, x'_{t+2}\right) \geq \frac{\varphi\left(s^t, x_{t+1}\right) \square \eta_{s^{t''}}\left(t+1\right)}{\beta} \geq \varphi\left(s^t, x_{t+1}, x''_{t+2}\right).$$

Hence,

$$\varphi\left(s^t, x'_{t+1}, x'_{t+2}\right) \geq \frac{\varphi\left(s^t\right) \square \eta_{s^{t''}}\left(t\right) \square \beta\eta_{s^{t''}}\left(t+1\right)}{\beta^2} \geq \varphi\left(s^t, x''_{t+1}, x''_{t+2}\right).$$

By proceeding in this way, we get

$$\begin{aligned} \varphi\left(s^t, x'_{t+1}, ..., x'_{t+n}\right) &\geq \frac{\varphi\left(s^t\right) \square \sum_{\tau=0}^{n \square 1} \beta^\tau \eta_{s^{t''}}\left(t+\tau\right)}{\beta^n} \\ &\geq \varphi\left(s^t, x''_{t+1}, ..., x''_{t+n}\right), \end{aligned} \tag{32}$$

If $\varphi\left(s^t\right) > \sum_{\tau \geq 0} \beta^\tau \eta_{s^{t''}}\left(t+\tau\right)$, then

$$\lim_n \frac{\varphi\left(s^t\right) \square \sum_{\tau=0}^{n \square 1} \beta^\tau \eta_{s^{t''}}\left(t+\tau\right)}{\beta^n} = +\infty,$$

while

$$\lim_n \frac{\varphi\left(s^t\right) \square \sum_{\tau=0}^{n \square 1} \beta^\tau \eta_{s^{t''}}\left(t+\tau\right)}{\beta^n} = \square\infty$$

if $\varphi\left(s^t\right) < \sum_{\tau \geq 0} \beta^\tau \eta_{s^{t''}}\left(t+\tau\right)$. In both cases, (32) leads to a contradiction because φ is bounded. We conclude that $\varphi\left(s^t\right) = \sum_{\tau \geq 0} \beta^\tau \eta_{s^{t''}}\left(t+\tau\right)$ for all $s^t \in \mathcal{S}$ with $s^t \geq s^{t''}$. ■

Next we refine the existence result of the previous section by providing sufficient conditions under which T is a contraction, and so it has a unique fixed point.

Theorem 5 *The operator $T : B\left(\mathcal{S}, I\right) \to B\left(\mathcal{S}, I\right)$ given by (31) has a unique fixed point provided $\in B_+\left(\mathcal{S}\right)$ and $\phi : \mathbb{R}_+ \to \mathbb{R}$ is twice differentiable on $\mathbb{R}_{++}$ and satisfies one of the following three conditions:*

(i) ϕ is IAAA;

(ii) ϕ is IRAA and $\inf_{s \in S} \quad (s) > 0$;

(iii) ϕ is SDAAA, concave and $\sup_{x \geq 0} \quad (x) < 1$.

Proof. The theorem and the proposition immediately below follow from collecting several uniqueness and global attractivity results proven in Marinacci and Montrucchio (2007). ■

Under conditions (i), (ii) or (iii), the unique fixed point can be found via a contraction argument, and this implies the following property useful for iteratively approaching the solution:

Proposition 8 *Suppose conditions (i), (ii) or (iii) of Theorem 5 hold. Given any initial condition $\varphi_0 \in B(\mathcal{S}, I)$, the sequence $T^n(\varphi)$ of iterates uniformly converges to φ^*, i.e., $\|T^n(\varphi) \square \varphi^*\|_\infty \to 0$.*

The final result we need shows that a unique fixed point must be monotonic in the sense we used in the main text. We proceed by first proving a key consequentialist property holds and then proving the monotonicity result.

