Learning How to Invest when Returns are Uncertain

Thomas A. Knox
Harvard University

November 20, 2002*

©2002 by Thomas A. Knox

Abstract

Most asset returns are uncertain, not merely risky: investors do not know the probabilities of different future returns. A large body of evidence suggests that investors are averse to uncertainty, as well as to risk. A separate, but similarly substantial, set of facts indicates that investors try to learn from the past. This paper analyzes the dynamic portfolio and consumption choices of an uncertainty-averse (as well as risk-averse) investor who tries to learn from historical data. Because an adequate general theory of uncertainty-averse dynamic choice does not exist, I build one up from axioms on preferences. I pay special attention to the existence and consistency of conditional preferences. My theory generalizes the major dynamic uncertainty-aversion models in the asset pricing literature, and reveals their underlying similarities. Working directly with preferences, I give natural conditions leading to learning. I then present analytical solutions to the continuous-time portfolio and consumption problems of two uncertainty-averse investors: one who dogmatically expects the worst, and another who has a preference for learning.

*I am indebted to John Campbell and Gary Chamberlain for guidance and encouragement. I am also grateful for the insights of Brian Hall, Parag Pathak, Jeremy Stein, and James Stock, and for the helpful comments of seminar participants at Harvard University and MIT.

1 Introduction

Most asset returns are uncertain, not merely risky: investors do not know the probabilities of different future returns. An aversion to uncertainty, or a preference for bets with known odds, has been used in recent studies, including Anderson, Hansen, and Sargent (2000), Chen and Epstein (2002), and Maenhout (2001), to explain the equity premium puzzle. However, as Maenhout (2001) notes, these explanations often rely on investors ignoring data and dogmatically expecting the worst. Whether uncertainty aversion remains a plausible explanation for the equity premium puzzle when learning is accounted for is an open question. Further, a debate is ongoing between the recursive multiple-priors school on one side and the robust control school on the other about how best to model uncertainty aversion. Models of learning in portfolio choice have also received significant attention recently from authors including Barberis (2000), Brennan (1998), Brennan and Xia (2001), Kandel and Stambaugh (1996), and Xia (2001). However, these papers do not account for uncertainty aversion. It is not clear that their results will continue to hold when uncertainty aversion is incorporated, since uncertainty aversion leads to behavior compatible with "extreme" priors, as shown below.

This paper takes the first steps toward resolving these issues: first, I give a simple example showing that the differences between the recursive multiple-priors and robust control schools are much smaller than they have been perceived to be by showing that the uncertainty-aversion models of the two schools have the same basic structure. I also show that the differences between the recursive multiple-priors approach and a modification of the robust control approach (a discrete-time version of the Maenhout (2001) modification) are smaller still. Next, I build up a theory of *model-based multiple-priors* as an alternative to both the recursive multiple-priors and robust control approaches, starting from axioms on preferences. I show that the model-based multiple-priors approach can incorporate learning and uncertainty aversion; this theoretical development further elucidates the basic similarities between the recursive multiple-priors and robust control schools. Finally, I consider, and solve in closed form, the canonical intertemporal consumption and portfolio choice problem of a power-utility investor under two different sorts of uncertainty aversion. Under the first sort of uncertainty aversion, learning never occurs: the investor dogmatically expects the worst. The portfolio and consumption decisions of this investor are structurally identical to those of the investors studied by Chen and Epstein (2002) and Maenhout (2001), which shows that the investors studied by these researchers are, behaviorally, also dogmatic. The second sort of uncertainty aversion, however, yields much more interesting and realistic preferences: under this sort of uncertainty aversion, the investor has a preference for learning, and attempts to use historical data in order to learn how to invest more effectively.

My consideration of uncertainty aversion is not motivated solely by the fact that uncertainty aversion is a plausible explanation for many financial phenomena: uncertainty aversion is also strongly supported in the experimental data. The most famous experimental demonstration of uncertainty aversion is the Ellsberg Paradox (Ellsberg (1961)): Consider two urns, each containing 100 balls. Each ball is either red or black. The first ("known") urn contains 50 red balls (and thus 50 black balls). The second ("ambiguous") urn contains an unknown number of red balls. One ball is drawn at random from each urn, and four bets based on the results of these draws are to be ranked; winning a bet results in a 100 dollar cash prize. The first bet is won if the ball drawn from the known urn is red; the second bet is won if the ball drawn from the known urn is black; the third bet is won if the ball drawn from the ambiguous urn is red; the fourth bet is won if the ball drawn from the ambiguous urn is black. Many decision makers are indifferent between the first and second bets and between the third and fourth bets, but strictly prefer either the first or second bet to either the third or fourth bet. Since no distribution on the number of red balls in the ambiguous urn can support these preferences through expected utility, this ranking of gambles violates the axioms of subjective expected utility theory.

An alternative experiment, with less built-in symmetry, could be performed using an ordinary tack. The idea of flipping tacks comes from Chapter 11 of Kreps (1988). The outcome in which the tack balances on its point and an edge of its head could be termed "tails" and the outcome in which it lands with its point up, flat on its head could be termed "heads." One could imagine asking a decision maker what she expects the probability of heads to be; call this expectation p. One could then introduce a (continuous) roulette wheel that pays off one dollar if the pointer on the wheel lands in an interval whose length is a fraction p of the circumference. If $p > \frac{1}{2}$ one could ask, "Would you rather bet a dollar that the outcome of a tack flip is heads, or would you rather gamble using the roulette wheel?" If $p \leq \frac{1}{2}$, one could ask the same question, substituting "tails" for "heads." If the decision maker prefers to gamble using the roulette wheel, which has known odds, then she is uncertainty averse. This is not nearly as clean as the Ellsberg experiment, but may be more related to real-world decisions, in that there is no explicit symmetry in the problem.

In a seminal response to the Ellsberg Paradox, Gilboa and Schmeidler (1989) provided an axiomatic foundation to support uncertainty aversion in static choice. They weakened the independence axiom and showed that, under this weakening, preferences could be represented by the minimum expected utility over a set of (prior) distributions. For example, an agent with Gilboa-Schmeidler preferences would exhibit Ellsberg-type behavior if the set of distributions on the number of red balls in the unknown urn included a distribution under which black balls are more numerous and one under which red balls are more numerous. The optimal choice under these assumptions is that which maximizes (over possible choices) the minimum (over the set of distributions) expected utility, leading to the label "maxmin expected utility."

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

The recursive multiple-priors approach began nonaxiomatically in discrete time with the work of Epstein and Wang (1994). In Chen and Epstein (2002) the approach was brought into a continuous-time framework, the portfolio choice problem was considered generally and solved analytically in some cases, and the separate effects of "ambiguity" (uncertainty) and risk were shown in equilibrium. The recursive multiple-priors approach was given axiomatic foundations in Epstein and Schneider (2001), and an attempt was made by Epstein and Schneider (2002) to incorporate learning. Despite the obvious importance of this strand of the literature, I will argue that Epstein and Schneider (2002) do not present a satisfactory model of learning under uncertainty.

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

The debate between the recursive multiple-priors school and the robust control school has focused on the "constraint preferences" generated by robust control. However, in applications the robust control school typically uses the "multiplier preferences" generated by robust control, which are acknowledged by both schools to differ from the constraint preferences (though they are observationally equivalent to the constraint preferences in any single problem; see Epstein and Schneider (2001) and Hansen, Sargent, Turmuhambetova, and Williams (2001)). I show that the set of priors generating the multiplier preferences is rectangular, so that the sets of priors being used in the recursive multiple-priors approach and in the robust control multiplier approach (implicitly, in the case of the robust control multiplier approach) have the same structure. By the results of Epstein and Schneider (2001) (which are themselves implied by my results), this implies that the multiplier preferences generated by robust control are dynamically consistent in the very narrow sense used by Epstein and Schneider (2001). Furthermore, I show that the set of priors implicit in a modification of the robust control multiplier method (a discrete-time version of the Maenhout (2001) modification) is even more closely linked to the set of priors used in recursive multiple-priors.

In the course of developing a theory sufficiently general to accommodate learning and uncertainty aversion, I am forced to axiomatize consistent conditional preferences in general. My characterization of necessary and sufficient conditions for the existence of sets of consistent conditional preferences will allow me to get at the roots of each of the major methods in the literature: Chamberlain's, the recursive multiple-prior school's, and the robust control school's.

The continuing debate regarding how to extend atemporal maxmin expected utility to intertemporal situations is essentially a debate about the structure of the set of prior distributions with which expected utility is evaluated. This paper shows the implications of the existence of consistent conditional preferences for the set of distributions used to represent utility. In this connection, a novel *restricted independence axiom* is introduced, and a class of distributions, termed *prismatic*, is characterized. The set of prismatic distributions is a strict superset of the set of rectangular distributions introduced by Epstein and Schneider (2001) (see below), and allows conditioning of a more general sort.

It is also shown that one cannot maintain uncertainty aversion in the face of too much consistent conditioning: if a sufficiently rich set of consistent conditional preferences exist, then the independence axiom must hold and the set of distributions must be a singleton. This is crucial in evaluating the usefulness of models that assume a certain set of consistent conditional preferences (such as those of Epstein and Schneider (2001)). If the consistent conditional preferences assumed are not those of interest to the researcher, it may be necessary to abandon them in order to maintain uncertainty aversion and have the consistent conditional preferences of interest.

Finally, I consider the consumption and portfolio choice problem of an investor who is averse to uncertainty, but believes there is something to be learned from the past. The investor has power utility over a stream of intermediate consumption. I solve this problem in closed form. Combining learning with uncertainty aversion in this way is of particular interest in dynamic portfolio choice problems, which have attracted much research on learning and uncertainty aversion separately. Anderson, Hansen, and Sargent (2000), Chen and Epstein (2002), Dow and Werlang (1992), Epstein and Schneider (2002), Epstein and

Wang (1994), Maenhout (2001), Liu, Pan, and Wang (2002), Pathak (2002), Uppal and Wang (2002), and Wang (2002) apply uncertainty-averse decision theories to study portfolio choice and some of its implications, while Barberis (2000), Brennan (1998), Brennan and Xia (2001), Kandel and Stambaugh (1996), and Xia (2001) (to mention just a few of the many researchers working productively in this area) examine portfolio choice with Bayesian learning. Only Epstein and Schneider (2002) and Miao (2001) attempt to include both learning and uncertainty aversion. Epstein and Schneider (2002) freely admit that their method is *ad hoc* and not founded upon any coherent set of principles. Miao (2001) attempts, again without the benefit of any coherent set of principles, to layer the κ-ignorance specification of Chen and Epstein (2002) on top of a standard (single-prior) model of Bayesian portfolio choice in continuous time. This implies that in Miao (2001), any learning which takes place is learning *not* about an uncertain variable, but about a variable the investor regards as merely risky (since there is a single prior in the underlying Bayesian model). Further, both Epstein and Schneider (2002) and Miao (2001) make use of the recursive multiple-priors method; it will be shown below that this method implies a very severe sort of nonstationarity, which may be unattractive in a model of learning.

The remainder of the paper proceeds as follows. Section 2 shows the similarities between the recursive multiple-priors approach and the robust control approach (as well as the Maenhout (2001) modification), and how they differ from model-based multiple-priors, in a two-period model with a risky (and uncertain) asset with only two possible returns. This binomial model is obviously oversimplified, and is presented only in order to clarify the basics of the three approaches; I present more realistic models later in the paper. In Section 3, I set forth the most general domain of preferences, that is, the set of bets or gambles over which the decision maker is to choose. Section 4 presents the axioms that I will use to formalize the decision maker's preference structure. Section 5 delivers the basic results concerning the existence of consistent conditional preferences, their link to the restricted independence axiom, and the associated shape of the set of priors used by the decision maker. It also provides an important negative result stating that a sufficiently rich set of consistent conditional preferences implies the full independence axiom, and thus implies that there is no uncertainty aversion. I view this result as cautionary, rather than discouraging. Section 6 shows how my basic results are easily adapted to a dynamic domain which includes consumption at various points in time. Section 7 puts my results in perspective by showing that they can be used to classify and understand the major existing models of uncertainty-averse decision-making. Section 8 solves two intertemporal portfolio and consumption choice problems. Working in a continuous-time setting, I analyze first a benchmark model of an investor who dogmatically expects the worst, which I show is related to the models of Chen and Epstein (2002) and Maenhout (2001) in its implications, and then a model in which the investor learns from historical data. Section 9 concludes. All proofs are placed in a separate appendix.

2 A Simple Two-period Example

In this section, I use an extremely simple model to illustrate the basic differences between the recursive multiple-priors approach, the robust control approach, and the model-based multiple-priors approach. There are two periods: $t = 0, 1$. Investment decisions are made at the beginning of each period, and the return for each period is realized at the end of that period. This corresponds to the event tree depicted in Figure 1. Initial wealth is $W_0 > 0$. Utility is of the power form over final wealth:

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

where W_2 denotes wealth at the *end* of period 1 (or the beginning of period 2). In the continuous-time models analyzed later in the paper, I consider intermediate consumption; I could include intermediate consumption here, but I omit it for the sake of simplicity. There is one riskless asset, and the (gross) riskless rate is denoted $R_f > 1$. There is one risky, and uncertain, asset; in each period, the gross return on this asset takes on one of two possible values. I denote the higher of these two values by H and the lower by L, and denote the (uncertain) gross return on the risky asset in period t by R_t. In order to avoid arbitrage, I assume that $H > R_f > L$. At each of the two periods, the investor chooses how much to invest in the risky (and uncertain) asset.

As a benchmark, I consider the standard approach to portfolio choice (as laid out by Ingersoll (1987)) in this context. In the standard approach, the next step would be to assume that there is some probability $p \in (0,1)$ such that

$$\Pr(R_0 = H, R_1 = H) = p^2 \tag{2}$$

$$\Pr(R_0 = H, R_1 = L) = p(1-p) \tag{3}$$

$$\Pr(R_0 = L, R_1 = H) = (1-p)p \tag{4}$$

$$\Pr(R_0 = L, R_1 = L) = (1-p)^2. \tag{5}$$

This assumption really has two pieces: first, the probability of each of the possible sequences of returns is known, and second, the return at time 1 is generated independently of the return at time 0, but from the identical distribution.

To solve the investor's choice problem in the standard setting of the previous paragraph, I will use dynamic programming. I denote by $J(W,t)$ the indirect utility of wealth W at the beginning of period $t = 0, 1$ (that is, J is the value function). Let φ_t denote the portfolio weight placed on the risky asset at the beginning of period $t = 0, 1$. The investor's problem can then be expressed as:

$$J(W,1) = \max_{\varphi} \left\{ \frac{W^{1-\gamma}}{1-\gamma} \left[\begin{array}{c} p(\varphi(H-R_f)+R_f)^{1-\gamma} \\ +(1-p)(\varphi(L-R_f)+R_f)^{1-\gamma} \end{array} \right] \right\} \tag{6}$$

$$J(W,0) = \max_{\varphi} \left\{ \begin{array}{c} pJ(W(\varphi(H-R_f)+R_f),1) \\ +(1-p)J(W(\varphi(L-R_f)+R_f),1) \end{array} \right\}. \tag{7}$$

The solution to this problem is given by Ingersoll (1987) (page 125); using a superscript "o" to denote optimality,

$$\text{for } t=0,1,\ \varphi_t^o \;=\; R_f \frac{(1-p)^{-\frac{1}{\gamma}}(R_f-L)^{-\frac{1}{\gamma}} - p^{-\frac{1}{\gamma}}(H-R_f)^{-\frac{1}{\gamma}}}{(1-p)^{-\frac{1}{\gamma}}(R_f-L)^{1-\frac{1}{\gamma}} + p^{-\frac{1}{\gamma}}(H-R_f)^{1-\frac{1}{\gamma}}}. \tag{8}$$

I raise the question, "From where does the investor obtain p?" The standard approach outlined above is, I feel, neither normatively nor descriptively satisfying, and I thus drop the key assumption embodied in equations (2) through (5). Indeed, I do not even make the more general assumption that each possible pair of returns has a known probability (which would relax the i.i.d. assumption above). The classical "subjective expected utility" theories of choice under uncertainty, as pioneered by Savage, show that there are conditions on the investor's preferences that, if true, guarantee that the investor has a prior probability distribution on the four possible pairs of returns. However, as initially pointed out by Ellsberg (1961), these theories cannot accommodate aversion to uncertainty, as distinct from aversion to risk. Aversion to uncertainty certainly seems descriptively significant, and I would argue that it also has normative appeal. For this reason, I do not accept the subjective expected utility approach: I seek an approach that incorporates uncertainty aversion.

Under the conditions on preferences given by Gilboa and Schmeidler (1989), the investor has a *set* of (subjective) prior probability distributions on the four possible pairs of returns. This set is closed and convex. The investor evaluates any potential portfolio choice by calculating the *minimum* expected utility of that portfolio choice, where the minimum is taken over the set of priors. The recursive multiple-priors approach, the robust control approach, and the model-based multiple-priors approach all attempt to put some economically-founded structure on this set of priors.

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

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

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

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

This set of priors is shown in Figure 2. One can recognize this as an extreme form of nonstationarity: it is essentially stating that the investor views the uncertainties in the return at time 0, the return at time 1 given a high return at time 0, and the return at time 1 given a low return at time 0 as mutually unrelated. The "κ-ignorance" specification of Chen and Epstein (2002) further specializes the above to sets in which $\underline{p} = \underline{p}^H = \underline{p}^L$ and $\overline{p} = \overline{p}^H = \overline{p}^L$, so that the three returns have the same ranges of uncertainty, but their uncertainties remain unrelated.

In particular, the recursive multiple-priors approach is incompatible with the usual Bayesian method of placing a prior on p in the standard i.i.d. model, and updating that prior as information arrives. This incompatibility is somewhat unsettling; one might prefer a generalization or relaxation of the Bayesian method over an incompatible alternative to it.

It is instructive to examine the dynamic programming problem of an investor whose preferences satisfy the recursive multiple-priors conditions. I denote by $J(W, t)$ the indirect utility of wealth W at the beginning of period $t = 0, 1$. Then the investor solves:

$$\begin{aligned} & J\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right),1\right) \\ & = \max_{\varphi_1} \min_{p\in\left[\underline{p}^H,\overline{p}^H\right]} \left\{ \begin{array}{c} \frac{\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right)\right)^{1-\gamma}}{1-\gamma} \times \\ \left[p\left(\varphi_1\left(H-R_f\right)+R_f\right)^{1-\gamma}+\left(1-p\right)\left(\varphi_1\left(L-R_f\right)+R_f\right)^{1-\gamma}\right] \end{array} \right\} \end{aligned} \tag{12}$$

$$\begin{aligned} & J\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right),1\right) \\ & = \max_{\varphi_1} \min_{p\in\left[\underline{p}^L,\overline{p}^L\right]} \left\{ \begin{array}{c} \frac{\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right)\right)^{1-\gamma}}{1-\gamma} \times \\ \left[p\left(\varphi_1\left(H-R_f\right)+R_f\right)^{1-\gamma}+\left(1-p\right)\left(\varphi_1\left(L-R_f\right)+R_f\right)^{1-\gamma}\right] \end{array} \right\} \end{aligned} \tag{13}$$

$$\begin{aligned} & J\left(W_0,0\right) \\ = \; & \max_{\varphi_0} \min_{p\in\left[\underline{p},\overline{p}\right]} \left\{pJ\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right),1\right)+\left(1-p\right)J\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right),1\right)\right\}. \end{aligned} \tag{14}$$

This shows in a more direct way the nonstationarity implicit in the recursive multiple-priors approach. In each of the three subproblems above, the investor behaves as if there were a separate asset whose return uncertainty is described by an interval of probabilities. Crucially, the *theory* behind the recursive multiple-priors approach does *not* give any guidance on how the interval endpoints of the three intervals should be related to each other.

A specific set of derivative assets can reveal some counterintuitive behavior on the part of a recursive multiple-priors investor: consider one derivative, A, on the risky asset that pays off 1,000 dollars at the end of period 1 if the return sequence is (H, L) and otherwise pays off one dollar at the end of period 1, and another derivative, B, that pays off 1,000 dollars at the end of period 1 if the return sequence is (L, H) and otherwise pays off one dollar at the end of period 1. A and B are essentially bets on the order of the high and the low return, if one high and one low return are realized. One might expect that an investor would be indifferent between holding a portfolio of only A and a holding a portfolio of only B. Indeed, it would seem reasonable that the investor would be indifferent between A, B, and flipping a fair coin, then holding a portfolio of only A if the coin came up heads, and only B if the coin came up tails. However, if $\overline{p} > \underline{p}$, so that the investor has uncertainty aversion over the time-0 return, then a recursive multiple-priors investor cannot be indifferent between A, B, and randomizing over A and B with equal probabilities.

In the robust control approach, one first specifies a baseline model. In the current setting, it is natural to suppose that the baseline model might be of the i.i.d. form specified in equations (2) through (5). I do so, denoting by p the probability of a high return in the baseline model (in which returns are i.i.d.). I shall discuss only the "multiplier preferences" arising in robust control, not the "constraint preferences," since the multiplier preferences are the ones emphasized by the robust control school. To implement the robust control multiplier approach, one uses dynamic programming. I fix a penalty parameter θ, and denote by $J(W,t)$ the indirect utility of wealth W at the beginning of period $t=0,1$. Then

$$J(W,1) = \max_{\varphi} \min_{g\in[0,1]} \left\{ \begin{array}{c} \frac{W^{1-\gamma}}{1-\gamma} \left[\begin{array}{c} g\left(\varphi\left(H-R_f\right)+R_f\right)^{1-\gamma} \\ +\left(1-g\right)\left(\varphi\left(L-R_f\right)+R_f\right)^{1-\gamma} \end{array} \right] \\ +\theta\left(g\ln\left(\frac{g}{p}\right)+\left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\} \tag{15}$$

$$J(W,0) = \max_{\varphi} \min_{g\in[0,1]} \left\{ \begin{array}{c} gJ\left(W\left(\varphi\left(H-R_f\right)+R_f\right),1\right) \\ +\left(1-g\right)J\left(W\left(\varphi\left(L-R_f\right)+R_f\right),1\right) \\ +\theta\left(g\ln\left(\frac{g}{p}\right)+\left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\}. \tag{16}$$

A key characteristic of the robust control multiplier method is that the multiplier θ remains constant over both states of the world and time. Apply the minimax theorem to switch the order of minimization and maximization above (see Ferguson (1967), page 85; the theorem's technical conditions are satisfied if one rules out portfolio choices that risk ruin):

$$J(W,1) = \min_{g\in[0,1]} \max_{\varphi} \left\{ \begin{array}{c} \frac{W^{1-\gamma}}{1-\gamma} \left[\begin{array}{c} g\left(\varphi\left(H-R_f\right)+R_f\right)^{1-\gamma} \\ +\left(1-g\right)\left(\varphi\left(L-R_f\right)+R_f\right)^{1-\gamma} \end{array} \right] \\ +\theta\left(g\ln\left(\frac{g}{p}\right)+\left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\} \tag{17}$$

$$J(W,0) = \min_{g\in[0,1]} \max_{\varphi} \left\{ \begin{array}{c} gJ\left(W\left(\varphi\left(H-R_f\right)+R_f\right),1\right) \\ +\left(1-g\right)J\left(W\left(\varphi\left(L-R_f\right)+R_f\right),1\right) \\ +\theta\left(g\ln\left(\frac{g}{p}\right)+\left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\}. \tag{18}$$

I now observe that the Lagrange multiplier theorem, which Hansen, Sargent, Turmuhambetova, and Williams (2001) use to motivate switching from the constraint approach to the multiplier approach, may also be employed to link the multiplier approach to a *different* set of constraints. From the above, I can apply the Lagrange multiplier theorem to each of the three problems that the investor may face: how to invest at the beginning of time 0, how to invest at the beginning of time 1 if the time-0 return was low, and how to invest at the beginning of time 1 if the time-0 return was high. Doing so results in:

$$\begin{aligned} & J\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right),1\right) \\ = & \min_{g\in[0,1],g\ln\left(\frac{g}{p}\right)+(1-g)\ln\left(\frac{1-g}{1-p}\right)\leq\eta_1\left(W_1^H\right)} \max_{\varphi_1} \\ & \left\{ \begin{array}{c} \frac{\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right)\right)^{1-\gamma}}{1-\gamma}\times \\ \left[g\left(\varphi_1\left(H-R_f\right)+R_f\right)^{1-\gamma}+\left(1-g\right)\left(\varphi_1\left(L-R_f\right)+R_f\right)^{1-\gamma}\right] \end{array} \right\} \\ & +\theta\eta_1\left(W_1^H\right) \end{aligned} \tag{19}$$

$$
\begin{aligned}
& J\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right),1\right) \qquad (20)\\
= & \min_{g\in[0,1],g\ln\left(\frac{g}{p}\right)+(1-g)\ln\left(\frac{1-g}{1-p}\right)\leq\eta_1\left(W_1^L\right)} \max_{\varphi_1} \\
& \left\{ \begin{array}{c} \frac{\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right)\right)^{1-\gamma}}{1-\gamma}\times \\ \left[g\left(\varphi_1\left(H-R_f\right)+R_f\right)^{1-\gamma}+(1-g)\left(\varphi_1\left(L-R_f\right)+R_f\right)^{1-\gamma}\right] \end{array} \right\} \\
& +\theta\eta_1\left(W_1^L\right)
\end{aligned}
$$

$$
\begin{aligned}
& J\left(W_0,0\right) \qquad (21)\\
= & \min_{g\in[0,1],g\ln\left(\frac{g}{p}\right)+(1-g)\ln\left(\frac{1-g}{1-p}\right)\leq\eta_0(W_0)} \max_{\varphi_0} \\
& \left\{gJ\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right),1\right)+(1-g)J\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right),1\right)\right\} \\
& +\theta\eta_0\left(W_0\right).
\end{aligned}
$$

These results follow from reasoning conceptually identical to that of Hansen, Sargent, Turmuhambetova, and Williams (2001), pages 10 to 12; note that the assumption required by those authors is not necessary here, since it has been observed above that the minimax theorem holds for this problem. In the display above, I have used the abbreviations $W_1^H \equiv W_0\left(\varphi_0\left(H-R_f\right)+R_f\right)$ and $W_1^L \equiv W_0\left(\varphi_0\left(L-R_f\right)+R_f\right)$. Observe that, in order to break the multiplier problem down into constraint problems, I was forced to introduce *different* constraints at each time *and* each level of wealth at the beginning of the period (so that constraints in the current period may depend on portfolio choices in past periods, though *not*, of course, on the portfolio choice in the current period). This is because the Lagrange multiplier theorem states only that each multiplier problem produces the same result as some constraint problem; it does not state that the constraints in each problem are the same, or that the constraints do not depend on wealth at the beginning of each period. In fact, it suggests that the constraints will *not* be the same, and *will* depend on the period and on wealth at the beginning of each period, since the shadow value of relaxing the constraint will differ across the problems and across different beginning-of-period wealth levels: it is more harmful to experience a reduction in investment opportunities in a low-wealth state of the world than in a high-wealth state. One can see, however, that the only differences between the constraint in period 1 after a high return in period 0 and the constraint in period 1 after a low return in period 0 will arise because of differences in wealth at the beginning of period 1 between these two possibilities. This follows because the two problems are identical except for their (potentially) different levels of beginning-of-period wealth.

One may now observe that the entropy constraint can be translated to an interval constraint, since the relative entropy function is convex in g and g is a scalar in each of the three problems (and for any level of wealth at the beginning of the period in each of the three problems). Either these do not bind, in which case one is left with the basic constraint $g \in [0,1]$ or they do bind, in which case the interval is a strict subset of $[0,1]$. In either case, one obtains:

$$\begin{aligned} & J\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right),1\right) \\ = & \min_{g\in\left[a_1\left(W_1^H\right),b_1\left(W_1^H\right)\right]} \max_{\varphi_1} \left\{ \begin{array}{c} \frac{\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right)\right)^{1-\gamma}}{1-\gamma}\times \\ \left[g\left(\varphi_1\left(H-R_f\right)+R_f\right)^{1-\gamma}+\left(1-g\right)\left(\varphi_1\left(L-R_f\right)+R_f\right)^{1-\gamma}\right] \end{array} \right\} \\ & +\theta\eta_1\left(W_1^H\right) \end{aligned} \tag{22}$$

$$\begin{aligned} & J\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right),1\right) \\ = & \min_{g\in\left[a_1\left(W_1^L\right),b_1\left(W_1^L\right)\right]} \max_{\varphi_1} \left\{ \begin{array}{c} \frac{\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right)\right)^{1-\gamma}}{1-\gamma}\times \\ \left[g\left(\varphi_1\left(H-R_f\right)+R_f\right)^{1-\gamma}+\left(1-g\right)\left(\varphi_1\left(L-R_f\right)+R_f\right)^{1-\gamma}\right] \end{array} \right\} \\ & +\theta\eta_1\left(W_1^L\right) \end{aligned} \tag{23}$$

$$\begin{aligned} & J\left(W_0,0\right) \\ = & \min_{g\in\left[a_0\left(W_0\right),b_0\left(W_0\right)\right]} \max_{\varphi_0} \left\{gJ\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right),1\right)+\left(1-g\right)J\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right),1\right)\right\} \\ & +\theta\eta_0\left(W_0\right). \end{aligned} \tag{24}$$

Note that the intervals are different at each time and for each level of wealth at the beginning of the period they apply to.

I may now apply the minimax theorem (Ferguson (1967), page 85) to each of the three subproblems faced by the investor. Note that the technical conditions of the theorem given by Ferguson (1967) are satisfied as long as one rules out obviously suboptimal investment decisions (one must exclude investment decisions which risk ruin). The result is:

$$\begin{aligned}
& J\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right),1\right) \\
= & \max_{\varphi_1} \min_{g\in\left[a_1\left(W_1^H\right),b_1\left(W_1^H\right)\right]} \left\{ \begin{array}{c} \frac{\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right)\right)^{1-\gamma}}{1-\gamma}\times \\ \left[g\left(\varphi_1\left(H-R_f\right)+R_f\right)^{1-\gamma}+\left(1-g\right)\left(\varphi_1\left(L-R_f\right)+R_f\right)^{1-\gamma}\right] \end{array} \right\} \\
& +\theta\eta_1\left(W_1^H\right)
\end{aligned} \tag{25}$$

$$\begin{aligned}
& J\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right),1\right) \\
= & \max_{\varphi_1} \min_{g\in\left[a_1\left(W_1^L\right),b_1\left(W_1^L\right)\right]} \left\{ \begin{array}{c} \frac{\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right)\right)^{1-\gamma}}{1-\gamma}\times \\ \left[g\left(\varphi_1\left(H-R_f\right)+R_f\right)^{1-\gamma}+\left(1-g\right)\left(\varphi_1\left(L-R_f\right)+R_f\right)^{1-\gamma}\right] \end{array} \right\} \\
& +\theta\eta_1\left(W_1^L\right)
\end{aligned} \tag{26}$$

$$\begin{aligned}
& J\left(W_0,0\right) \\
= & \max_{\varphi_0} \min_{g\in\left[a_0\left(W_0\right),b_0\left(W_0\right)\right]} \left\{gJ\left(W_0\left(\varphi_0\left(H-R_f\right)+R_f\right),1\right)+\left(1-g\right)J\left(W_0\left(\varphi_0\left(L-R_f\right)+R_f\right),1\right)\right\} \\
& +\theta\eta_0\left(W_0\right).
\end{aligned} \tag{27}$$

But comparing the above display to equations (12) through (14), one sees that this is similar to the problem faced by a recursive multiple-priors investor, who has rectangular priors; here, the set of priors is:

$$\Pr\left(R_0=H\right) \in \left[a_0\left(W_0\right),b_0\left(W_0\right)\right] \tag{28}$$

$$\Pr\left(R_1=H\middle| R_0=H\right) \in \left[a_1\left(W_1^H\right),b_1\left(W_1^H\right)\right] \tag{29}$$

$$\Pr\left(R_1=H\middle| R_0=L\right) \in \left[a_1\left(W_1^L\right),b_1\left(W_1^L\right)\right]. \tag{30}$$

The main difference in the results obtained by the recursive multiple-priors school and the robust control school can, then, be traced to differences in how the (continuous-time analogs of) the endpoints of the probability intervals are chosen. Indeed, the "κ-ignorance" specification of Chen and Epstein (2002) sets all of the upper endpoints to the same value and all of the lower endpoints to the same value, which will tend to generate results different from those of the robust-control multiplier preferences, in which the endpoints will, in general, vary over time, states of the world, and wealth levels.

The above analysis shows that if one includes wealth (along with the stock price) as a state variable, then the set of priors implicit in the robust control multiplier approach is, in fact, rectangular. Of course, the recursive multiple-priors sets of priors, which are rectangular when only the stock price is considered as a state variable, remain rectangular with this augmented set of state variables.

I now consider a discrete-time version of the modified multiplier robust control method proposed in continuous time by Maenhout (2001). I temporarily turn attention to an extension of the current example which may have more than two periods (but still possesses binary returns, power utility over terminal wealth, etc.), in order to demonstrate that the discrete-time version of the Maenhout (2001) transformation I propose is not limited to two-period problems. Solely for expositional ease, I assume that $\gamma > 1$ throughout this treatment of the modified multiplier robust control method. Although it is not entirely obvious how to translate the Maenhout (2001) transformation into discrete time, it turns out that the appropriate translation is as follows: rather than a constant, state-independent multiplier θ on relative entropy, consider the following state-dependent multiplier on the relative entropy in period $t+1$:

$$\theta\left(1-\gamma\right)J\left(W_t, t+1\right).$$

Note the offset in timing: the value function is evaluated at period $t+1$, but at the wealth level of period t.

The modified robust control investor's dynamic programming problem at period t is thus:

$$\begin{aligned} & J\left(W_t, t\right) \\ = \; & \max_{\varphi} \min_{g \in [0,1]} \left\{ \begin{array}{c} E_g\left[J\left(W_{t+1}, t+1\right)\right] \\ +\theta\left(1-\gamma\right) J\left(W_t, t+1\right)\left(g \ln\left(\frac{g}{p}\right) + \left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\}, \end{aligned} \tag{31}$$

where $E_g\left[\cdot\right]$ denotes the expectation with respect to the distribution under which $\Pr\left(R_{t+1} = H\right) = g$.

First use a backward induction argument to show that

$$\forall X > 0,\ \forall t \in \{1, \ldots, T\} \;\; J\left(X, t\right) = f\left(t\right)\frac{X^{1-\gamma}}{1-\gamma} \tag{32}$$

for some function f such that $f\left(t\right) > 0$. Since $J\left(X, T\right) = \frac{X^{1-\gamma}}{1-\gamma}$, the assertion clearly holds for $t = T$, with $f\left(T\right) = 1$. Now, suppose the assertion holds for $t = s+1, \ldots, T$. I will show that it holds for $t = s$. Return to the dynamic programming problem at time s:

$$\begin{aligned} & J\left(W_s, s\right) \\ = \; & \max_{\varphi} \min_{g \in [0,1]} \left\{ \begin{array}{c} E_g\left[J\left(W_{s+1}, s+1\right)\right] \\ +\theta\left(1-\gamma\right) J\left(W_s, s+1\right)\left(g \ln\left(\frac{g}{p}\right) + \left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\} && (33) \\ = \; & \max_{\varphi} \min_{g \in [0,1]} \left\{ \begin{array}{c} E_g\left[f\left(s+1\right)\frac{W_{s+1}^{1-\gamma}}{1-\gamma}\right] \\ +\theta\left(1-\gamma\right) f\left(s+1\right)\frac{W_s^{1-\gamma}}{1-\gamma}\left(g \ln\left(\frac{g}{p}\right) + \left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\} && (34) \\ = \; & f\left(s+1\right) W_s^{1-\gamma} \max_{\varphi} \min_{g \in [0,1]} \left\{ \begin{array}{c} \frac{1}{1-\gamma} E_g\left[\left(\frac{W_{s+1}}{W_s}\right)^{1-\gamma}\right] \\ +\theta\left(g \ln\left(\frac{g}{p}\right) + \left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\} && (35) \\ = \; & f\left(s+1\right) W_s^{1-\gamma} \max_{\varphi} \min_{g \in [0,1]} \left\{ \begin{array}{c} \frac{1}{1-\gamma} E_g\left[\left(\varphi\left(R_{s+1} - R_f\right) + R_f\right)^{1-\gamma}\right] \\ +\theta\left(g \ln\left(\frac{g}{p}\right) + \left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\} && (36) \\ = \; & f\left(s\right)\frac{W_s^{1-\gamma}}{1-\gamma}, && (37) \end{aligned}$$

where I define

$$f\left(s\right) = \left(1-\gamma\right) f\left(s+1\right) \max_{\varphi} \min_{g \in [0,1]} \left\{ \begin{array}{c} \frac{1}{1-\gamma} E_g\left[\left(\varphi\left(R_{s+1} - R_f\right) + R_f\right)^{1-\gamma}\right] \\ +\theta\left(g \ln\left(\frac{g}{p}\right) + \left(1-g\right)\ln\left(\frac{1-g}{1-p}\right)\right) \end{array} \right\}, \tag{38}$$

which is valid since W_s appears nowhere in the definition of $f(s)$, and, in fact, nothing in the definition of $f(s)$ depends on anything except the period, s. Note that the second equality in the display above follows from the inductive hypothesis regarding the structure of $J(X, s+1)$. Finally, $f(s) > 0$ must be verified. By the inductive hypothesis, $f(s+1) > 0$. By assumption, $\gamma > 1$. Thus $g = p$ results in the relative entropy term being zero, while the expectation term is (due to the leading $\frac{1}{1-\gamma}$) negative. Regardless of φ, taking the minimum over $g \in [0,1]$ can only result in a lower value, so the "maxmin" term is negative. However, $(1-\gamma) < 0$ for $\gamma > 1$, so the product of the two negative quantities is positive and $f(s) > 0$ (since $f(s+1) > 0$).

Since this logic holds for any $W_s > 0$, I have shown that

$$\forall X > 0,\ \forall t \in \{1, \ldots, T\} \quad J(X,t) \quad = \quad f(t) \frac{X^{1-\gamma}}{1-\gamma} \tag{39}$$

with $f(t) > 0$ as desired, and I am now able to characterize the set of priors implicit in this discrete-time version of the Maenhout (2001) modification of the multiplier robust control method. Return once more to the dynamic programming problem faced by the investor at some time t:

$$\begin{aligned} & J(W_t, t) \\ = \quad & \max_{\varphi} \min_{g \in [0,1]} \left\{ \begin{array}{c} E_g \left[J(W_{t+1}, t+1) \right] \\ +\theta (1-\gamma) J(W_t, t+1) \left(g \ln\left(\frac{g}{p}\right) + (1-g) \ln\left(\frac{1-g}{1-p}\right) \right) \end{array} \right\} && (40) \\ = \quad & f(t+1) W_t^{1-\gamma} \max_{\varphi} \min_{g \in [0,1]} \left\{ \begin{array}{l} \frac{1}{1-\gamma} E_g \left[\left(\varphi (R_{t+1} - R_f) + R_f \right)^{1-\gamma} \right] \\ +\theta \left(g \ln\left(\frac{g}{p}\right) + (1-g) \ln\left(\frac{1-g}{1-p}\right) \right) \end{array} \right\}, && (41) \end{aligned}$$

by following exactly the same steps that we followed in our inductive demonstration above, since we have shown that $J(X, t+1)$ has the necessary structure (at this stage it is no longer an inductive hypothesis, but instead is a proven fact).

Exactly as in the case of the (unmodified) multiplier robust control method, applying the minimax theorem and the Lagrange multiplier theorem, then the minimax theorem again transforms the multiplier "maxmin" problem above into a constrained "maxmin" form:

$$f\left(t+1\right)W_t^{1-\gamma}\left[\max_{\varphi}\min_{g\in[a,b]}\left\{\frac{1}{1-\gamma}E_g\left[\left(\varphi\left(R_{t+1}-R_f\right)+R_f\right)^{1-\gamma}\right]\right\}+\eta\theta\right]. \tag{42}$$

The crucial difference between this modified multiplier robust control method and the original, unmodified multiplier robust control method is that in the modified version, the interval endpoints a and b do not depend on the period or on wealth at the beginning of the period, and neither does the η term. This lack of dependence can be traced to the fact that nothing *inside* the modified multiplier maxmin problem depends on beginning-of-period wealth or on the period itself (functions of beginning-of-period wealth and of the period multiply the solution of the maxmin problem, but do not affect the optimized value or the optimizing controls in the problem), whereas in the original version, beginning-of-period wealth and the period appear inside the maxmin problem.

The fact that a and b are not dependent on the period, on beginning-of-period wealth, or on the stock price implies that the same a and b are the implicit interval endpoints in each period and for any wealth levels. This, in turn, implies not only that the sets of priors implicit in the modified multiplier robust control method are rectangular, but that the implicit sets of priors satisfy the even stronger κ-ignorance specification of Chen and Epstein (2002). Thus, the modified multiplier robust control method discussed here, which is a natural discrete-time analog of the continuous-time method of Maenhout (2001), is structurally identical to one of the most narrowly-defined recursive multiple-priors methods, the κ-ignorance specification. The link between the modified multiplier robust control method and the recursive multiple-priors method therefore seems even tighter than the link between the original multiplier robust control method and the recursive multiple-priors method.

In contrast to either the recursive multiple-priors approach or the robust control multiplier approach, which I have shown to be *structurally* equivalent (though their results may differ due to differences in intervals), model-based multiple-priors relies on a novel structure.

Consider first the classical model of equations (2) to (5). This model has only one unknown parameter, p. I could follow the Savage approach by placing a (subjective) prior distribution on p, but this would not account for uncertainty aversion.

Instead, model-based multiple-priors (relative to the classical model) is based on placing a *set* of prior distributions on the parameter p. The investment chosen is the one that results from maximizing (over portfolio weights) the minimum (over priors in the set) expected utility. The name "model-based multiple-priors," comes from the fact that the method places multiple priors on the parameters of an economic model. Depending on how the economic model is written, model-based multiple-priors may lead to learning; when multiple priors are placed on p in the standard i.i.d. model as discussed above, learning will certainly result if every prior in the set considered has full support. On the other hand, I might have begun with an economic model that, in contrast to the classical model, provided for a different probability of a high return at each node of the binomial tree. Then learning would not have resulted (since each period and each state of the world would have involved a new parameter); in fact, the set of priors would be rectangular. Thus, the model-based multiple-priors method leads to sets of priors whose structure is more general than that of rectangular priors, and this greater generality allows the model-based multiple-priors method to accomodate learning.

3 The General Domain of Preferences

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

Let S be a non-empty set of *states*, let Σ be an algebra of subsets of S, and define

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

4 Axioms

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

4.1 The Restricted Independence Axiom

My key, novel axiom is the *restricted independence axiom*. It is stated relative to a subset $A \in \Sigma$ (recall that Σ is the algebra of subsets of S with respect to which the functions in L_0 are measurable), and I thus refer to it as restricted independence relative to A, or $R(A)$ for compactness of notation.

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

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

Suppose two gambles give the same payoff for each state in some set of states. It seems, normatively, that changing each of the gambles in the same way on that set of states (so that they continue to agree on that set) should not reverse preferences between them. The restricted independence axiom (relative to the set on which the two gambles differ) goes slightly beyond this, by saying that preferences are still preserved if, in addition to being changed on their set of agreement as described above, each gamble is also mixed with the same roulette lottery (or gamble that does not depend on the state) on the set on which they disagree. Still, this seems quite plausible normatively. The descriptive usefulness of restricted independence is best judged from the examples given below.