Lemma 6 *Let A be a subset of $B(\mathcal{S})$. Suppose the operator $T : B(\mathcal{S}, I) \to B(\mathcal{S}, I)$ given by (31) has a unique fixed point, φ , for each $\in A$. Then, for any $_1, _2 \in A$, $_1(s) = _2(s)$ for all $s \geq s^t$ implies $\varphi_{_1}(s^t) = \varphi_{_2}(s^t)$.*

Proof. Fix s^t and suppose to the contrary that $_1(s) = _2(s)$ for all $s \geq s^t$ but $\varphi_{_1}(s^t) \neq \varphi_{_2}(s^t)$. Now construct $\hat{\varphi}_{_1} \in B(\mathcal{S}, I)$ as follows:

$$\hat{\varphi}_{_1}(s) = \begin{cases} \varphi_{_1}(s) & \text{if} \quad s \not\geqslant s^t \text{ and } s^t \not\geqslant s \\ \varphi_{_2}(s) & \text{if} \quad s \geqslant s^t \\ \vartheta(s) & \text{if} \quad s^t \geqslant s \text{ and } s \neq s^t \end{cases}$$

where

$$\vartheta\left(s^{t'}\right) \equiv {}_1\left(s^{t'}\right) + \beta\phi^{\square 1}\left(\int_\Theta \phi \left(\int_{\mathcal{X}_{t'+1}} \hat{\varphi}_{_1}\left(s^{t'}, x_{t'+1}\right) d\pi_\theta\left(x_{t'+1}; s^{t'}\right)\right) d\mu_{s^{t'}}\right),$$

for $s^t > s^{t'}$ are the unique values determined by the recursion from the already specified values of $\hat{\varphi}_{_1}$. By construction, $\hat{\varphi}_{_1}(s)$ is a fixed point of T for $_1$, but $\hat{\varphi}_{_1}(s^t) \neq \varphi_{_1}(s^t)$ contradicting the assumed uniqueness. ■

Proposition 9 *Let A be a subset of $B(\mathcal{S})$. Suppose the operator $T : B(\mathcal{S}, I) \to B(\mathcal{S}, I)$ given by (31) has a unique fixed point, φ , for each $\in A$. Then, for any $_1, _2 \in A$, $_1(s) \geq _2(s)$ for all $s \geq s^t$ implies $\varphi_{_1}(s^t) \geq \varphi_{_2}(s^t)$. If, in addition, $_1$ and $_2$ are deterministic and $_1(s) > _2(s)$ for at least some $s \geq s^t$, then $\varphi_{_1}(s^t) > \varphi_{_2}(s^t)$.*

Proof. For deterministic (i.e., such that the value at a node s depends only on the time at which the node occurs), one may verify that $\varphi(s^t) = \sum_{\tau \geq 0} \beta^\tau (s^{t+\tau})$ is a fixed point, and, therefore, given the assumption of uniqueness, must be the unique fixed point. Thus, for deterministic $_1$ and $_2$ with $_1(s) \geq _2(s)$ for all $s \geq s^t$ and $_1(s) > _2(s)$ for at least some $s \geq s^t$, $\varphi_{_1}(s^t) > \varphi_{_2}(s^t)$. More generally, consider $_1, _2 \in A$ and the associated unique fixed points $\varphi_{_1}$ and $\varphi_{_2}$. Suppose that for some node s^t, $_1(s) \geq _2(s)$ for all $s \geq s^t$. If $_1(s) = _2(s)$ for all $s \geq s^t$, the result follows from Lemma 6. Suppose then that $_1(s) > _2(s)$ for at least one $s \geq s^t$. $\mathcal{S}$ is countable, so let $\mathcal{D} = \{s_1, s_2, \ldots\}$ be an enumeration of the successors of s^t where $_1(s) > _2(s)$. We may approach $_1$ from $_2$ on $s \geq s^t$ through a countable number of steps by changing the value at one node at a time. That is, construct a sequence, $\{\omega_n\}$, such that $\omega_0 = _2$ and, for $n > 0$, ω_n has the same

value as $\omega_{n\square 1}$ at all but the node s_n, with $\omega_n(s_n) = {}_1(s_n)$. Note that if we start from a for which T has a unique fixed point and change the value of at only one node, there will still be a unique fixed point (by recursion backward from the time of the change). Let $\{\varphi_{\omega_n}\}$ be the associated sequence of unique fixed points. Observe that since $\omega_n = \omega_{n\square 1}$ at all nodes except s_n, Lemma 6 implies $\varphi_{\omega_n} = \varphi_{\omega_{n\square 1}}$ at all strict successors of s_n and at all successors of s^t that are neither predecessors nor successors of s_n. Therefore, at s_n,