I will typically be concerned with preferences that satisfy the restricted independence axiom relative to a collection of sets. I shall state that preferences satisfy $R(\mathcal{A})$ if $\mathcal{A} = \{A_1, \ldots, A_k\}$, and preferences satisfy $R(A_i)$ for each $i \in \{1, \ldots, k\}$.

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

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

$$(l;h)_i = \begin{cases} l & \text{for } s \in A_i, \\ h & \text{for } s \in A_i^C, \end{cases} \tag{43}$$

and

$$(q;h)_i = \begin{cases} q & \text{for } s \in A_i, \\ h & \text{for } s \in A_i^C. \end{cases} \tag{44}$$

Then

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

4.2 The Gilboa-Schmeidler Axioms

I group the axioms of Gilboa and Schmeidler (1989) together into the following axiom.

Axiom 3

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

The labels attached to each portion of the axiom are those used by Gilboa and Schmeidler (1989). Of these portions of the axiom, Weak Order, Continuity, Monotonicity, and Non-degeneracy are completely standard in the axiomatic literature on choice under uncertainty. The usual independence axiom is strictly stronger than (implies, and is not implied by) the Certainty Independence and Uncertainty Aversion portions of the axiom above. Descriptively, it seems more plausible that Certainty Independence would hold than full-blown independence: Decision makers may find it easier to work through the implications of mixing with a roulette lottery, which yields the same thing in every state, than to see the implications of mixing with a general act. Normatively, Certainty Independence seems appealing, but it seems much less clear that full independence is desirable.

These axioms are the standard "maxmin expected utility" axioms. They form the base from which I build up a theory of consistent conditional preferences.

4.3 The Consistent Conditioning Axiom

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

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

The axiom below formalizes the notion that conditional preferences should be genuinely conditional; that is, if two horse lotteries (elements of L_0) have identical payoffs on some set of states, then a preference relation conditional on the state being in that set ought to display indifference between the two horse lotteries. This property is known as *consequentialism*, and has been extensively investigated by Hammond (1988). There should also be some minimal link between conditional and unconditional preference relations: if every conditional preference relation in an exhaustive set displays a weak preference for one horse lottery (element of L_0) over another, then the unconditional preference relation ought to display a weak preference for the first horse lottery, too. If, in addition, one of the conditional preference relations, conditional on a set of states that is not null, displays a strict preference for the first horse lottery, then the unconditional preference relation ought to display a strict preference for the first horse lottery, too. This is a notion of *consistency*, and is part of the axiom below. Finally, the following axiom allows conditional preferences to display uncertainty aversion by assuming that each conditional preference relation satisfies an appropriate modification of Axiom 3.

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

for each $A_i \in \mathcal{A}$, and these conditional preference orderings satisfy:

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

Consistency: $f \succsim_i g \;\; \forall i \in \{1, \ldots, k\} \Rightarrow f \succsim g.$

If, in addition, $\exists A_i \in \mathcal{A}$ such that $f \succ_i g$ and A_i is not null, then $f \succ g$.

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

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

Of course, the appeal of consistency as a property of conditional preferences does not speak to the appeal of conditional preferences themselves, or, more precisely, to the appeal of the "consequentialism" section of the axiom above. Normatively, the attraction of certain sets of conditional (and consequentialist) preferences seems clear. For example, suppose one is offered a choice among bets on a horse race. One's preferences over bets are likely to be quite different depending on whether or not one is told which horse finished first; in fact, if one is told which horse finished first, one is likely not to care about what payoffs various bets offer when that horse does not finish first. However, it is important to note that the normative appeal of consequentialism can depend on what is being conditioned upon; for cases in which consequentialism may be much less attractive, see Machina (1989). This issue is revisited in Section 7.

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

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

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

5 General Results

5.1 The Relation of Restricted Independence to Consistent Conditional Preferences

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

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

5.2 Prismatic Sets of Priors

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

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

The name "prismatic" was chosen because of what sets of priors on four points look like when they satisfy the above condition. (A set of priors on *three* points that is prismatic with respect to a partition into one set of two elements and one set of a single element is a trapezoidal subset of the simplex on three points.) Suppose that a set of priors on four points is prismatic with respect to a partition into one set of two points and two sets, each of one point. Without loss of generality, suppose that the two singleton sets in the partition correspond to points 1 and 4, while the set of two points contains points 2 and 3. Label the probabilities $[q_1 \ \ q_2 p \ \ q_2(1-p) \ \ q_3]$, where the fact that the set of priors is prismatic with respect to the given partition implies that $p \in [\underline{p}, \overline{p}]$ and $[q_1 \ \ q_2 \ \ q_3] \in Q$ for some convex subset Q of the simplex on three points (so, because of the constraint that $q_1 + q_2 + q_3 = 1$, the set Q has dimension two). In three dimensions, plot q_1 on the x-axis, q_3 on the y-axis, and p on the z-axis. The solid formed will be a generalization of a right prism, in which the shapes of the parallel top and bottom "caps" (in this case, these are projections of the set Q onto the xy plane) are convex, but need not be polyhedral.

Now suppose that the set of priors on four points is prismatic with respect to a partition into two sets, each containing two elements (this might arise, for example, if the four points are formed from a two-period dynamic model with a binomial branching in each period). Without loss of generality, suppose that points 1 and 2 are in the first set, while points 3 and 4 are in the second. Label the probabilities $[qp_1 \;\; q(1-p_1) \;\; (1-q)\,p_2 \;\; (1-q)(1-p_2)]$, where the fact that the set of priors is prismatic with respect to the given partition implies that $q \in [\underline{q}, \overline{q}]$, $p_1 \in [\underline{p}_1, \overline{p}_1]$, and $p_2 \in [\underline{p}_2, \overline{p}_2]$. In three dimensions, plot p_1 on the x-axis, p_2 on the y-axis, and q on the z-axis. The solid formed will be a right prism; in fact, it will be a box.

Since, for a set of priors on four points, being prismatic with respect to a partition into two sets of two points is equivalent to being "rectangular" (Epstein and Schneider (2001)) in the two-period dynamic model with binomial branching mentioned above, a set of "rectangular" priors looks (as it should) like a box. This is an example of the way in which prismatic priors generalize rectangular priors.

5.3 The Basic Representation Result

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

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

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

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

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

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

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

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

This theorem obtains the representation that will be useful in applying my theory. The contribution I make here is specifying the shape of the set of distributions $\mathcal{P}$. Again, one sees a natural generalization of the situation explored by Epstein and Schneider (2001). The key to the shape of the set of distributions is that any conditional may be selected from the set $\mathcal{C}_i$, regardless of how other conditionals or the marginals are chosen. Intuitively, it is sensible that this "independence" in the selection of the distribution corresponds to the restricted independence of Axiom 1.

5.4 Restricted Independence and Full Independence

Definition 3 *Suppose that* $|S| < \infty$ *and* $\Sigma = 2^S$. *Then two partitions,* $\mathcal{A} = \{A_1, \ldots, A_k\}$ *and* $\mathcal{B} = \{B_1, \ldots, B_n\}$, *will be said to be* **interlaced** *if and only if:*

$$(1)\ \forall i \in \{1, \ldots, k\},\ \exists j_1, \ldots, j_{|A_i|} \in \{1, \ldots, n\}$$

$$\text{such that } B_{j_k} \cap A_i = \{x_{j_k}\} \text{ and } |B_{j_k}| = 1 \text{ for at most one } j_k$$

$$(2)\ \forall j \in \{1, \ldots, n\},\ \exists i_1, \ldots, i_{|B_j|} \in \{1, \ldots, k\}$$

$$\text{such that } A_{i_m} \cap B_j = \{x_{i_m}\} \text{ and } |A_{i_m}| = 1 \text{ for at most one } i_m$$

$$(3)\ \forall \text{ proper, nonempty } I \subset \{1, \ldots, k\},$$

$$\exists B_j \in \mathcal{B} \text{ such that } B_j \cap (\cup_{i \in I} A_i) \neq \phi \text{ and } B_j \cap (\cup_{i \in I} A_i)^C \neq \phi$$

$$(4)\ \forall \text{ proper, nonempty } J \subset \{1, \ldots, n\},$$

$$\exists A_i \in \mathcal{A} \text{ such that } A_i \cap (\cup_{j \in J} B_j) \neq \phi \text{ and } A_i \cap (\cup_{j \in J} B_j)^C \neq \phi.$$

Being interlaced is a very strong condition; however, it is often satisfied for partitions of interest. An example of special significance is that of repeated multinomial event trees, in which each node has the same number of branches emanating from it. Form one partition by separating sample paths into sets based on the node they pass through in the penultimate period. Form another partition in which all sample paths with the same numbers of each type of transition (for example the highest transition, the middle transition, the lowest transition) are grouped together (this is grouping based on what would be the sufficient statistic in an i.i.d. multinomial problem). Then these two partitions are interlaced.

Theorem 3 *Assume that Axioms 5 and 3 hold, that $|S| < \infty$, that $\Sigma = 2^S$, and that every nonempty $F \subset S$ is non-null. Then the following two conditions are equivalent:*

(1) There exist interlaced partitions $\mathcal{A}, \mathcal{B} \subset \Sigma$ such that $\succsim$ satisfies Axiom 1 relative to both $\mathcal{A}$ and $\mathcal{B}$.

(2) $\succsim$ satisfies the full independence axiom over acts in L_0.

This is an important negative result: uncertainty aversion cannot be maintained in the face of a very rich set of consistent conditional preferences. Put another way, this theorem states that enough restricted independence adds up to full independence.

6 A Dynamic Domain for Preferences

I consider a dynamic setting in which time is discrete and the horizon is finite: $t \in \{0, 1, \ldots, T\}$.

To maintain continuity of notation, let the state space be denoted S as before (rather than the notation Ω more typical in the stochastic-process literature, and used in, e.g., Epstein and Schneider (2001)). The natural filtration (the filtration representing the revelation of information about the state of the world over time) is $\{\mathcal{F}_t\}_{t=0}^T$, where $\mathcal{F}_0$ is trivial (includes only S and ϕ). For any $t \in \{0, 1, \ldots, T\}$, I will be most concerned with the *atoms* of the σ-algebra $\mathcal{F}_t$. The atoms of a σ-algebra are the sets in that σ-algebra such that any other set in the σ-algebra is the union of some collection (possibly empty) of atoms. The atoms of $\mathcal{F}_t$, taken together, thus make up the finest partition of S that can be formed using sets in $\mathcal{F}_t$. When Epstein and Schneider (2001) state, "We assume that for each finite t, $\mathcal{F}_t$ corresponds to a finite partition," (page 4) the finite partition they are referring to is, in fact, the set of atoms of $\mathcal{F}_t$. I make the same assumption: that there is a finite number of atoms in each $\mathcal{F}_t$ for $t \in \{0, 1, \ldots, T\}$. Any filtration $\{\mathcal{F}_t\}_{t=0}^T$ satisfying this assumption has an event-tree representation, in which each atom of $\mathcal{F}_t$ is identified with a set of terminal nodes that originate from some node at the time-t level of the tree. The branches from a node at time t to nodes at time $t+1$ can be thought of as connecting an atom in $\mathcal{F}_t$ to the atoms in $\mathcal{F}_{t+1}$ which partition it. Given such a finite filtration $\{\mathcal{F}_t\}_{t=0}^{T}$, the distinction between a σ-algebra and an algebra is vacuous, and the algebra Σ on S is related to that filtration by $\Sigma = \mathcal{F}_T$.

The set X of *consequences* or *prizes* will be the set of $(T+1)$-long sequences of consumption bundles, $(c_0, c_1, \ldots, c_T)$, such that each $c_t \in C$ for some set C (which might, for example, be the positive reals). It is now tempting to proceed with L_0 equal, as usual, to the set of all Σ (or $\mathcal{F}_T$) measurable finite step functions from S, the set of states, into Y, the set of simple probability distributions on X. This is problematic because, as noted by Epstein and Schneider (2001), such a definition would fail to account for the order in which information is revealed according to the filtration $\{\mathcal{F}_t\}_{t=0}^{T}$, since L_0 would then include, for example, acts in which c_0, consumption at time 0, was $\mathcal{F}_T$ measurable but not $\mathcal{F}_0$ measurable (in other words, constant, since $\mathcal{F}_0$ is trivial). This would mean that consumption at time 0 was dependent on some possible outcome not "known" (according to the filtration) until some time in the future, making it difficult to use the filtration for its customary purpose: to represent the information structure of the environment.

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

These are exactly the set of adapted horse lotteries and the set of roulette lotteries that Epstein and Schneider (2001) work with. Following their notation, I term the set of adapted horse lotteries $\mathcal{H}$.

In order to directly apply the results of Section 5, I must obtain some set of consequences, denoted X^U, paired with the set of simple probability distributions on it, which I label Y^U, such that the set of Σ-measurable finite step functions $f : S \to Y^U$ is equivalent, from a preference perspective, to $\mathcal{H}$. I do so by showing that under Axiom 3, preferences over roulette lotteries are representable by a von Neumann-Morgenstern utility function, which is, moreover, additively time-separable. The vN-M utility function is also, as usual, unique up to a positive affine transformation. Then I take $X^U \subset \mathbb{R}$ to be the set of all vN-M utility values arising from consumption lotteries. Preferences over adapted acts naturally induce a preference relation on $f : S \to Y^U$, which then satisfies Axiom 3.

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

Now I will use Proposition 1 to show that the results obtained in Section 5 continue to hold when preferences are defined on the dynamic domain of the current section.

Theorem 4 *Theorems 1, 2, and 3 continue to hold when $\succsim$ is defined on the dynamic domain of this section. Moreover, the function u in Theorem 2 is additively time-separable.*

7 Understanding Uncertainty Aversion Frameworks

Using the dynamic framework I have now developed, I will be able to classify and analyze the approaches to uncertainty aversion that have been used in the asset pricing literature. The recursive multiple-priors approach and the robust control multiplier approach both admit consistent conditional preferences, so they both lead to prismatic (and, in fact, rectangular) sets of priors. Likewise, model-based multiple-priors admits consistent conditional preferences, though relative to a different partition, and thus leads to prismatic priors (again, relative to a different partition). However, as the negative result of Theorem 3 makes clear, assuming one set of consistent conditional preferences, while enforcing uncertainty aversion, may well rule out the possibility that some other set of conditional preferences is consistent.

To more effectively see the differences between the three approaches that I will consider, note that, for any Σ-measurable Euclidean-valued finite-ranged function $\chi(s)$, the level sets of χ deliver a finite partition of S into Σ-measurable sets. I shall call this partition $\mathcal{A}_\chi$.

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

The partition generated by the level sets of the projection χ is model-based: states of the world are partitioned according to the values of an economic model's parameters. If one assumes Axiom 1, the restricted independence axiom, and Axiom 2 with respect to this partition, and also assumes Axiom 3, then Theorem 2 implies that there is a maxmin expected utility representation for preferences, and that the set of distributions on the state of the world is prismatic with respect to the partition generated by the level sets of χ. (In place of

Axioms 1 and 2, one could assume Axioms 4 and 5 with respect to the partition.) Because χ projects the set of states of the world onto the parameter space, the prismatic structure of the set of distributions on the state of the world is, in fact, a "multiple-priors multiple-likelihoods" structure: there is a set of distributions on the parameters of the economic model and, given any value of the parameters of the model, there is a set of distributions on the vector of data.

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

Some of the recent work of Epstein and Schneider (Epstein and Schneider (2001) and Epstein and Schneider (2002)), on the other hand, uses several functions χ_{it}. Indeed, in any event tree, the Epstein and Schneider (2001) assumptions state that consistent conditional preferences exist conditional on any node in the tree. Thus, there is one χ_{it} function for each node in the tree, and that function takes on as many values as there are branches from the node, plus one. Each χ_{it} function is constant on the complement of the set of successors to its designated node. This setup evidently leads to dynamic consistency, since then a (very) full set of consistent conditional preferences exists. One could prove the Epstein-Schneider theorem by recursively applying Theorem 2. In fact, the condition on the set of distributions that they term "rectangularity" is a special case of the prismatic condition.

The intuition behind rectangularity can be captured in a simple example. Suppose that the Ellsberg example were extended to include two draws, in sequence and with replacement, from the unknown urn. Rectangularity would imply that the decision-maker's preferences correspond to a situation in which the first draw is taken from one unknown urn, and, if the ball drawn is black, the second draw is taken from a second, separate unknown urn; if the first ball is red, the second draw is taken from a third, again distinct, unknown urn. Thus, rectangularity implies a very strong form of nonstationarity even when the physical realities of the situation are homogeneous over time. Although Epstein and Schneider (2001) contend that rectangularity is compatible with learning and that their method leads to distribution-by-distribution updating via Bayes rule, the extreme nonstationarity implied by rectangularity means that Bayesian updating of a rectangular set of distributions bears little resemblance to, say, Bayesian updating of a set of posterior distributions on a model parameter. Indeed, the recursive multiple-priors approach is incompatible with model-based multiple-priors (where the model is i.i.d.) on simple multinomial repeated event trees, as an application of Theorem 3 shows. The two may coexist, but only if there is no uncertainty aversion.

The structural link between robust control multiplier preferences and recursive multiple-priors preferences has been discussed in Section 2; because of this connection, and because recursive multiple-priors has already been discussed, a separate discussion of the robust control multiplier method is not included here.

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

Given the impossibility of "having it all" in terms of conditional preferences and uncertainty aversion, what approach is most attractive? I believe that model-based multiple-priors is most promising. This is due to the fact that it can naturally induce learning, which is certainly one of the key features of dynamic decision making. As Epstein and Schneider (2002) have shown, it is possible to generate something that looks somewhat like learning within a recursive multiple-priors framework, but they freely admit that it is completely *ad hoc*. The sort of learning they prescribe is the kind that would occur if a decision maker re-minimized over the set of priors in model-based multiple-priors each period, but was aware that this would occur and planned for it. This is not what one typically thinks of as learning. As a consequence, such intuitive characteristics as exchangeability are incompatible with the Epstein and Schneider (2002) method.

8 The Investor's Consumption and Portfolio Choice Problem

I consider an investor who faces an intertemporal choice problem: at each point in time, the investor chooses what fraction of wealth to consume, and decides on how to allocate the remaining wealth between a risky asset and a riskless asset. I depart from the discrete-time framework of Section 6, and follow the seminal work of Merton (1969) by setting the problem in continuous time. Although discrete-time, discrete-state settings allow for axioms that are clearer and less cluttered by technical requirements, solving the investor's problem in continuous time shows the generality of the concepts developed above. A continuous-time model also permits me to compare my results with those of Anderson, Hansen, and Sargent (2000), Chen and Epstein (2002), and Maenhout (2001), all of whom set their models of uncertainty aversion in continuous time, and to contrast my findings with those of Brennan (1998), Brennan and Xia (2001), Cvitaniȅ, Lazrak, Martellini, and Zapatero (2002), and Xia (2001), all of whom set their models of Bayesian learning in continuous time.

The conditional preferences I am most interested in are those based on some key parameter, such as the expected return in the specific model considered below. The idea is similar to that explored in an econometric setting by Chamberlain (2000): if the investor knew what the expected return was, all of the uncertainty (though not necessarily all of the risk) would be resolved. In a model such as this, a low-dimensional parameter captures all of the uncertainty, though not necessarily all of the risk, perceived by the investor. However, it is worth noting that the complexity of the model considered below is *not* constrained by the model-based multiple-priors approach, which could deal with a model that allowed endogenous structural breaks and other nonstationarities, or even a model in which the parameter was infinite-dimensional (so that the setting was "nonparametric"). The model treated in this section was chosen for its simplicity, and because it demonstrates how model-based multiple-priors can accomodate learning of a familiar and intuitive sort.

As I have explored at length above, consistent conditional preferences based on conditioning upon the parameters of an economic model are not, in general, compatible with the recursive multiple-priors approach of Chen and Epstein (2002), Epstein and Schneider (2001), and Epstein and Schneider (2002). Because I impose the requirement that the conditional preferences (given the value of the expected return) satisfy not only Axioms 4 and 5 but also full independence (capturing the notion that the uncertainty in the expected return is the only uncertainty perceived by the investor), these preferences are also incompatible with those of a robust-control investor.

Despite these differences, I will show that model-based multiple-priors can, for certain types of sets of priors, deliver the same portfolio and consumption rules that Chen and Epstein (2002) and Maenhout (2001) derive under the much less transparent recursive multiple-priors and modified robust-control methods. Moreover, I will show that the model-based multiple-priors approach is more general: under different types of sets of priors, a model-based multiple-priors investor learns from the past.

The investor has power utility over consumption at each moment in time. The coefficient of relative risk aversion is assumed to be greater than one (though the log case is the natural limit as $\gamma \to 1$ for both of the models):

$$\gamma > 1.$$

The (uncertainty-averse) investor allocates wealth between a riskless asset, which follows the process

$$dB_t = rB_t dt, \tag{45}$$

and a risky (and indeed, uncertain) asset, which follows a geometric Brownian motion,

$$dS_t = \mu S_t dt + \sigma S_t dw_t. \tag{46}$$

The investor is *uncertain* about the value of the expected return parameter μ. In a Bayesian (subjective expected utility theory) model, the investor would place a prior on μ. Here, however, the investor has a (closed, convex) *set* of prior distributions on the expected return. Preferences over consumption processes (and thus over consumption and portfolio choices) are represented by the minimum (over the set of priors on μ) of the expected utility (where expectations are taken over μ, using some prior in the set, and over the process $\{S_t\}_{t\in[0,T]}$, using the geometric Brownian motion "likelihood") of the consumption process.