$$\begin{aligned}
\varphi_{\omega_n}(s_n) &= {}_1(s_n) + \beta\phi^{\square 1}\left(\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \varphi_{\omega_{n\square 1}}(s_n, x_{t+1})\, d\pi_\theta(x_{t+1}; s_n)\right) d\mu_{s_n}\right) \\
&> {}_2(s_n) + \beta\phi^{\square 1}\left(\int_\Theta \phi\left(\int_{\mathcal{X}_{t+1}} \varphi_{\omega_{n\square 1}}(s_n, x_{t+1})\, d\pi_\theta(x_{t+1}; s_n)\right) d\mu_{s_n}\right) \\
&= \varphi_{\omega_{n\square 1}}(s_n).
\end{aligned}$$

Thus, working backwards by recursion, $\varphi_{\omega_n}(s) > \varphi_{\omega_{n\square 1}}(s)$ for all successors of s^t that are predecessors of s_n (the inequality is strict since all finite paths were assumed to have positive probability). In particular, $\varphi_{\omega_n}(s^t) > \varphi_{\omega_{n\square 1}}(s^t)$. This holds for all n in the sequence, so $\varphi_{{}_2}(s^t) = \varphi_{\omega_0}(s^t) < \varphi_{\omega_n}(s^t)$ for all $n > 0$. Since $\lim_{n\to\infty} \varphi_{\omega_n}(s^t) \leq \varphi_{{}_1}(s^t)$, we have $\varphi_{{}_2}(s^t) \leq \varphi_{{}_1}(s^t)$. ■

10.4 Proofs for Section 4.2

Proof of Theorem 2. Follows from Proposition 6.. ■

Proof of Proposition 3. Follows from Proposition 7 and Proposition 9. ■

Proof of Theorem 3. Follows from Theorem 5 and Proposition 9. ■

References

Abel, A. (1990): "Asset Prices under Habit Formation and Catching up with the Joneses," *The American Economic Review*, 80(2), 38–42.

Backus, D., B. Routledge, and S. Zin (2004): "Exotic Preferences for Macroeconomists," Working paper 10597, NBER.

Camerer, C. (1999): "Ambiguity-Aversion and Non-Additive Probability: Experimental Evidence, Models and Applications," in *Uncertain Decisions: Bridging Theory and Experiments*, ed. by L. Luini, pp. 53–80. Kluwer Academic Publishers.

Chen, Z., and L. G. Epstein (2002): "Ambiguity, Risk and Asset Returns in Continuous Time," *Econometrica*, 70(4), 1403–1443.

Durrett, R. (1991): *Probability: Theory and Examples*. Wadsworth and Brooks, Pacific Grove.

Eichberger, J., S. Grant, and D. Kelsey (2005): "CEU Preferences and Dynamic Consistency," *Mathematical Social Sciences*, 49, 143–151.

Epstein, L. G. (1983): "Stationary cardinal utility and optimal growth under uncertainty," *Journal of Economic Theory*, 31, 133–152.

Epstein, L. G., and M. LeBreton (1993): "Dynamically consistent beliefs must be bayesian," *Journal of Economic Theory*, 61(1), 1–22.

Epstein, L. G., and M. Schneider (2003a): "IID: independently and indistinguishably distributed," *Journal of Economic Theory*, 113(1), 32–50.

——— (2003b): "Recursive Multiple-Priors," *Journal of Economic Theory*, 113(1), 1–31.

Epstein, L. G., and T. Wang (1994): "Intertemporal Asset Pricing under Knightian Uncertainty," *Econometrica*, 62(3), 283–322.

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

Ergin, H., and F. Gul (2004): "A Subjective Theory of Compound Lotteries," mimeo., Princeton University.

Ghirardato, P. (2002): "Revisiting Savage in a Conditional World," *Economic Theory*, 20, 83–92.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with a Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

Gollier, C. (2005): "Does ambiguity aversion reinforce risk aversion? Applications to portfolio choices and asset prices," mimeo., University of Toulouse.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

Hanany, E., and P. Klibanoff (2007a): "Updating Ambiguity Averse Preferences," mimeo., Northwestern Univ.

——— (2007b): "Updating Preferences with Multiple Priors," *Theoretical Economics*, 2(3), 261–298.