Let the closed, convex set of prior distributions on μ be denoted Π, so that:

$$\mu \sim \pi \in \Pi. \tag{47}$$

Also, let

$$\begin{aligned} & F(W_0) \qquad (48) \\ = \; & \left\{ \{c_t\}_{t=0}^T, \{\varphi_t\}_{t=0}^T \; : \; dW_t = (\varphi_t(\mu - r) + r) W_t dt + \varphi_t \sigma W_t dw_t - c_t dt \text{ and } W_T \geq 0 \right\}, \end{aligned}$$

where φ_t is the weight on the risky asset in the investor's portfolio at the moment t, so that $F(W_0)$ is the set of feasible consumption and portfolio choices given an initial wealth of W_0. Then the investor's problem is:

$$\max_{c_t, \varphi_t \in F(W_0)} \left\{ \min_{\pi \in \Pi} \left\{ E_\pi \left[E_S \left[\int_{t=0}^T e^{-\rho t} \frac{c_t^{1-\gamma}}{1-\gamma} dt \middle| \mu \right] \right] \right\} \right\}, \tag{49}$$

where the operator "$E_\pi[\cdot]$" denotes expectation over different values of μ using the prior π, and the operator "$E_S[\cdot|\mu]$" denotes expectation over possible stock price paths using the measure generated by the geometric Brownian motion with mean return μ.

In this form, the problem is not tractable. However, one may follow the logic of Chamberlain (2000) and Chamberlain (2001) in applying the minimax theorem (Ferguson (1967), page 85) to conclude that

$$\max_{c_t,\varphi_t\in F(W_0)}\left\{\min_{\pi\in\Pi}\left\{E_\pi\left[E_S\left[\int_{t=0}^{T}e^{-\rho t}\frac{c_t^{1-\gamma}}{1-\gamma}dt\middle|\,\mu\right]\right]\right\}\right\} \tag{50}$$

$$= \min_{\pi\in\Pi}\left\{\max_{c_t,\varphi_t\in F(W_0)}\left\{E_\pi\left[E_S\left[\int_{t=0}^{T}e^{-\rho t}\frac{c_t^{1-\gamma}}{1-\gamma}dt\middle|\,\mu\right]\right]\right\}\right\}. \tag{51}$$

Note that the technical requirements of the theorem presented by Ferguson (1967) are satisfied in this case if one rules out obviously suboptimal behavior (such as investments which risk ruin). On the second line above, the problem *within* the minimization (over $\pi \in \Pi$) is a standard one: it is a Bayesian problem of consumption and portfolio choice under the prior $\pi \in \Pi$. Equivalently, it is a problem of the sort that a subjective expected utility maximizer might encounter in this setting.

At this stage, I will need to put more structure on the set of prior distributions on μ in order to proceed. I will first consider a case in which the investor will not learn from the past; although this is evidently not my ultimate goal, it is both a useful benchmark and a link to the results of Chen and Epstein (2002) and Maenhout (2001).

8.1 A Benchmark Model

Consider some interval $\left[\underline{\lambda},\overline{\lambda}\right]$ of possible expected returns. Let

$$\Pi \;\equiv\; \left\{\pi \,:\, \pi\left(\left[\underline{\lambda},\overline{\lambda}\right]\right)=1\right\}. \tag{52}$$

Note, in particular, that Π includes all of the Dirac delta functions on the interval $\left[\underline{\lambda},\overline{\lambda}\right]$.

In order to solve the maxmin expected utility problem I have laid out, it will be helpful to consider first the (much simpler) problem

$$\max_{c_t,\varphi_t\in F(W_0)}\left\{E_\pi\left[E_S\left[\int_{t=0}^{T}e^{-\rho t}\frac{c_t^{1-\gamma}}{1-\gamma}dt\middle|\,\mu\right]\right]\right\}, \tag{53}$$

where π is a delta function at some $\lambda \in \left[\underline{\lambda},\overline{\lambda}\right]$. But this is simply the usual model of continuous-time portfolio choice with a constant investment opportunity set; it was first solved by Merton (1969). I reproduce the solution for convenience. Let $X = \frac{\lambda - r}{\sigma}$ (the Sharpe ratio associated with the expected return λ).

Proposition 2 *Let*

$$a \equiv \frac{1}{\gamma}\left\{\rho - (1-\gamma)\left[\frac{X^2}{2\gamma} + r\right]\right\}. \tag{54}$$

The solution to the problem (53) is:

$$\frac{c_t^o}{W_t^o} = \frac{a}{1 - e^{a(t-T)}} \tag{55}$$

$$\varphi_t^o = \frac{X}{\gamma\sigma}. \tag{56}$$

The indirect utility function is

$$J(W,t) = \frac{W^{1-\gamma}}{1-\gamma} e^{-\rho t}\left(\frac{1 - e^{a(t-T)}}{a}\right)^{\gamma}. \tag{57}$$

I now show that an uncertainty-averse investor whose set of priors is Π will behave as though investment opportunities were the worst possible constant investment opportunities in the set, and will not attempt to learn from the data.

Theorem 5 *An uncertainty-averse investor with the set of priors* Π *will optimally invest and consume as if* $X = X^{LF}$, *where* $X^{LF} = \frac{\arg\min_{\lambda \in [\underline{\lambda}, \overline{\lambda}]}\left\{\left(\frac{\lambda - r}{\sigma}\right)^2\right\} - r}{\sigma}$.

This theorem has several interesting consequences. The most obvious, perhaps, is that the uncertainty-averse investor does not attempt to learn from the data; the interpretation of this result is that the investor is so fully convinced that investment opportunities are very poor that no amount of data could alter her beliefs. Another important consequence is that this result is virtually identical to that obtained by Chen and Epstein (2002) under a very different set of priors (a rectangular set which they describe as representing "κ-ignorance"). Further, it is of the same sort (though not precisely identical) to the result obtained by Maenhout (2001) under his modification of the robust control method.

It is of interest to note that one can obtain the same results using the model-based multiple-priors approach that have been derived in the past only by using much more elaborate methods. However, the next subsection shows that the model-based multiple-priors approach has another, more important advantage over existing methods: an investor who uses the model-based multiple-priors approach can have a preference for learning.

8.2 A Model of Robust Learning

I now consider an investor who rules out dogmatic behavior: such an investor is never willing to act as if investment opportunities were completely known. It seems likely that this assumption is a better description of reality than the assumption that investors act as though they are completely certain of poor investment opportunities.

To represent the preferences of a non-dogmatic investor, each prior in the set of priors used by the investor must have non-singleton support. It seems natural to go to the opposite end of the spectrum, and assume that each prior in the set has full support. Although sets of normal distributions are quite special, sets of *mixtures* of normal distributions can approximate many distributions quite well. Brennan and Xia (2001) make use of mixture-of-normals priors for just this reason. Indeed, sets of mixtures of normal distributions can include multi-modal and skewed distributions. I thus fix an interval $\left[\underline{\lambda}, \overline{\lambda}\right]$ and set

$$\Pi \equiv \overline{\text{conv}} \left\{ N\left(\lambda, \nu^2\right) \; : \; \lambda \in \left[\underline{\lambda}, \overline{\lambda}\right] \right\}, \tag{58}$$

where the notation "$\overline{\text{conv}}$" should be read as "the closed convex hull of." By definition, then, Π is closed and convex, as it should be. As mentioned above, Π contains a diverse collection of probability measures, including skewed and multi-modal distributions. The assumption that Π takes the form above is, then, not very restrictive. The one notable constraint I have placed on Π is that there is no mixing over the scale parameter ν. This restriction is relaxed in the following two subsections.

In order to solve the maxmin expected utility problem I have laid out, it will be helpful to consider first the (much simpler) problem

$$\max_{c_t, \varphi_t \in F(W_0)} \left\{ E_\pi \left[E_S \left[\int_{t=0}^{T} e^{-\rho t} \frac{c_t^{1-\gamma}}{1-\gamma} dt \middle| \mu \right] \right] \right\}, \tag{59}$$

where π is a normal distribution with mean λ and variance ν^2:

$$\pi = N\left(\lambda, \nu^2\right). \tag{60}$$

This Bayesian problem, and variants of it, have been extensively investigated. Detemple (1986), Dothan and Feldman (1986), and Gennotte (1986) demonstrated that the so-called *separation principle* holds in this problem (the investor can solve the problem by considering the estimation problem and the portfolio choice problem separately). Later, Lakner (1995), Lakner (1998), Karatzas and Zhao (2001), and Zohar (2001) attacked the problem, but no analytical solution was discovered for an investor with power utility. Rogers (2001) presented (without proof) the first analytical solution of a problem of this sort under power utility, but with utility over terminal wealth. Subsequently, Cvitanič, Lazrak, Martellini, and Zapatero (2002) gave, with proof, a generalization of the solution of Rogers (2001) to the case of multiple assets, but still only under power utility over terminal wealth. Rogers (2001) also contains a partial treatment, again without proof, of the Bayesian problem of an investor with power utility over intermediate consumption, but gives only time-zero optimal wealth for this problem. Optimal wealth at times $t > 0$, the indirect utility of wealth function, the optimal consumption-wealth ratio, and the optimal allocation to the risky asset are not provided by Rogers (2001).

I will now define the posterior expected Sharpe ratio at time t,

$$\overline{X}_t \equiv \frac{\tau^2 t}{1+\tau^2 t} \left(\frac{1}{\sigma} \left(\frac{\ln(S_t) - \ln(S_0)}{t} - \left(r - \frac{1}{2}\sigma^2 \right) \right) + \frac{m}{\tau^2 t} \right), \tag{61}$$

where the parameters are defined by $\tau^2 = \frac{\nu^2}{\sigma^2}$ and $m = \frac{\lambda - r}{\sigma}$. This definition takes the average log return, applies the usual concavity correction term in σ, strips the riskless rate out of the average log return, and scales by the standard deviation, giving a measure of the sample mean Sharpe ratio. The definition then takes a weighted average of this sample mean Sharpe ratio and the prior expected Sharpe ratio, m, in order to obtain the posterior expected Sharpe ratio.

Chapter 5 of Campbell and Viceira (2002) shows that the investor's problem as I have laid it out above is a single-state-variable model of stochastic investment opportunities. The state variable is the posterior expected Sharpe ratio, as defined above. Since innovations to the posterior expected Sharpe ratio are perfectly *positively* correlated with instantaneous returns, one expects that investors will have a *negative* hedging demand, and that hedging demand will grow as the investment horizon lengthens.

Although the general structure of such problems has been understood, I go beyond previous work on parameter uncertainty by giving the solution to the investor's problem in closed form.

Proposition 3 *Define*

$$\delta(t) \equiv \frac{\tau^2}{1+\tau^2 t} \tag{62}$$

$$A(s-t,\delta(t)) \equiv \left[\frac{\left(1-\frac{1}{\gamma}\right)(s-t)}{1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)}\right] \tag{63}$$

$$B(s-t,\delta(t)) \equiv e^{-\frac{1}{\gamma}\rho(s-t)}e^{-\left(1-\frac{1}{\gamma}\right)r(s-t)}\frac{\left(1+\delta(t)(s-t)\right)^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)\right)^{\frac{1}{2}}}. \tag{64}$$

Then the solution to the problem (59) is

$$\frac{c_t^o}{W_t^o} = \frac{1}{\int_t^T B(s-t,\delta(t))\exp\left\{-\frac{1}{2\gamma}A(s-t,\delta(t))\overline{X}_t^2\right\}ds} \tag{65}$$

$$\varphi_t^o = \frac{\overline{X}_t}{\gamma\sigma}\times \left(1-\delta(t)\frac{\int_t^T B(s-t,\delta(t))\exp\left\{-\frac{1}{2\gamma}A(s-t,\delta(t))\overline{X}_t^2\right\}A(s-t,\delta(t))\,ds}{\int_t^T B(s-t,\delta(t))\exp\left\{-\frac{1}{2\gamma}A(s-t,\delta(t))\overline{X}_t^2\right\}ds}\right). \tag{66}$$

The indirect utility function is

$$J\left(W,\overline{X}_t,t\right) = \frac{W^{1-\gamma}}{1-\gamma}e^{-\rho t}\left(\int_t^T B\left(s-t,\delta\left(t\right)\right)\exp\left\{-\frac{1}{2\gamma}A\left(s-t,\delta\left(t\right)\right)\overline{X}_t^2\right\}ds\right)^{\gamma}.(67)$$

In particular, the expected utility at time $t = 0$ *as a function of initial wealth and the parameter values is:*

$$J\left(W,m,0\right) = \frac{W^{1-\gamma}}{1-\gamma}\left(\int_0^T B\left(s,\tau^2\right)\exp\left\{-\frac{1}{2\gamma}A\left(s,\tau^2\right)m^2\right\}ds\right)^{\gamma}. \tag{68}$$

Hedging demand is negative (as long as $\overline{X}$ is positive), as expected for a state variable whose shocks are perfectly positively correlated with those of the risky asset. These results are consistent with the numerical results of Brennan (1998) in a model with utility over terminal wealth and with the results of Merton (1971) in a model with exponential utility and an infinite horizon.

Horizon effects are evidently present as well: holding fixed the posterior expected Sharpe ratio, at longer horizons hedging demand has a larger impact in decreasing the magnitude of demand for the risky asset. This is in contrast to the sort of horizon effect that has been observed by researchers who study the dividend yield's influence on investment opportunities and portfolio choice: Barberis (2000), Wachter (2002), and Xia (2001) all find that the allocation to the stock market increases with the investment horizon in these models. As Wachter (2002) shows, this is due to the very negative contemporaneous correlation between stock market returns and innovations to the dividend yield. In contrast, in a learning model such as the one I consider here, stock market returns and innovations to the state variable (here, the posterior expected Sharpe ratio) are perfectly *positively* correlated.

Wachter (2002) pointed out, in a different context, that hedging demand had a weighted-average form. The same is true in this setting, and hedging demand can thus be interpreted as the duration of the investor's consumption stream with respect to the state variable (in this case, the posterior expected Sharpe ratio $\overline{X}_t$). This interpretation of the investor's hedging demand exploits the analogy between wealth and a bond: wealth can be thought of as a bond which pays a consumption coupon.

To my knowledge, mine is the first complete analytical solution of the problem of a Bayesian, constant relative risk aversion investor with intermediate consumption who faces parameter uncertainty. (A different, and complementary, problem of "steady-state" learning about time-varying investment opportunities has been solved analytically by Rodriguez (2002).) As noted above, however, Rogers (2001) and Cvitanič, Lazrak, Martellini, and Zapatero (2002) have analytical solutions of the problem of a Bayesian, constant relative risk aversion investor with utility over terminal wealth who faces parameter uncertainty; Rogers (2001) does not include a proof, while Cvitanič, Lazrak, Martellini, and Zapatero (2002) do provide a proof of their solution's optimality. Further, Rogers (2001) contains a partial treatment, again without proof, of the intermediate consumption case (only the optimal time-zero wealth is given; the optimal wealth at times $t > 0$, the optimal consumption-wealth ratio, the optimal portfolio weight on the risky asset, and the indirect utility of wealth are not provided).

As an aside, it is evident upon inspection that (due to the exponentially decaying terms in the function B and the fact that $\gamma > 1$) the limit as $T \to \infty$ of the expected utility exists and is strictly negative. Thus, an infinite horizon presents no special problems: the infinite-horizon solution can be obtained from the above simply by taking the limit as $T \to \infty$.

Now, in order to solve a maxmin expected utility investor's problem, I must find the least favorable prior in the set Π. This can be accomplished with the aid of the martingale method (Cox and Huang (1989) and Cox and Huang (1991)), as the theorem below shows.

Theorem 6 *An uncertainty-averse investor whose set of priors is* Π *optimally invests as though the prior distribution were the normal distribution with mean* $\arg\min_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} \left\{ \left(\frac{\lambda - r}{\sigma} \right)^2 \right\}$. *Thus, the solution given in Proposition 3 above applies to the optimal consumption and portfolio choice under maxmin expected utility, with the minimizing parameter substituted* *into the given formulae.*

8.3 Richer Sets of Prior Distributions

As mentioned in the previous subsection, we may be interested in richer sets of prior distributions than the set Π specified above. In particular, each of the normal distributions from which Π is built up (by taking the closed convex hull) has the same variance. Although this implies that Π includes bimodal distributions and skewed distributions, it does not imply that Π includes heavy-tailed distributions. To include heavy-tailed distributions, the set of priors must permit the building-block normal distributions to differ in variance as well as differing in mean. To this end, I fix intervals $\left[\underline{\lambda}, \overline{\lambda}\right]$ and $[\underline{\nu}^2, \overline{\nu}^2]$ and define:

$$\Gamma \equiv \overline{\text{conv}} \left\{ N\left(\lambda, \nu^2\right) : \lambda \in \left[\underline{\lambda}, \overline{\lambda}\right], \nu^2 \in \left[\underline{\nu}^2, \overline{\nu}^2\right] \right\}. \qquad (69)$$

Finding the least-favorable prior over the set Γ is generally more challenging than finding the least-favorable prior over the set Π. In order to obtain analytical answers, I consider the more tractable problem of an investor who has utility over terminal wealth, rather than over intermediate consumption.

Define

$$\begin{aligned} & G\left(W_0\right) \qquad (70) \\ = & \left\{ \{\varphi_t\}_{t=0}^{T} : dW_t = \left(\varphi_t \left(\mu - r\right) + r\right) W_t + \varphi_t \sigma W_t dw_t \text{ and } W_T \geq 0 \right\}, \end{aligned}$$

where φ_t is the weight on the risky asset in the investor's portfolio at the moment t, so that $G\left(W_0\right)$ is the set of feasible portfolio choices given an initial wealth of W_0. Then the problem of an uncertainty-averse investor with utility over terminal wealth is

$$\max_{\varphi_t \in G(W_0)} \left\{ \min_{\pi \in \Gamma} \left\{ E_\pi \left[E_S \left[\frac{W_T^{1-\gamma}}{1-\gamma} \middle| \mu \right] \right] \right\} \right\}, \qquad (71)$$

and the analysis proceeds very similarly to that in Subsections 8.2 and 8.3.

As was the case for intermediate consumption, it is first convenient to get the solution for a single-prior Bayesian investor. Let the investor have a normal prior distribution with mean λ and variance ν^2. Then the investor's problem is

$$\max_{\varphi_t \in G(W_0)} \left\{ E_\pi \left[E_S \left[\frac{W_T^{1-\gamma}}{1-\gamma} \middle| \mu \right] \right] \right\}. \tag{72}$$

This problem has been solved, without proof, by Rogers (2001), and solved with proof by Cvitaniċ, Lazrak, Martellini, and Zapatero (2002). I reproduce their solution here for convenience, though I rewrite the solution in a way that I believe to be more conducive to interpretation, since it disentangles hedging demand and myopic demand.

Proposition 4 *The solution to the problem (72) is:*

$$\varphi_t^o = \frac{\overline{X}_t}{\gamma\sigma}\left(1 - \frac{\frac{\gamma-1}{\gamma}\delta(t)(T-t)}{1+\frac{\gamma-1}{\gamma}\delta(t)(T-t)}\right). \tag{73}$$

The indirect utility of wealth is

$$\begin{aligned} J\left(W, \overline{X}_t, t\right) &= \frac{W^{1-\gamma}}{1-\gamma} \frac{\left(1+\delta(t)(T-t)\right)^{\frac{\gamma-1}{2}}}{\left(1+\frac{\gamma-1}{\gamma}\delta(t)(T-t)\right)^{\frac{\gamma}{2}}} \\ &\quad \times \exp\left\{ -\frac{\overline{X}_t^2}{2} \frac{\frac{\gamma-1}{\gamma}(T-t)}{1+\frac{\gamma-1}{\gamma}\delta(t)(T-t)} - r(\gamma-1)(T-t) \right\}. \end{aligned} \tag{74}$$

This solution is quite intuitive; the expression for φ_t^o shows that hedging demand is best interpreted here, as in the intermediate consumption problem, as relative (that is, as a fraction of myopic demand). Since the relative hedging demand is of the form $\frac{y}{1+y}$, where $y = \frac{\gamma-1}{\gamma}\delta(t)(T-t)$, it is increasing in the quantity $\frac{\gamma-1}{\gamma}\delta(t)(T-t)$. This means that relative hedging demand increases in risk aversion (being zero, of course, for logarithmic utility), in posterior variance (and hence in prior variance), and in the remaining investment horizon $T-t$.

The absence of integrals over time makes this problem more tractable than the intermediate consumption problem. This tractability delivers the following theorem regarding the least favorable prior and the maxmin expected utility investment rule when the set of priors is Γ, that is, when the normal distributions from which the set of priors is built up differ both in mean and in variance.

Theorem 7 *Define*

$$\lambda^{LF} \equiv \arg\min_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} \left\{ \left(\frac{\lambda - r}{\sigma} \right)^2 \right\} \tag{75}$$

$$m^{LF} \equiv \frac{\lambda^{LF} - r}{\sigma} \tag{76}$$

$$\tau^{2,*} \equiv \frac{1}{2} \left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma} T} + \sqrt{\left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma} T} \right)^2 + \frac{4}{T} \left(m^{LF}\right)^2} \right). \tag{77}$$

An uncertainty-averse investor whose set of priors is Γ and who has power utility over terminal wealth optimally invests as though the prior distribution were the normal distribution with mean λ^{LF} and variance

$$\nu^{2,\,LF} \equiv \max\left\{ \underline{\nu}^2, \min\left\{ \overline{\nu}^2, \sigma^2 \tau^{2,*} \right\} \right\}. \tag{78}$$

Thus, the solution given in Proposition 4 above applies to the optimal portfolio choice under maxmin expected utility, with the minimizing parameter substituted into the given formulae.

In particular, Theorem 7 shows that the least favorable prior is *not* necessarily "extreme" with respect to prior variance. In other words, the least-favorable prior variance is *not* always $\underline{\nu}^2$ or always $\overline{\nu}^2$. Indeed, letting $\tau^{2,\,LF} = \frac{\nu^{2,\,LF}}{\sigma^2}$, we have the following proposition.

Proposition 5

$$\tau^{2,\,LF} \in \left[\frac{\gamma - 1}{\gamma} \left(m^{LF}\right)^2, \left(m^{LF}\right)^2 \right], \tag{79}$$

so $\underline{\nu}^2 \leq \frac{\gamma-1}{\gamma} \sigma^2 \left(m^{LF}\right)^2$ and $\overline{\nu}^2 \geq \sigma^2 \left(m^{LF}\right)^2$ are sufficient (though not necessary) for $\underline{\nu}^2 < \nu^{2,\,LF} < \overline{\nu}^2$, that is, an interior solution for the variance of the least favorable prior.

Furthermore, $\sigma^2\tau^{2,*}$ *satisfies*

$$\lim_{T\to\infty}\left\{\sigma^2\tau^{2,*}\right\} = \sigma^2\left(m^{LF}\right)^2 \quad (80)$$

$$\lim_{T\to 0}\left\{\sigma^2\tau^{2,*}\right\} = \frac{\gamma-1}{\gamma}\sigma^2\left(m^{LF}\right)^2 \quad (81)$$

$$\lim_{\gamma\to\infty}\left\{\sigma^2\tau^{2,*}\right\} = \sigma^2\left(m^{LF}\right)^2. \quad (82)$$

Finally,

$$m^{LF} = 0 \;\Rightarrow\; \sigma^2\tau^{2,*} = 0. \quad (83)$$

Qualitatively, this proposition shows that a large enough interval for ν^2, coupled with a nonzero m^{LF}, always leads to an interior solution for the least-favorable prior variance. The intuition behind this result is that there are two forces influencing the least-favorable prior variance, and they work in opposite directions. Typically, neither completely outweighs the other, and an interior solution exists. The first force is risk aversion: the investor dislikes higher prior variances, since they imply that investment opportunities are less certain. This very intuitive force tends to make the least-favorable prior variance higher. The second force is less obvious, but crucially important: because the investor learns, and engages in portfolio rebalancing, a prior under which the probability of a high-absolute-value Sharpe ratio is large may be quite favorable from the investor's perspective. A prior with an extremely large variance implies the belief that the expected return is either very positive or very negative with extremely high probability. If it is true that the expected return is large in absolute value, then the investor will eventually learn its sign and its magnitude, and will then face an attractive investment opportunity set (note that if the expected return is large and negative, the investor will simply short the risky asset). Thus, a sufficiently high prior variance implies a belief, on the part of the investor, that attractive investment opportunities will be available in the future after some learning has taken place. This force tends to make the least-favorable prior variance lower. It is interesting to note that this force would be absent in a buy-and-hold problem; thus, one might expect that the least-favorable prior variance in a buy-and-hold problem would always be at the upper boundary of its constraint interval.

8.4 Explicit Results and Discussion

Although studying the above results can yield a great deal of insight, there is no substitute for examining explicit quantitative results. In recognition of this, I provide Figures 3 through 6, which depict the hedging demand, consumption-wealth ratio, and allocation to the risky asset of an uncertainty-averse investor with a preference for learning. The variation in these key quantities as the prior expected Sharpe ratio, m, and the investment horizon, T, are changed is examined.

Throughout the figures, I fix the relative risk aversion $\gamma = 5$, the rate of time preference $\rho = 0.05$, and the riskless rate $r = 0.05$.

A note on interpretation is in order. As Kandel and Stambaugh (1996) were the first to recognize in finance, natural conjugate priors allow the interpretation of a prior in terms of the years of data one would need to observe, starting from a diffuse prior, in order to arrive at the given prior as a posterior. That is, with a natural conjugate prior, one can interpret prior beliefs in terms of a fictitious "prior sample." Since I have proven, in Theorem 6 above, that the least-favorable prior is of the natural conjugate form in my problem, I take advantage of this to use a prior variance, $\tau^2 = 0.05$, that I interpret as the same level of confidence that an investor who had started with a diffuse prior and then seen 20 years of data would feel. The minimal prior expected Sharpe ratio can then be interpreted as the sample mean Sharpe ratio that was observed over this pessimistic, fictitious 20-year "prior sample."

The results in Figures 3 through 6 conform nicely to intuition. Figure 3 shows that hedging demand is more important as the minimal prior expected Sharpe ratio grows in magnitude, and that hedging demand is more important as the investment horizon grows. Figure 4 shows that the consumption wealth ratio increases in the magnitude of the minimal prior expected Sharpe ratio and decreases with the investment horizon. Figure 5 demonstrates that the allocation to the risky asset is increasing in the minimal prior expected Sharpe ratio and is decreasing in the investment horizon. Figure 6 documents the same qualitative relationships, but shifts the focus from the variation induced by the minimal prior expected Sharpe ratio to the variation induced by the investment horizon.

In Figure 7, I show how an investor of the type described in Subsection 8.1 and an investor of the sort described in Subsection 8.2 compare behaviorally. The data are monthly returns (with distributions) on the CRSP value-weighted market portfolio from January 1926 to December 2001. The results are exactly as expected: the dogmatic investor's allocation to the risky asset is constant, leading to a horizontal line in the figure, while the investor with a preference for learning optimally allocates wealth to the risky asset in a way that responds to the arrival of new information. I also show the myopic allocation to the risky asset, which isolates the hedging demand of the investor with a preference for learning.

I now turn to the figures that pertain to the richer set of priors considered (with utility over terminal wealth) in Subsection 8.3. Figure 8 illustrates the results of Proposition 5 regarding the effects of investment horizon and the least-favorable prior expected Sharpe ratio on the (unconstrained) least-favorable prior variance of the Sharpe ratio. Figure 9 captures the effect of the least-favorable prior expected Sharpe ratio on hedging demand for two different investment horizons under the unconstrained least-favorable prior variance (which is permitted to vary with the least-favorable prior expected Sharpe ratio and with the investment horizon), while Figures 10 and 11 show how the least-favorable prior expected Sharpe ratio and the investment horizon, respectively, influence the investor's hedging demand under the unconstrained least-favorable prior variance (which, again, is permitted to vary with the least-favorable prior expected Sharpe ratio and with the investment horizon in each figure).

9 Conclusion

In order to analyze the problem of learning under uncertainty aversion, I have built up a theory of dynamic choice under uncertainty aversion from axiomatic foundations. This theory has allowed me to write a general and well-justified model of learning under uncertainty aversion. It has also paid dividends in other ways: it shows the essential similarities between the existing theories of decision-making under uncertainty aversion by pointing out the common set of consistent conditional preferences the different theories imply.

I have examined, and solved in closed form, the canonical intertemporal consumption and portfolio choice problem of a power-utility investor under two different sorts of uncertainty aversion. The differences in uncertainty aversion are represented by differences in the sets of priors used by the investor. Under the first set of priors, learning never occurs: the investor dogmatically expects the worst. The portfolio and consumption decisions of this investor are structurally identical to those of the investors studied by Chen and Epstein (2002) and Maenhout (2001), which demonstrates that the investors studied by these researchers are, behaviorally, also dogmatic. The second set of priors, however, yields much more interesting, and realistic, preferences: under this second set of priors, the investor has a preference for learning, and attempts to use historical data in order to learn how to invest more effectively.

References

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

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

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

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

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

Campbell, J. Y. and L. M. Viceira (2002). *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. New York, New York: Oxford University Press.

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

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

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

Cox, J. C. and C. Huang (1989). Optimum consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory 49*, 33–83.

Cox, J. C. and C. Huang (1991). A variational problem arising in financial economics. *Journal of Mathematical Economics 20*, 465–487.

Cvitanič, J., A. Lazrak, L. Martellini, and F. Zapatero (2002). Revisiting Treynor and Black (1973): An intertemporal model of active portfolio management. University of Southern California working paper.

Detemple, J. (1986). Asset pricing in a production economy with incomplete information. *Journal of Finance 41*, 383–391.

Dothan, M. U. and D. Feldman (1986). Equilibrium interest rates and multiperiod bonds in a partially observable economy. *Journal of Finance 41*, 369–382.

Dow, J. and S. Werlang (1992). Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica 60*, 197–204.

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

Epstein, L. and M. Schneider (2001). Recursive multiple-priors. Unpublished manuscript, University of Rochester.

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

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

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. San Diego, California: Academic Press.

Gennotte, G. (1986). Optimal portfolio choice under incomplete information. *Journal of Finance 41*, 733–746.

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

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

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

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

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

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

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

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

Karatzas, I. and X. Zhao (2001). Bayesian adaptive portfolio optimization. In J. Cvitanič, E. Jouini, and M. Musiela (Eds.), *Handbooks in Mathematical Finance: Option Pricing, Interest Rates and Risk Management*, pp. 632–669. Cambridge University Press.

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

Lakner, P. (1995). Utility maximisation with partial information. *Stochastic Processes and their Applications 56*, 247–273.

Lakner, P. (1998). Optimal trading strategy for an investor: the case of partial information. *Stochastic Processes and their Applications 76*, 77–97.

Liu, J., J. Pan, and T. Wang (2002). An equilibrium model of rare event premia. Unpublished manuscript, MIT.

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

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

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. *Review of Economics and Statistics 51*, 247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory 3*, 373–413.

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

Pathak, P. (2002). Notes on robust portfolio choice. Unpublished manuscript, Harvard University.

Rodriguez, J. F. (2002). Hedging demands under incomplete information. Unpublished manuscript, Massachusetts Institute of Technology.

Rogers, L. (2001). The relaxed investor and parameter uncertainty. *Finance and Stochastics 5*, 131–154.

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

Uppal, R. and T. Wang (2002). Model misspecification and under-diversification. Unpublished manuscript, London Business School.

Wachter, J. A. (2002). Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets. *Journal of Financial and Quantitative Analysis 37*, 63–91.

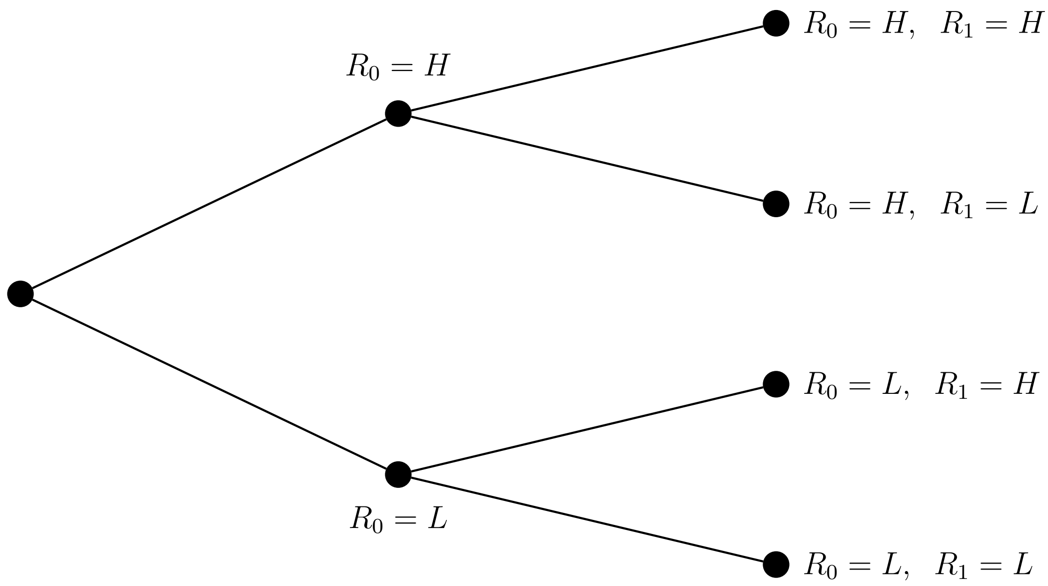
Wang, T. (1999). Updating rules for non-Bayesian preferences. Unpublished manuscript, University of British Columbia.

Wang, Z. (2002). A shrinkage approach to model uncertainty and asset allocation. Unpublished manuscript, Columbia Business School.

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

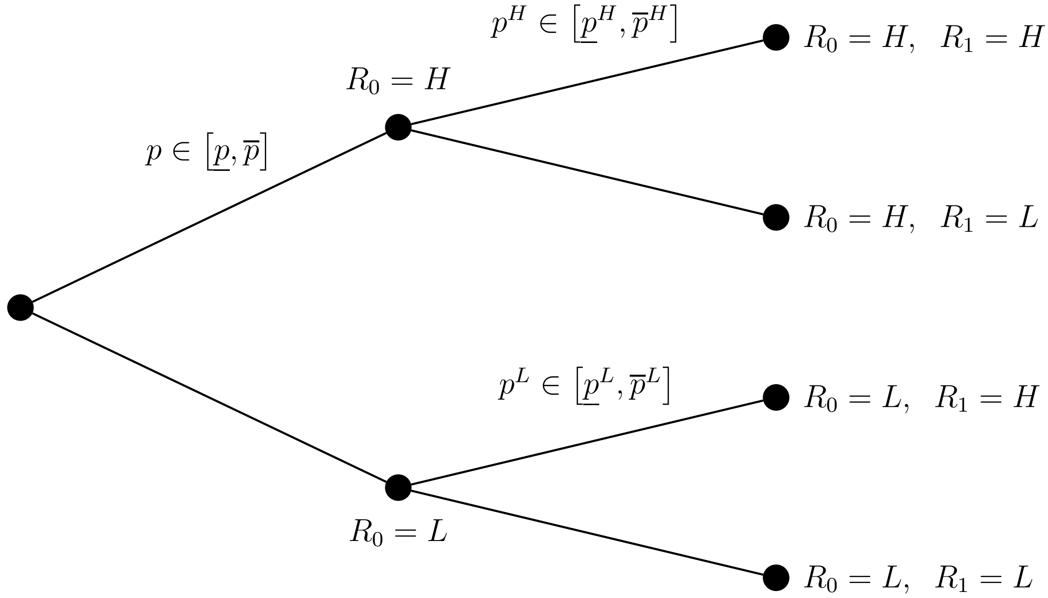
Zohar, G. (2001). Generalized Cameron-Martin formula with applications to partially observed dynamic portfolio optimization. *Mathematical Finance 11*, 475–494.

Figure 1: **The Two-period Binomial Model**



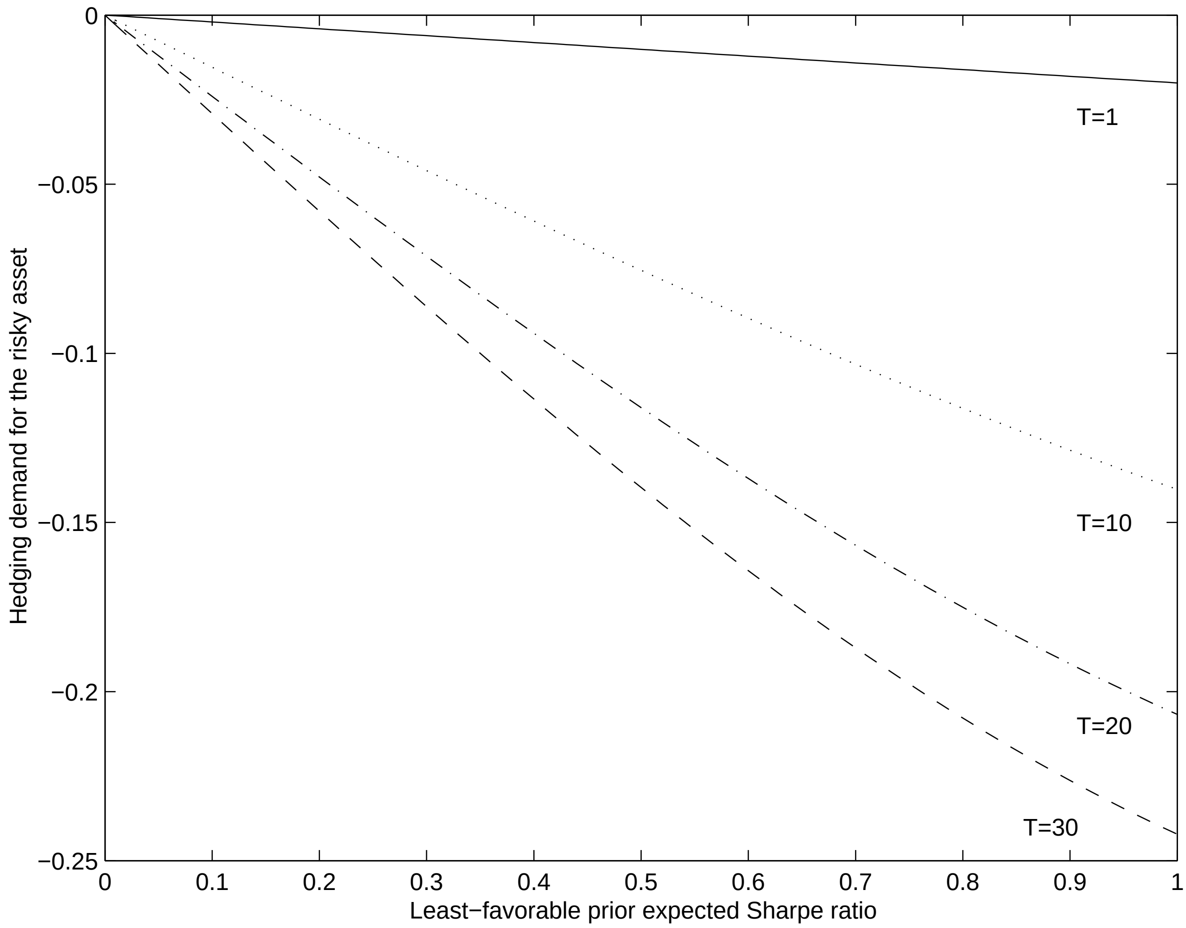
This figure depicts an event-tree representation of the two-period model with a binomial risky asset, which is described in Section 2.

Figure 2: **The Recursive Multiple-Priors "Rectangular" Set of Priors in the Two-period Binomial Model**



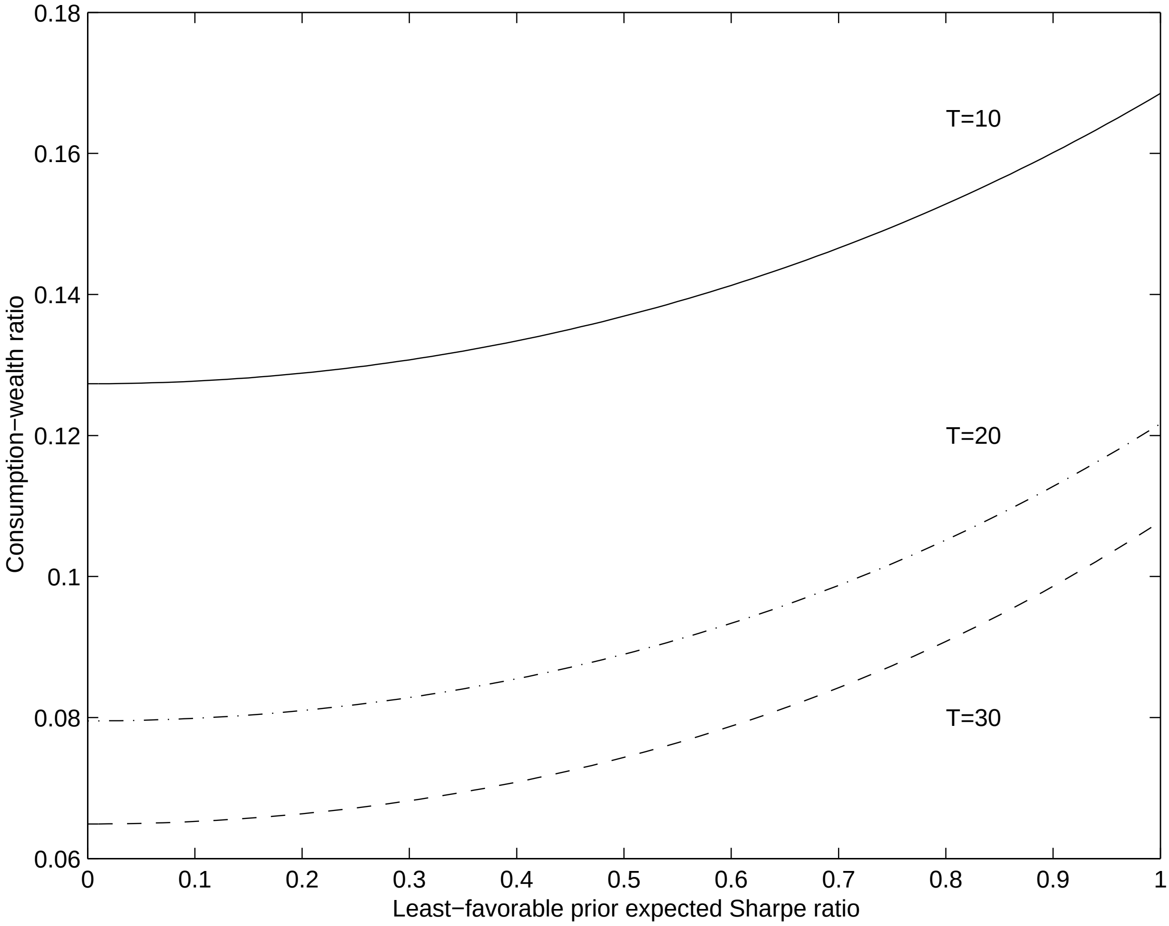
Under the Epstein-Schneider axioms, the investor has a "rectangular" set of priors, which we show in the figure above. The quantity p shown in the figure is the probability that the time-0 return on the risky asset is H. Because of uncertainty aversion, this probability is not fixed: the investor is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

Figure 3: **The Hedging Demand of an Uncertainty-averse Investor with a Preference for Learning**