Hansen, L., and T. Sargent (2006): "Recursive Robust Estimation and Control Without Commitment," mimeo., University of Chicago.

Hansen, L. P. (2007): "Beliefs, Doubts and Learning: The Valuation of Macroeconomic Risk," *American Economic Review*, 97(2), 1–30.

Hansen, L. P., and T. Sargent (2001): "Robust control and model uncertainty," *American Economic Review*, 91, 60–66.

Hayashi, T. (2005): "Intertemporal substitution, risk aversion and ambiguity aversion," *Economic Theory*, 25, 933–956.

Jewitt, I., and S. Mukerji (2006): "Ordering Ambiguous Acts," mimeo., Oxford University.

Johnsen, T. H., and J. B. Donaldson (1985): "The Structure of Intertemporal Preferences under Uncertainty and Time Consistent Plans," *Econometrica*, 53(6), 1451–58.

Ju, N., and J. Miao (2007): "Ambiguity, Learning, and Asset Returns," mimeo., Hong Kong University of Science and Technology.

Karni, E., and D. Schmeidler (1991): "Atemporal Dynamic Consistency and Expected Utility Theory," *Journal of Economic Theory*, 54, 401–408.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005): "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73(6), 1849–1892.

Kreps, D., and E. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46(1), 185–200.

Lijoi, A., I. Prunster, and S. Walker (2007): "Bayesian Consistency for Stationary Models," *Econometric Theory*, 23, 749–759.

Lucas, R. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429–1445.

Maccheroni, F., and M. Marinacci (2005): "A Strong Law of Large Numbers for Capacities," *Annals of Probability*, 33, 1171–1178.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006a): "Ambiguity Aversion, Robustness and the Variational Representation of Preferences," *Econometrica*, 74(6), 1447–1498.

——— (2006b): "Dynamic Variational Preferences," *Journal of Economic Theory*, 128, 4–44.

Machina, M. (1989): "Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty," *Journal of Economic Literature*, 27, 1622–1668.

Machina, M., and D. Schmeidler (1992): "A More Robust Definition of Subjective Probability," *Econometrica*, 60, 745–780.

Marinacci, M., and L. Montrucchio (2007): "Unique Solutions of Some Recursive Equations in Economic Dynamics," mimeo., Collegio Carlo Alberto.

Nau, R. (2006): "Uncertainty Aversion with Second-Order Utilities and Probabilities," *Management Science*, 52(1), 136–145.

Neilson, W. S. (1993): "Ambiguity Aversion: An Axiomatic Approach Using Second Order Probabilities," mimeo., Texas A&M University.

Nishimura, K., and H. Ozaki (2003): "A Simple Axiomatization of Iterated Choquet Objectives," mimeo., University of Tokyo.

Pollak, R. (1970): "Habit Formation and Dynamic Demand Functions," *The Journal of Political Economy*, 78(4), 745–763.

Sarin, R., and P. Wakker (1998): "Dynamic Choice and Nonexpected Utility," *J. of Risk and Uncertainty*, 17, 87–119.

Schmeidler, D. (1989): "Subjective Probability and Expected Utility Without Additivity," *Econometrica*, 57(3), 571–587.

Segal, U. (1987): "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach," *International Economic Review*, 28, 175–202.

Seidenfeld, T., and L. Wasserman (1993): "Dilation for sets of probabilities," *Annals of Statistics*, 21, 1139–1154.

Seo, K. (2006): "Ambiguity and Second-Order Belief," mimeo., University of Rochester.

Siniscalchi, M. (2004): "Dynamic choice under ambiguity," mimeo., Northwestern Univ.

Strotz, R. (1955-6): "Myopia and inconsistency in dynamic utility maximization," *Review of Economic Studies*, 23, 165–180.

Tarski, A. (1955): "A lattice-theoretical fixpoint theorem and its applications," *Pacific Journal of Mathematics*, 5, 285–309.

Wakker, P. (1989): *Additive representations of preferences, a new foundation of decision analysis*. Kluwer (Academic Publishers), Dordrecht.

Wang, T. (2003): "Conditional Preferences and Updating," *Journal of Economic Theory*, 108, 286–321.

West, D. B. (2001): *Introduction to Graph Theory*. Prentice-Hall, New Jersey, 2nd edn.