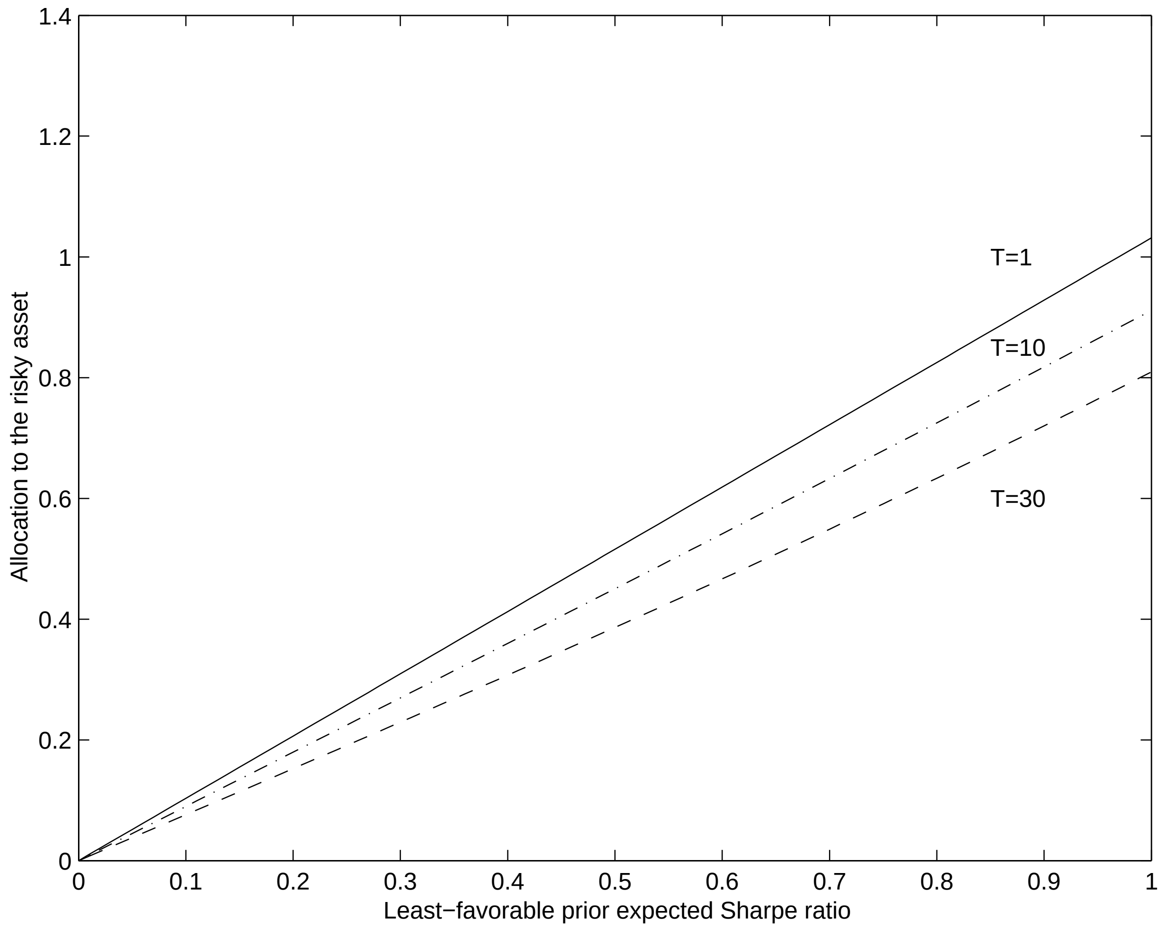
Due to uncertainty aversion, the investor has a set of priors; the least-favorable prior expected Sharpe ratio over that set is on the abscissa above. The investor has power utility over consumption, with a rate of time preference of $\rho = 0.05$ and a constant relative risk aversion of $\gamma = 5$. The riskless rate is $r = 0.05$. The volatility of the risky asset is 0.1902, the maximum likelihood estimate of volatility, quoted at an annual frequency, from data on the CRSP value-weighted market portfolio with distributions (January 1926 to December 2001, monthly). The prior variance of the Sharpe ratio is $\tau^2 = 0.05$, the level of confidence that an investor who had started with a diffuse prior and observed 20 years of data would feel. The investment horizon is T. See Subsection 8.2 for the formulae used to compute these results.

Figure 4: **The Consumption-Wealth Ratio of an Uncertainty-averse Investor with a Preference for Learning**



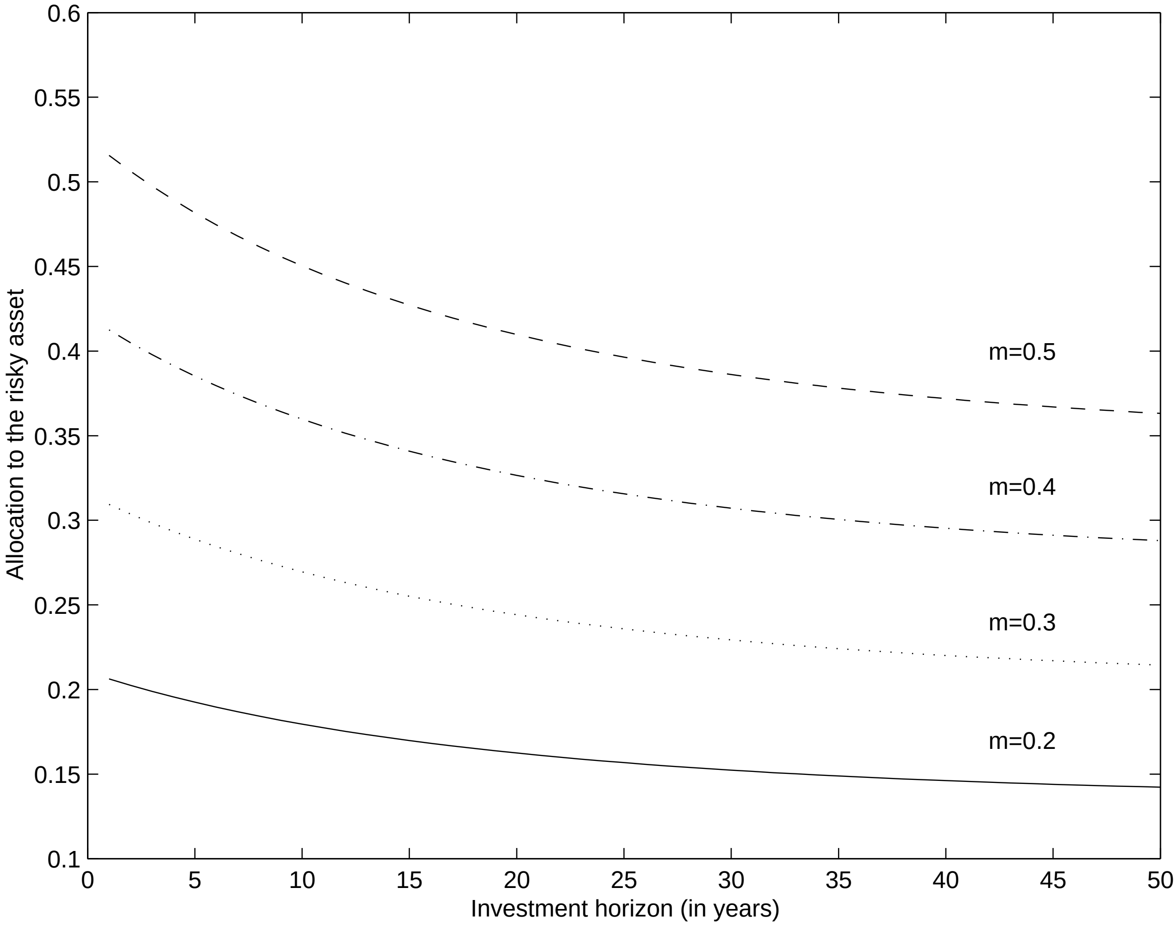
Due to uncertainty aversion, the investor has a set of prior distributions; the least-favorable prior expected Sharpe ratio over that set is on the abscissa above. The investor has power utility over consumption, with a rate of time preference of $\rho = 0.05$ and a constant relative risk aversion of $\gamma = 5$. The riskless rate is $r = 0.05$. The volatility of the risky asset is 0.1902, the maximum likelihood estimate of the volatility, quoted at an annual frequency, from data on the CRSP value-weighted market portfolio with distributions (January 1926 to December 2001, monthly). The prior variance of the Sharpe ratio is $\tau^2 = 0.05$, the level of confidence that an investor who had started with a diffuse prior and observed 20 years of data would feel. The investment horizon is T. See Subsection 8.2 for the formulae used to compute these results.

Figure 5: **An Uncertainty-Averse, Learning Investor's Allocation to the Risky Asset, as a Function of the Least-favorable Prior Expected Sharpe Ratio**



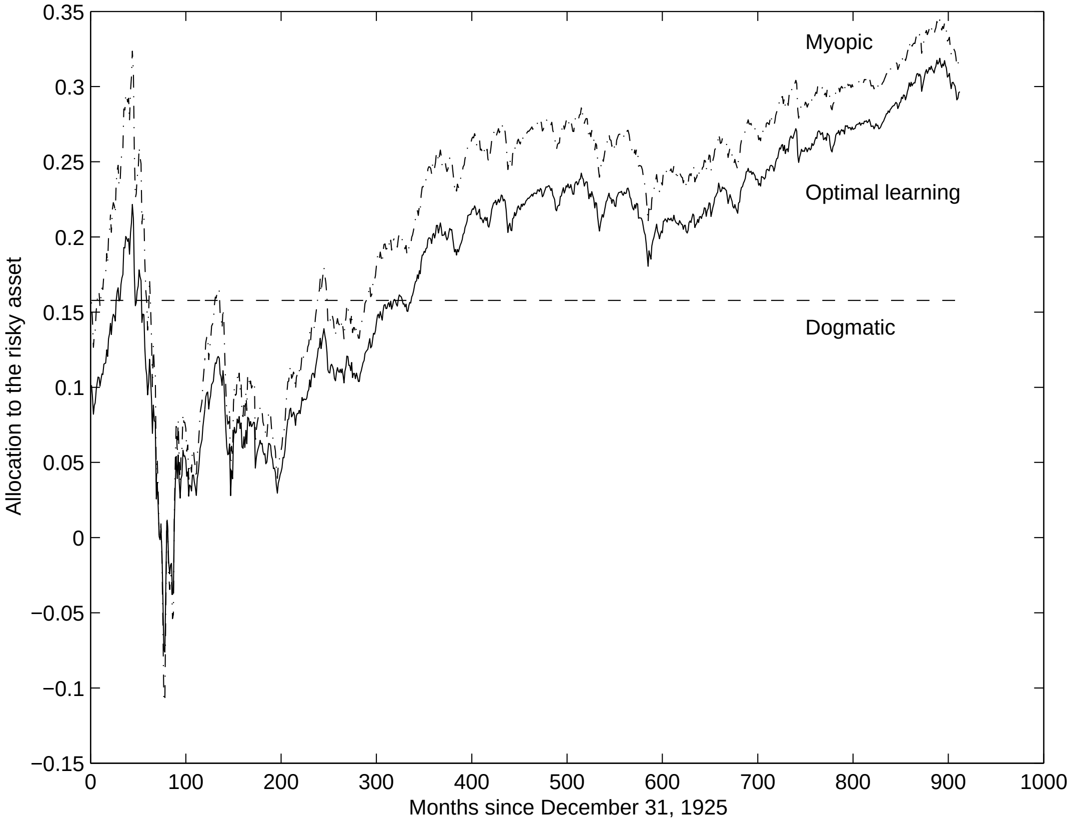
Due to uncertainty aversion, the investor has a set of prior distributions; the least-favorable prior expected Sharpe ratio over that set is on the abscissa above. The investor has power utility over consumption, with a rate of time preference of $\rho = 0.05$ and a constant relative risk aversion of $\gamma = 5$. The riskless rate is $r = 0.05$. The volatility of the risky asset is set to 0.1902, the maximum likelihood estimate of the volatility, quoted at an annual frequency, from data on the CRSP value-weighted market portfolio with distributions (January 1926 to December 2001, monthly). The prior variance of the Sharpe ratio is $\tau^2 = 0.05$, the level of confidence that an investor who had started with a diffuse prior and observed 20 years of data would feel. The investment horizon is T. See Subsection 8.2 for the formulae used to compute these results.

Figure 6: **An Uncertainty-Averse, Learning Investor's Allocation to the Risky Asset, as a Function of the Investment Horizon**



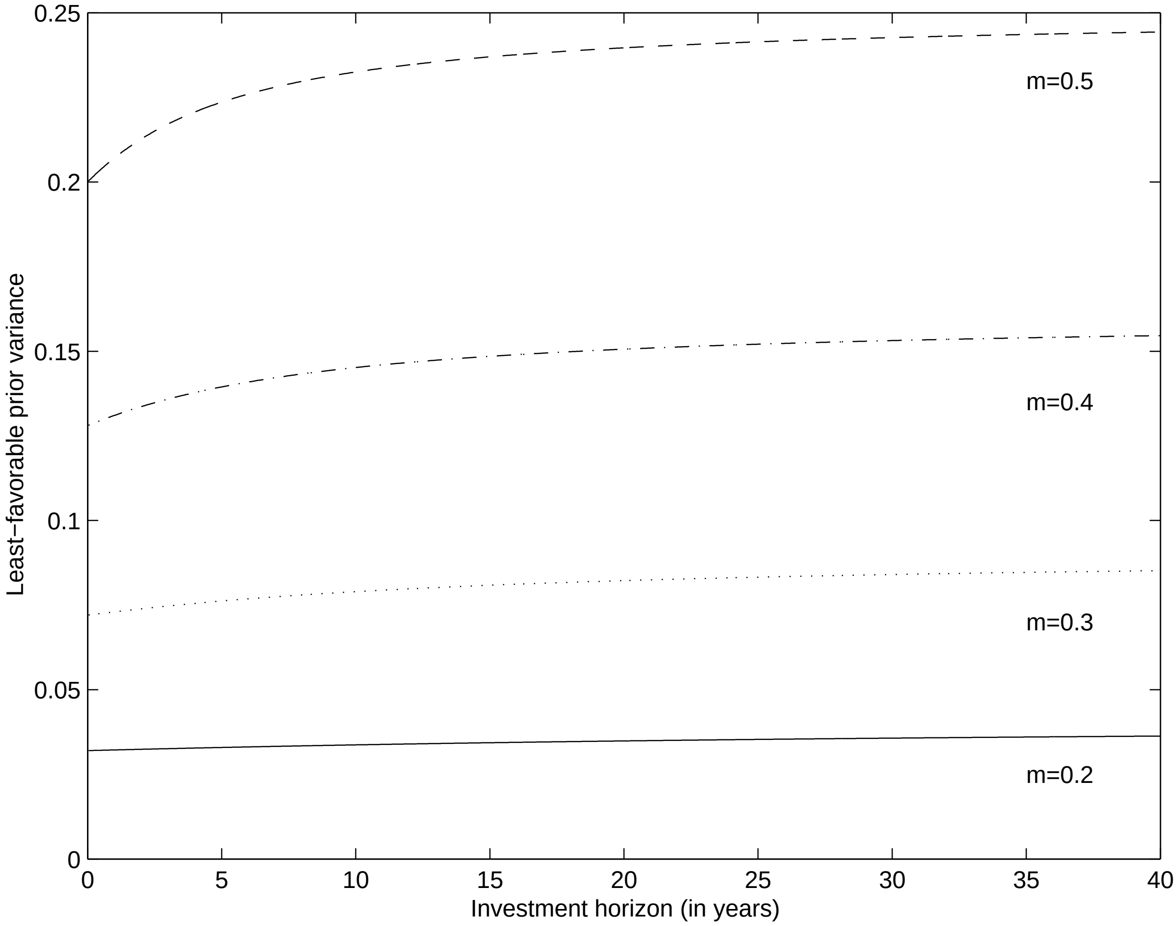
The investor has power utility over consumption, with a rate of time preference of $\rho = 0.05$ and a constant relative risk aversion of $\gamma = 5$. The riskless rate is $r = 0.05$. The volatility of the risky asset is 0.1902, the maximum likelihood estimate of the volatility, quoted at an annual frequency, from data on the CRSP value-weighted market portfolio with distributions (January 1926 to December 2001, monthly). The prior variance of the Sharpe ratio is $\tau^2 = 0.05$, the level of confidence that an investor who had started with a diffuse prior and observed 20 years of data would feel. Due to uncertainty aversion, the investor has a set of prior distributions; the least-favorable prior expected Sharpe ratio over that set is m. See Subsection 8.2 for the formulae used to compute these results.

Figure 7: **The Dynamic Asset Allocation of Uncertainty-Averse Investors with and without a Preference for Learning**



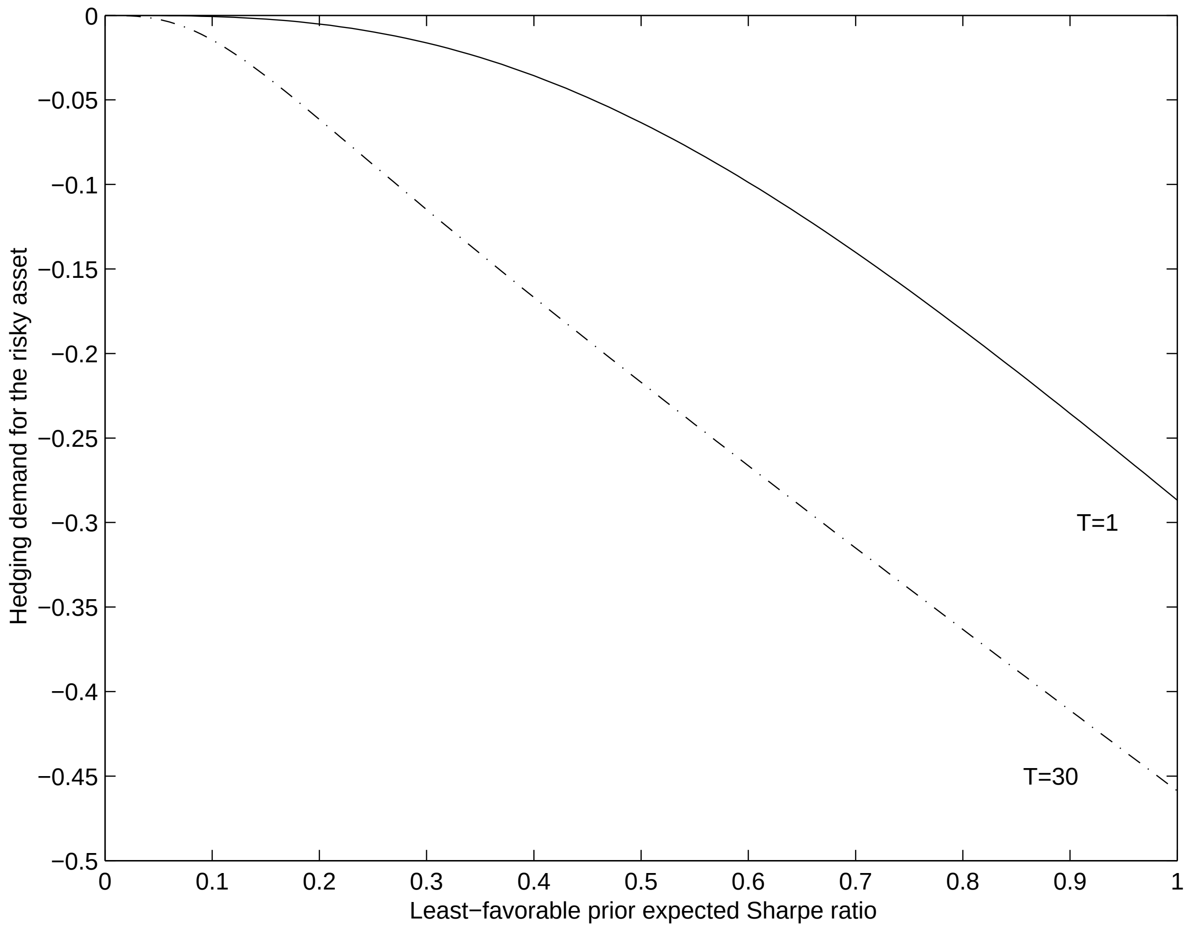
This figure shows the optimal portfolio weight on the risky (and uncertain) asset of one uncertainty-averse investor with a preference for learning and one uncertainty-averse investor who dogmatically expects the worst; see Subsections 8.1 and 8.2, respectively, for details. The uppermost, dotted line represents the myopic demand of the investor who has a preference for learning. Each investor has power utility over intermediate consumption, with constant relative risk aversion $\gamma = 5$. All rates are continuously compounded and quoted in annual units. The rate of time preference is $\rho = 0.05$, and the riskless rate is $r = 0.05$. The volatility of the risky asset is set to 0.1902 (the maximum likelihood estimate of volatility in my data; see below). Each investor's horizon is 100 years. Motivated by the results of Subsection 8.2, I consider a least-favorable prior in which the prior variance of the Sharpe ratio (on December 31, 1925) is set at $\tau^2 = 0.05$, (as though 20 years of data had been observed, beginning from a diffuse prior) and the prior mean of the Sharpe ratio is set at $m = 0.15$ (slightly less that half of the maximum-likelihood estimate of the Sharpe ratio in my data, which is 0.3437) for the investor who has a preference for learning. For the dogmatic investor, Subsection 8.1 shows that we can assume that the least-favorable prior is a point mass on the lowest possible prior Sharpe ratio, which we assume is $m = 0.15$ (for comparability to the learning case). The risky asset is the value-weighted CRSP market portfolio, including distributions. The data span the period from January 1926 to December 2001 at a monthly frequency, for a total of 912 months of data.

Figure 8: **The Least-favorable Prior Variance for an Investor with Utility over Terminal Wealth**



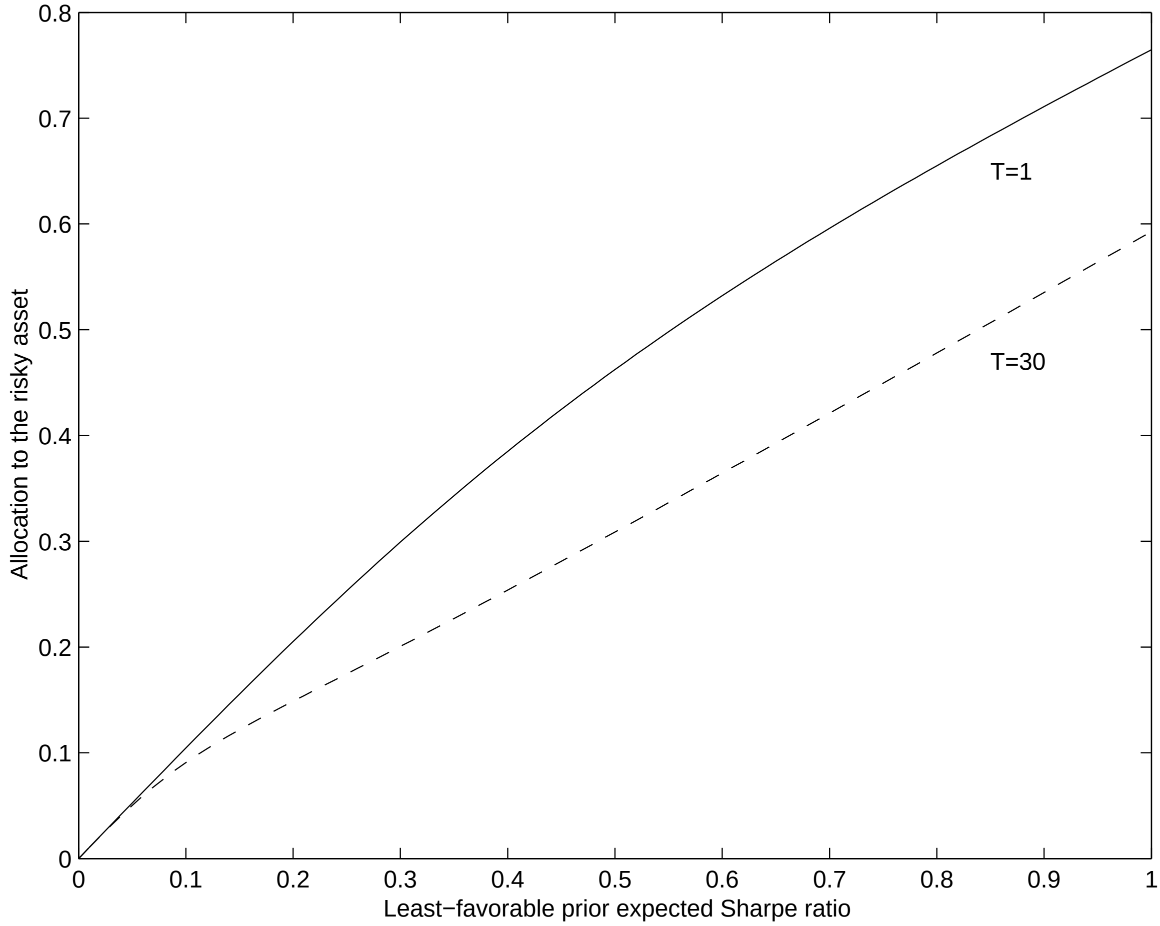
This figure shows the dependence of the least-favorable prior variance of the Sharpe ratio (for an investor with power utility over terminal wealth) on the least-favorable prior expected Sharpe ratio (denoted m in the figure) and on the investment horizon. The investor's constant relative risk aversion is $\gamma = 5$. See Subsection 8.3 for the formula used to create this graph. Proposition 5 makes statements about the limits of the least-favorable prior variance of the Sharpe ratio as $T \to 0$ and as $T \to \infty$; this figure conforms precisely to the statements regarding the limit as $T \to 0$, and shows the beginnings of convergence to the limit as $T \to \infty$.

Figure 9: **The Hedging Demand of an Investor with Utility over Terminal Wealth**



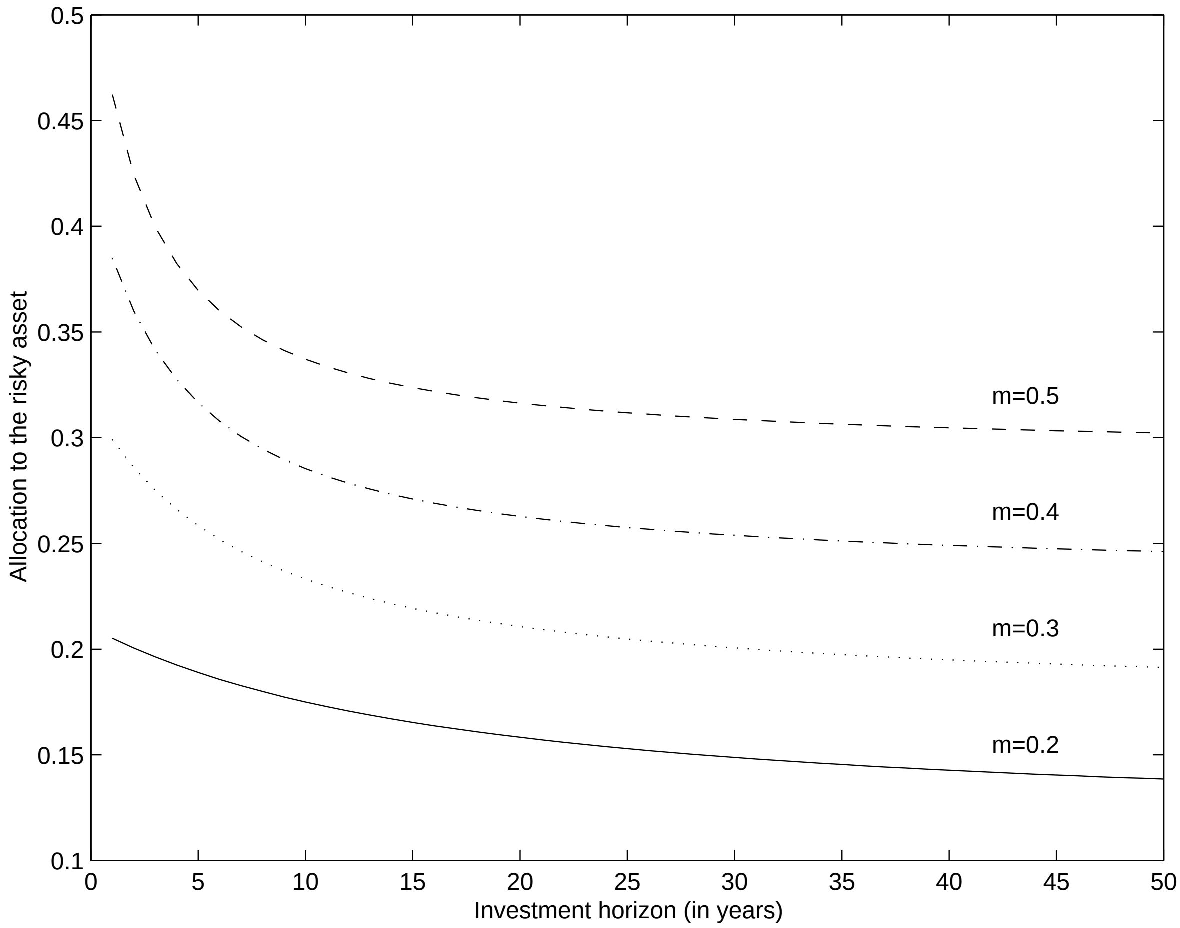
This figure shows the dependence of hedging demand, for an investor with utility over terminal wealth, on the least-favorable prior expected Sharpe ratio and on the investment horizon (T, in years). The investor's constant relative risk aversion is $\gamma = 5$. The riskless rate is $r = 0.05$. The volatility of the risky asset is 0.1902, the maximum likelihood estimate of the volatility, quoted at an annual frequency, from data on the CRSP value-weighted market portfolio with distributions (January 1926 to December 2001, monthly). See Subsection 8.3 for the formula used to create this graph. Note that, for each least-favorable prior expected Sharpe ratio and each investment horizon, the least-favorable prior variance of the Sharpe ratio is calculated and used in creating this figure.

Figure 10: **The Allocation to the Risky Asset of an Investor with Utility over Terminal Wealth, as a Function of the Least-favorable Prior Expected Sharpe Ratio**



This figure shows the dependence of the allocation to the risky asset, for an investor with utility over terminal wealth, on the least-favorable prior expected Sharpe ratio and on the investment horizon (T, in years). The investor's constant relative risk aversion is $\gamma = 5$. The riskless rate is $r = 0.05$. The volatility of the risky asset is 0.1902, the maximum likelihood estimate of the volatility, quoted at an annual frequency, from data on the CRSP value-weighted market portfolio with distributions (January 1926 to December 2001, monthly). See Subsection 8.3 for the formula used to create this graph. Note that, for each least-favorable prior expected Sharpe ratio and each investment horizon, the least-favorable prior variance of the Sharpe ratio is calculated and used in creating this figure.

Figure 11: **The Allocation to the Risky Asset of an Investor with Utility over Terminal Wealth, as a Function of the Investment Horizon**



This figure shows the dependence of the allocation to the risky asset, for an investor with utility over terminal wealth, on the investment horizon and on the least-favorable prior expected Sharpe ratio (denoted m in the figure). The investor's constant relative risk aversion is $\gamma = 5$. The riskless rate is $r = 0.05$. The volatility of the risky asset is 0.1902, the maximum likelihood estimate of the volatility, quoted at an annual frequency, from data on the CRSP value-weighted market portfolio with distributions (January 1926 to December 2001, monthly). See Subsection 8.3 for the formula used to create this graph. Note that, for each least-favorable prior expected Sharpe ratio and each investment horizon, the least-favorable prior variance of the Sharpe ratio is calculated and used in creating this figure.

Appendix to
Learning How to Invest when Returns are Uncertain

Thomas A. Knox
Harvard University

November 20, 2002*

©2002 by Thomas A. Knox

*I am indebted to John Campbell and Gary Chamberlain for guidance and encouragement. I am also grateful for the insights of Brian Hall, Parag Pathak, Jeremy Stein, and James Stock, and for the helpful comments of seminar participants at Harvard University and MIT.

1 Introduction

This Appendix contains proofs of the propositions and theorems stated in the paper "Learning How to Invest when Returns are Uncertain." To avoid confusion between equations in the main text of the paper and equations in this Appendix, I number equations in this Appendix (A1), (A2), etc. Throughout the proofs below, $\mathcal{A} = \{A_1, \ldots, A_k\}$ is a partition of S, and we adopt the notation: given any $A_i \in \mathcal{A}$, let $(f; g)_i = \begin{cases} f & \text{for } s \in A_i, \\ g & \text{for } s \in A_i^C. \end{cases}$.

2 Proofs

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

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

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

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

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

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

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

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

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

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

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

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

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

Lemma 4 *If $|S| < \infty$ and $\Sigma = 2^S$, and if $\mathcal{A}$ and $\mathcal{B}$ are two interlaced partitions, then given any $s, w \in S$, there is a sequence of finite length, $(x_1, \ldots, x_N)$ such that: 1) $x_1 = s$ and $x_N = w$ and 2) $\forall i \in \{1, \ldots, N-1\}$, $\{x_i, x_{i+1}\} \subset B_j$ for some $j \in \{1, \ldots, n\}$ or $\{x_i, x_{i+1}\} \subset A_h$ for some $h \in \{1, \ldots, n\}$.*

Proof of Lemma 4 Suppose not. Then there is some set $M(s)$ of all the $x \in S$ such that a sequence of the sort described in the statement of the lemma exists between s and x. This set is nonempty, since s itself obviously belongs to it. s is in $A_i \in \mathcal{A}$ for some $i \in \{1, \ldots, k\}$ since $\mathcal{A}$ is a partition. Then $A_i \subset M(s)$ must hold, since we can connect s to any $x \in A_i$ using a sequence of length two that satisfies the stated requirements.

More generally, $\forall h \in \{1, \ldots, k\}$, either $A_h \cap M(s) = \phi$ or $A_h \subset M(s)$. To see this, observe that either $A_h \cap M(s) = \phi$, in which case the statement is true, or $A_h \cap M(s) \neq \phi$. If $A_h \cap M(s) \neq \phi$, then there exists some sequence of finite length and satisfying the stated requirement such that $x_1 = s$ and $x_N = y \in A_h$. But then any $z \in A_h$ can be reached by a sequence of the desired sort simply by appending it to the end of the sequence connecting s and $y \in A_h$. Thus, $A_h \subset M(s)$, as claimed.

Define an index set $I = \{h \in \{1, \ldots, k\} : A_h \subset M(s)\}$. Then I is nonempty, since $i \in I$ as shown above. Also, $I \neq \{1, \ldots, k\}$, since if equality held the A_e such that $w \in A_e$ (such an A_e exists because $\mathcal{A}$ is a partition) would be a subset of $M(s)$, implying that a sequence of the stated type exists between s and w and contradicting the assumption.

Since we have shown that I is a nonempty and proper subset of $\{1, \ldots, k\}$, we can apply the intersection portion of the definition of interlaced partitions to obtain a B_j such that $B_j \cap (\cup_{h \in I} A_h) \neq \phi$ and $B_j \cap (\cup_{h \in I} A_h)^C \neq \phi$. Since $(\cup_{h \in I} A_h) \subset M(s)$, we have that $B_j \cap M(s) \neq \phi$. Thus, $B_j \subset M(s)$ (since either $B_j \cap M(s) = \phi$ or $B_j \subset M(s)$; the argument proving this is identical to the one used above to prove the analogous statement for the A_i). But $B_j \cap (\cup_{h \in I} A_h)^C \neq \phi$. Since $B_j \subset M(s)$, this implies that $M(s) \cap (\cup_{h \in I} A_h)^C \neq \phi$. But $M(s) \cap (\cup_{h \in I} A_h)^C = \phi$ by the definition of I and by the separation, shown above, of the A_h in to $A_h \subset M(s)$ and A_h such that $A_h \cap M(s) = \phi$. Thus, we have a contradiction, and our assumption that a sequence does not exist must have been mistaken. Therefore, a sequence of the stated sort exists. □

Lemma 5 *If $Y \sim N(a, b^2)$ and $\beta > -\frac{1}{2b^2}$,*

$$E\left[\exp\left\{-\beta (Y-\alpha)^2\right\}\right] = \frac{1}{\sqrt{2\beta b^2 + 1}} \exp\left\{\frac{-\beta}{2\beta b^2 + 1}(a-\alpha)^2\right\}. \tag{A1}$$

Proof of Lemma 5:

This is an exercise in integration. Multiply the exponential whose expectation we are taking by the normal density given in the statement of the lemma, complete the square on Y inside the exponent, and use the fact that a normal density integrates to one to finish the calculation.

□

Proof of Theorem 1: First we prove that Axiom 4 implies Axiom 1 (each being relative to $\mathcal{A}$). Suppose $f, g \in L_0$ are such that $f(s) = g(s) \quad \forall s \in A_i^C$, and that $f \succ g$. Then A_i cannot be a null set; if it were, then (since f and g are identical on its complement) $f \sim g$ would have to hold by the definition of a null set. Now, $\forall j \neq i,\ f(s) = g(s) \quad \forall s \in A_j$, since the $A_j,\ j \neq i$, partition A_i^C. Thus, by the "consequentialism" portion of Axiom 4, $f \sim_j g \ \ \forall j \neq i$. If we had

$f \precsim_i g$, then we would have $f \precsim_l g \ \ \forall l \in \{1, \ldots, k\}$ (since $f \sim_l g \Rightarrow f \precsim_l g$ by definition), so by the "consistency" portion of Axiom 4 we would have $f \precsim g$, which does not hold. Thus, we must have $f \succ_i g$. Now, given h such that h is constant on A_i, we have by definition that $\exists y \in Y$ such that $h(s) = y \ \ \forall s \in A_i$. By the "consequentialism" portion of Axiom 4, and letting $l \in L_c$ be such that $l(s) = y \ \ \forall s \in S$, we have that $l \sim_i h$, since l and h are equal (with value y) at each element of S. Now, by the "multiple priors" portion of Axiom 4, and the "certainty independence" portion of Axiom 2, $\forall \alpha \in (0,1)\,,\ \alpha f + (1-\alpha) l \succ_i \alpha g + (1-\alpha) l$. Since $(\alpha f + (1-\alpha) l)(s) = (\alpha f + (1-\alpha) h)(s) \quad \forall s \in A_i$, the "consequentialism" portion of Axiom 4 implies that $\alpha f + (1-\alpha) l \sim_i \alpha f + (1-\alpha) h$. Likewise, since $(\alpha g + (1-\alpha) l)(s) = (\alpha g + (1-\alpha) h)(s) \quad \forall s \in A_i$, the "consequentialism" portion of Axiom 4 implies that $\alpha g + (1-\alpha) l \sim_i \alpha g + (1-\alpha) h$. The transitivity of $\succsim_i$ is implied by Axiom 4 and the "weak order" portion of Axiom 2. By this transitivity, then,

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

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

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

From the above, we have that $\forall \alpha \in (0,1)\,,\ f \succ g \Rightarrow \alpha f + (1-\alpha) h \succ \alpha g +$

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

Given $\alpha\in(0,1)$, suppose that f,g,h are as described in the previous section of the proof, and that $\alpha f+(1-\alpha)\,h\succ\alpha g\,(1-\alpha)\,h$. The steps to follow are quite similar to those above, but we include them for the sake of completeness. A_i cannot be a null set; if it were, then (since $\alpha f+(1-\alpha)\,h$ and $\alpha g+(1-\alpha)\,h$ are identical on its complement) $\alpha f+(1-\alpha)\,h\sim\alpha g+(1-\alpha)\,h$ would have to hold by the definition of a null set. Now, $\forall j\neq i,\ (\alpha f+(1-\alpha)\,h)\,(s)=(\alpha f+(1-\alpha)\,h)\,(s)\quad\forall s\in A_j$, since the $A_j,\ j\neq i$, partition A_i^C. Thus, by the "consequentialism" portion of Axiom 4, $\alpha f+(1-\alpha)\,h\sim_j\alpha g+(1-\alpha)\,h\ \ \forall j\neq i$. If we had $\alpha f+(1-\alpha)\,h\precsim_i\alpha g+(1-\alpha)\,h$, then we would have $\alpha f+(1-\alpha)\,h\precsim_l\alpha g+(1-\alpha)\,h\quad\forall l\in\{1,\ldots,k\}$ (since $\alpha f+(1-\alpha)\,h\sim_l\alpha g+(1-\alpha)\,h\Rightarrow\alpha f+(1-\alpha)\,h\precsim_l\alpha g+(1-\alpha)\,h$ by definition), so by the "consistency" portion of Axiom 4 we would have $\alpha f+(1-\alpha)\,h\precsim\alpha g+(1-\alpha)\,h$, which does not hold. Thus, we must have $\alpha f+(1-\alpha)\,h\succ_i\alpha g+(1-\alpha)\,h$. We have by definition that $\exists y\in Y$ such that $h\,(s)=y\ \ \forall s\in A_i$. Letting $l\in L_c$ be such that $l\,(s)=y\ \ \forall s\in S$, we have that $(\alpha f+(1-\alpha)\,l)\,(s)=(\alpha f+(1-\alpha)\,h)\,(s)\quad\forall s\in A_i$. Thus, the "consequentialism" portion of Axiom 4 implies that $\alpha f+(1-\alpha)\,l\sim_i\alpha f+(1-\alpha)\,h$. Likewise, $(\alpha g+(1-\alpha)\,l)\,(s)=(\alpha g+(1-\alpha)\,h)\,(s)\quad\forall s\in A_i$. Thus, the "consequentialism" portion of Axiom 4 implies that $\alpha g+(1-\alpha)\,l\sim_i\alpha g+(1-\alpha)\,h$. The transitivity of $\succsim_i$ is implied by Axiom 4 and the "weak order" portion of Axiom 2. By this transitivity and the above observations,

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

implies that $\alpha f + (1-\alpha)\, l \succ_i \alpha g + (1-\alpha)\, l$. By the "multiple priors" portion of Axiom 4, and the "certainty independence" portion of Axiom 2, $\alpha f + (1-\alpha)\, l \succ_i \alpha g + (1-\alpha)\, l \Rightarrow f \succ_i g$. We therefore have that $f \succ_i g$.

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

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

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

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

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

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

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

Now we verify that $\succsim_i$ satisfies each of the portions of Axiom 3. These follow because $\succsim$ satisfies Axiom 3 and by the definition of $\succsim_i$. First we show that $\succsim_i$ is a weak order (that is, that $\succsim_i$ is complete and transitive). Suppose that $f \succsim_i g$

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

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

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

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

Consider the "uncertainty aversion" portion of Axiom 3. Suppose that $f, g \in L_0$ satisfy $f \sim_i g$. Then, by definition of $\succsim_i$, $\exists h \in L_0$ such that $(f;h)_i \sim (g;h)_i$.

Since $\succsim$ satisfies uncertainty aversion (by Axiom 3), this implies that $\alpha\left(f;h\right)_i + \left(1-\alpha\right)\left(g;h\right)_i \succsim \left(g;h\right)_i \;\; \forall \alpha \in (0,1)$. By the definition of $\succsim_i$, this, in turn, implies that $\alpha f + (1-\alpha)\, g \succsim_i g \;\; \forall \alpha \in (0,1)$, confirming that $\succsim_i$ satisfies the uncertainty aversion property.

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

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

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

for all $i \in \{1, \ldots, k\}$. Then we have, for $i \in \{2, \ldots, k\}$, $\left(g;h_i\right)_i(s) = h_i\left(s\right) = \left(f;h_{i-1}\right)_{i-1}(s)$. This can be seen by considering the values of each of the above expressions on each $A_j \in \mathcal{A}$. Now, since $\left(f;h_i\right)_i \succsim \left(g;h_i\right)_i$ for each $i \in \{1, \ldots, k\}$, we can use the equality above to conclude that $h_{i+1} = \left(f;h_i\right)_i \succsim \left(g;h_i\right)_i = h_i$

for $i \in \{1, \ldots, k-1\}$, and thus that $h_{i+1} \succsim h_i$ for $i \in \{1, \ldots, k-1\}$. Applying transitivity repeatedly, this implies that $h_k \succsim h_1$. By definition, $h_1(s) = g(s) \ \ \forall s \in S$. We also have $f(s) = (f; h_k)_k(s) \ \ \forall s \in S$, $(f; h_k)_k \succsim (g; h_k)_k$, and $(g; h_k)_k(s) = h_k(s) \forall s \in S$. Combining these facts, we obtain $f \succsim h_k$. A final application of transitivity yields $f \succsim g$, and thus verifies the main part of the "consistency" portion of Axiom 3.

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

□

Proof of Theorem 2:

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

Now suppose that condition (2) holds. Apply Theorem 1 to prove that Axiom 1 holds. Then Axiom 2, combined with the consequentialism property of the conditional preference relations, implies Axiom 5 directly. This shows that (2) ⇒

(1), and combining this with the above yields (1) $\Leftrightarrow$ (2).

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

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

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

We must verify that we may take $u_i = u$ w.l.o.g. This is implied directly by Axiom 5: since all of the conditional preference relations agree on the roulette lotteries L_c, and since any preference relation on L_c implies a u that is unique up to a positive affine transformation, the u_i differ by at most a positive affine transformation. Since any positive affine transformation of u_i represents preferences over L_c, we can let $u_i = u_1 \;\; \forall i \in \{2, \ldots, k\}$ w.l.o.g.

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

$\int_{s\in A_i} f(s)\, d(P_i^*(f))(s)$. Then

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Suppose not; then $\exists j \in \{1, \ldots, k\}$ and $P \in \mathcal{P}$ such that $\mathcal{P}\left(\mathcal{A}_{|}\right) = \prime$. By the non-degeneracy condition, there exist two roulette lotteries $l, q \in L_c$ such that $l \succ_j q$. Since we can select any positive affine transformation of u in the representation result, and since $u\left(l\right) > u\left(q\right)$, we can w.l.o.g. choose u such that $u\left(l\right) > 0$ and $u\left(q\right) = 0$. We do so. Obviously, $q \sim_i q$ for all $i \in \{1, \ldots, k\}$. Consider the act

$f = (l;q)_j$. Using our selection of u to evaluate its utility, we have:

$$\begin{aligned} & \min_{P \in \mathcal{P}} \left\{ \int_{s \in S} u(f(s))\, dP(s) \right\} \\ = & \min_{P \in \mathcal{P}} \left\{ \sum_{i=1, i \neq j}^{k} u(q) P(A_i) + u(l) P(A_j) \right\} \\ = & \min_{P \in \mathcal{P}} \{ u(l) P(A_j) \} \\ = & \ 0, \end{aligned}$$

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

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

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

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

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

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

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

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

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

Also,

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

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

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

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

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

□

Proof of Theorem 3:

We first prove (2) $\Rightarrow$ (1). Full independence obviously implies restricted independence with respect to any $F \subset S$, and since $|S| \geq 3$, we can construct a pair of interlaced partitions as follows. Label the elements of S so that $S = \{x_1, \ldots, x_N\}$. If N is odd, define

$$\mathcal{A} = \{\{x_1, x_2\}, \{x_3, x_4\}, \ldots, \{x_{N-2}, x_{N-1}\}, \{x_N\}\}$$

and

$$\mathcal{B} = \{\{x_1\}, \{x_2, x_3\}, \{x_4, x_5\}, \ldots, \{x_{N-1}, x_N\}\}.$$

If N is even, define

$$\mathcal{A} = \{\{x_1, x_2\}, \{x_3, x_4\}, \ldots, \{x_{N-1}, x_N\}\}$$

and

$$\mathcal{B} = \{\{x_1\}, \{x_2, x_3\}, \{x_4, x_5\}, \ldots, \{x_{N-2}, x_{N-1}\}, \{x_N\}\}.$$

In either case, $\mathcal{A}$ and $\mathcal{B}$ are easily seen to be interlaced, and restricted independence holds with respect to each because full independence holds. Thus, (2) $\Rightarrow$ (1) is proven.

To show that (1) $\Rightarrow$ (2), first apply Theorem 2 to obtain the fact that $\succsim$ is represented by a set of priors, $\mathcal{P}$, that is prismatic with respect to both $\mathcal{A}$ and $\mathcal{B}$. We will use the fact that $\mathcal{P}$ is prismatic with respect to both $\mathcal{A}$ and $\mathcal{B}$ to show that it must be a singleton, which will then imply full independence. Denote the set of conditionals of $P \in \mathcal{P}$ conditioned on $A_i,\ i \in \{1, \ldots, k\}$ by $\mathcal{C}_i,\ i \in \{1, \ldots, k\}$, and denote the set of conditionals of $P \in \mathcal{P}$ conditioned on $B_j,\ j \in \{1, \ldots, n\}$ by $\mathcal{D}_j,\ j \in \{1, \ldots, n\}$.

Our first step will be to show that $\mathcal{C}_i,\ i \in \{1, \ldots, k\}$ and $\mathcal{D}_j,\ j \in \{1, \ldots, n\}$ are all singletons. Suppose not. Then there is some nonsingleton $\mathcal{C}_i$ or $\mathcal{D}_j$; assume w.l.o.g. that it is some $\mathcal{C}_i$. Since $\mathcal{C}_i$ is nonsingleton, A_i must have at least two distinct elements and $\exists P, Q \in \mathcal{C}_i$ such that $\exists x, y \in A_i$ with $x \neq y$ and $P(\{x\}) \neq Q(\{x\}), P(\{y\}) \neq Q(\{y\})$. Then, since $\mathcal{A}$ and $\mathcal{B}$ are interlaced, we have that $\exists B_j, B_h \in \mathcal{B}$ such that $B_j \cap A_i = \{x\}$ and $B_h \cap A_i = \{y\}$, and at least one of B_j, B_h is nonsingleton. Assume w.l.o.g. that B_j is nonsingleton.

Consider $f \in L_0$ such that $u(f(s)) = 1\{s \in A_i\}$. Such an f can be constructed due to the nondegeneracy condition; see Gilboa and Schmeidler (1989), page 151. Let us characterize the set of probability distributions that achieve

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

We will denote this set $\mathcal{P}^{LF}$. Since $\mathcal{P}$ is prismatic with respect to $\mathcal{A}$,

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

which shows that the conditional distributions of s given A_h are irrelevant in the minimization problem. Because $\mathcal{P}$ is prismatic with respect to $\mathcal{A}$, this means that if $\sum_{h=1}^{k} P_h(\cdot| A_h) P(A_h) \in \mathcal{P}^{LF}$ for some $P \in \mathcal{P}, P_1 \in \mathcal{C}_1, \ldots, P_k \in \mathcal{C}_k$, then $\sum_{h=1}^{k} Q_h(\cdot| A_h) P(A_h) \in \mathcal{P}^{LF}$ for any $Q_1 \in \mathcal{C}_1, \ldots, Q_k \in \mathcal{C}_k$ (note, however, that the marginal probabilities of the A_h do matter; only the conditionals can be freely selected).

In contrast, since $u(f(x)) = 1$ and $u(f(y)) = 0 \quad \forall y \in B_j$ such that $y \neq x$, we have that $\forall P, Q \in \mathcal{P}^{LF}, P(\{x\}| B_j) = Q(\{x\}| B_j)$. Otherwise, $\mathcal{P}^{LF}$ would include P, Q such that, w.l.o.g., $P(\{x\}| B_j) < Q(\{x\}| B_j)$. But, since $\mathcal{P}$ is prismatic with respect to $\mathcal{B}$, we would then be free to alter Q by replacing $Q(\{x\}| B_j)$ by $P(\{x\}| B_j)$ without changing Q in any other way. This would lower the Q-expected utility of f, contradicting the fact that Q already minimized the expected utility of f. Thus, all distributions in $\mathcal{P}^{LF}$ assign the same probability to the occurrence of x conditional on B_j.

Now let us compute the conditional probability $P(\{x\}| B_j)$ for an arbitrary $P \in$

$\mathcal{P}^{LF}$ using the partition $\mathcal{A}$:

$$\begin{aligned}
& P\left(\{x\}|\, B_j\right) \\
= \;& \frac{P\left(\{x\} \cap B_j\right)}{P\left(B_j\right)} \\
= \;& \frac{P\left(\{x\}\right)}{P\left(B_j\right)} \\
= \;& \frac{P\left(\{x \cap A_i\}\right)}{P\left(B_j\right)} \\
= \;& \frac{P\left(\{x|\, A_i\}\right) P\left(A_i\right)}{P\left(B_j\right)} \\
= \;& \frac{P\left(\{x|\, A_i\}\right) P\left(A_i\right)}{\sum_{h=1}^{k} P\left(B_j|\, A_h\right) P\left(A_h\right)} \\
= \;& \frac{P\left(\{x|\, A_i\}\right) P\left(A_i\right)}{\sum_{h=1,h\neq i}^{k} P\left(B_j|\, A_h\right) P\left(A_h\right) + P\left(\{x\}|\, A_i\right) P\left(A_i\right)},
\end{aligned}$$

where the first equality is by the definition of conditional probability, the second equality is due to the fact that $\{x\} \cap B_j = \{x\}$, the third equality is due to the fact that $\{x\} \cap A_i = \{x\}$, the fourth equality is by the standard separation of a joint probability into a conditional and a marginal, the fifth inequality is by exhaustion and by the same standard decomposition of a joint probability into a conditional and a marginal (this time performed for each term in the sum), and the fifth equality separates the term concerned with A_i, and then recognizes that $P\left(B_j|\, A_i\right) = P\left(\{x\}|\, A_i\right)$ because $B_j \cap A_i = \{x\}$.

Note that we can (and do) assume $\sum_{h=1,h\neq i}^{k} P\left(B_j|\, A_h\right) P\left(A_h\right) > 0$ w.l.o.g., as we now show. B_j has at least two elements by construction. Thus, $\exists z \in B_j$ such that $z \neq x$, and $\exists A_g \in \mathcal{A}$ such that $A_g \cap B_j = \{z\}$. Furthermore, the assumption that no event is null implies that $P\left(A_g\right) > 0 \;\; \forall P \in \mathcal{P}$ (as shown in the proof of Theorem 2 above; otherwise, the "if, in addition" clause of the consistency condition in Axiom 4 is violated). Another implication of non-nullity is that $\exists P_g \in \mathcal{C}_g$ such that $P_g\left(\{z\}|\, A_g\right) > 0$. If not, then $\{z\}$ would evidently be null: it would occur with probability zero for every $P \in \mathcal{P}$. Since we are free to select any conditionals from the sets $\mathcal{C}_h,\ i \in \{1, \ldots, k\}$, we can and do select $P_g \in \mathcal{C}_g$ w.l.o.g. Note that $A_g \cap B_j = \{z\}$ implies $P_g(B_j|\, A_g) = P_g(\{z\}|\, A_g) > 0$. But then $P(B_j|\, A_g)\, P(A_g) > 0$, and all the other terms are nonnegative, so the sum is positive.

Now, we have

$$\begin{aligned} & P(\{x\}|\, B_j) \\ = & \frac{P(\{x|\, A_i\})\, P(A_i)}{\sum_{h=1,h\neq i}^{k} P(B_j|\, A_h)\, P(A_h) + P(\{x\}|\, A_i)\, P(A_i)} \\ \neq & \frac{Q(\{x|\, A_i\})\, P(A_i)}{\sum_{h=1,h\neq i}^{k} P(B_j|\, A_h)\, P(A_h) + Q(\{x\}|\, A_i)\, P(A_i)} \\ = & P(\{x\}|\, B_j), \end{aligned}$$

where the first equality is proven above, and the inequality is due to the facts that, by our assumption, $\exists P, Q \in \mathcal{C}_i$ such that $P(\{x\}) \neq Q(\{x\})$ (the conditioning is added explicitly above because P and Q are conditional distributions, by definition) and that, as we showed above, we can assume w.l.o.g. that $\sum_{h=1,h\neq i}^{k} P(B_j|\, A_h)\, P(A_h) > 0$. The last equality follows from the facts that, as demonstrated above, we are free to alter the $\mathcal{C}_i$-component of $P \in \mathcal{P}^{LF}$ to any $Q \in \mathcal{C}_i$ while remaining in $\mathcal{P}^{LF}$, as long as the marginals are unaffected and that, again as demonstrated above, $P(\{x\}|\, B_j)$ is constant over all of $\mathcal{P}^{LF}$.

But this is a contradiction, since it states that $P(\{x\}|\, B_j) \neq P(\{x\}|\, B_j)$, so it must be that our initial assumption was false. Therefore every $\mathcal{C}_i,\ i \in \{1, \ldots, k\}$ and every $\mathcal{D}_j,\ j \in \{1, \ldots, n\}$ is a singleton. Thus, we can speak of the unique conditional probabilities $P(\{s\}|\, A_h)$ and $P(\{s\}|\, B_e)$ for all $s \in S,\ \ h \in \{1, \ldots, k\}$, and $e \in \{1, \ldots, n\}$. If $\{s\} \cap B_e \neq \phi$, then $P(\{s\}|\, B_e) > 0$ (otherwise, $\{s\}$ is a null event, which we have assumed is not true). Likewise, if $\{s\} \cap A_h \neq \phi$, then $P(\{s\}|\, A_h) > 0$ by non-nullity. Recall that $P(A_h) > 0$ and

$P(B_e) > 0$ for all $h \in \{1,\ldots,k\}$, $e \in \{1,\ldots,n\}$, and $P \in \mathcal{P}$. Combining the positivity of conditionals and marginals, we have that $P(\{s\}) > 0$ for all $s \in S$ and for all $P \in \mathcal{P}$.

We now complete the proof by showing that, in fact, $\mathcal{P}$ is a singleton. Suppose not. Then $\exists P, Q \in \mathcal{P}$ such that $P \neq Q$. This implies that $\exists P, Q \in \mathcal{P}$ such that, for some $s, w \in S$ we have $\frac{P(\{s\})}{P(\{w\})} \neq \frac{Q(\{s\})}{Q(\{w\})}$. (Otherwise, $P = Q$ for every $P, Q \in \mathcal{P}$ because every such P and Q would be finitely additive and would assign the same probability to every singleton in a finite set.) Note that, by the positivity shown in the preceding paragraph, the fractions considered are strictly positive and finite.

Apply Lemma 4 to conclude that there exists a sequence $(x_1,\ldots,x_N)$, where $x_1 = s$, $x_N = w$, and $\forall i \in \{1,\ldots,N-1\}$, either $\{x_i, x_{i+1}\} \subset A_h$ for some $h \in \{1,\ldots,k\}$ or $\{x_i, x_{i+1}\} \subset B_e$ for some $e \in \{1,\ldots,n\}$. Suppose w.l.o.g. that $\{x_i, x_{i+1}\} \subset A_h$ for some $h \in \{1,\ldots,k\}$. Then for any $M \in \mathcal{P}$ and for any $i \in \{1,\ldots,N-1\}$,

$$\begin{aligned}
\frac{M(\{x_i\})}{M(\{x_{i+1}\})} &= \frac{M(\{x_i\}|\, A_h)\, M(A_h)}{M(\{x_{i+1}\}|\, A_h)\, M(A_h)} \\
&= \frac{M(\{x_i\}|\, A_h)}{M(\{x_{i+1}\}|\, A_h)}, \\
&= c_i \in (0,\infty),
\end{aligned}$$

where the first equality follows from the usual conditional-marginal decomposition of the joint after the observations that $\{x_i\} \cap A_h = \{x_i\}$, and likewise for x_{i+1}, so that $M(\{x_i\}) = M(\{x_i\} \cap A_h)$, and likewise for x_{i+1}. Note again that the fractions are well-defined due to the positivity of any $M \in \mathcal{P}$ on every singleton set. The second equality follows by cancelling the (positive, as shown above) marginal, and the third equality, in which c_i is a constant over all $M \in \mathcal{P}$, follows from the fact that $M(\cdot|\, A_h) \in \mathcal{C}_h$, which is a singleton as shown above, so that

$M\left(\cdot\,|\, A_h\right)$ must be the same for any $M \in \mathcal{P}$. Again, the ratio of the conditionals is strictly positive and finite due to the positivity fact shown earlier in the proof.

Iterating the above step, we see that, for any $M \in \mathcal{P}$,

$$\begin{aligned} M\left(\{s\}\right) &= M\left(\{x_1\}\right) \\ &= c_1 M\left(\{x_2\}\right) \\ &= c_1 c_2 M\left(\{x_3\}\right) \\ &\vdots \\ &= \left(\prod_{i=1}^{N-1} c_i\right) M\left(\{x_N\}\right) \\ &= \left(\prod_{i=1}^{N-1} c_i\right) M\left(\{w\}\right), \end{aligned}$$

where the first equality is by the definition of the sequence, the second equality is by one application of the argument in the preceding paragraph, the third equality is by another application, ..., the penultimate equality is by one more application, and the final equality is, again, by the definition of the sequence. (More formally, an induction argument could be used to establish the equality above.)

This shows that $\frac{P(\{s\})}{P(\{w\})} = \frac{Q(\{s\})}{Q(\{w\})}$. This contradicts the assumption that $\exists P, Q \in \mathcal{P}$ such that $P \neq Q$, so that assumption must have been false. We have now completed the proof by demonstrating that (1) $\Rightarrow$ (2).

□

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

Proof of Theorem 4: First observe that Axiom 3 is assumed in each of Theorems 1, 2, and 3. Thus we may apply Proposition 1 to conclude that there is a von Neumann-Morgenstern utility function v that represents $\succsim$ in comparing roulette lotteries. v evidently maps adapted acts to functions from S to Y^U, since if $f \in \mathcal{H}$, then $v(f(s))$ is in $Y^U = X^U$ for every $s \in S$. If we can show that v is an isomorphism *on indifference classes* of $f \in \mathcal{H}$, then we can define a new preference relation $\succsim^V$ by $\forall f, g \in \mathcal{H},\ f \succsim g \Leftrightarrow v(f) \succsim^V v(g)$ and, since v is an isomorphism on indifference classes, the new preference relation $\succsim^V$ will be well-defined. By the definition of $Y^U = X^U$ (and the mixture linearity of v), we can see that v is onto. To show that it is one-to-one *as a mapping of indifference classes*, note that $v(f) = v(g) \quad \forall s \in S$ implies that $f(s) \sim g(s) \quad \forall s \in S$, since v represents $\succsim$ on roulette lotteries (and the constant act with value $f(s)$ is a roulette lottery). By monotonicity, then, $f \sim g$. This implies that v is one-to-one as a mapping of indifference classes.

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

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

Proof of Proposition 3: We shall follow the techniques of Cvitanič, Lazrak, Martellini, and Zapatero (2002) and of Wachter (2002) in order to solve the Bayesian problem of the proposition. First, note that the volatility of risky-asset returns, σ, is presumed known (and even if not known, in continuous time it would be learned immediately). Since the riskless rate is assumed to be constant, and the prior on μ is Gaussian, we may equivalently assume a Gaussian prior on the Sharpe ratio $X = \frac{\mu - r}{\sigma}$. If the prior on the Sharpe ratio has mean m and variance τ^2, then $m = \frac{\lambda - r}{\sigma}$ and $\tau^2 = \frac{\nu^2}{\sigma^2}$, in terms of the parameters of the original prior on μ. It is considerably more convenient to deal with the Sharpe ratio, and we shall do so throughout the remainder of the proof.

Define

$$w_t^* \quad = \quad w_t + Xt, \tag{A2}$$

so that w_t^* is the risk-neutral Brownian motion. It will be crucial for us, since it generates the price filtration, which is all of the information the investor has. We shall define $\overline{X}_s = \frac{\tau^2}{1+\tau^2 s}\left(w_t^* - \frac{1}{\tau^2} m\right)$, which is the usual Bayesian conditional mean of the Sharpe ratio. Filtering theory provides a device using $\overline{X}_s$, the *innovation process*, which we can use to greatly simplify the problem (Liptser and Shiryayev (2001)). The innovation process is

$$\overline{w}_t \quad = \quad w_t - \int_0^t \left(\overline{X}_s - X\right) ds \tag{A3}$$

$$\quad = \quad w_t^* - \int_0^t \overline{X}_s. \tag{A4}$$

The budget constraint is:

$$dW_t = (rW_t - c_t)\,dt + \varphi_t\left(\sigma X dt + \sigma dw_t\right) \tag{A5}$$

$$= (rW_t - c_t)\,dt + \varphi_t\left(\sigma \overline{X}_t dt + \sigma d\overline{w}_t\right), \tag{A6}$$

so we can put everything in terms of the innovation process. Thus, like the Bayesian investor, we can view the problem as a full-information problem with a stochastic Sharpe ratio, $\overline{X}_s$. As pointed out by Cvitaniȅ, Lazrak, Martellini, and Zapatero (2002), we can use any full-information method for solving the investor's problem just by replacing w_t with $\overline{w}_t$ and X by $\overline{X}_t$. We will denote expectations taken under the innovation-process measure by "$\overline{E}\left[\cdot\right]$."

We will make use of the "martingale method" for solving consumption and portfolio choice problems, due to Cox and Huang (1989) and Cox and Huang (1991). To use the martingale method, we must first check Novikov's condition (which should hold with respect to the investor's subjective expectation); Novikov's condition is easily seen to hold in the current problem, and we omit its verification. Our treatment will closely parallel that of Cvitaniȅ, Lazrak, Martellini, and Zapatero (2002), though they do not consider intermediate consumption. Consider the risk-neutral density process

$$\overline{Z}_t = \exp\left(-\int_0^t \overline{X}_s d\overline{w}_s - \frac{1}{2}\int_0^t \overline{X}_s^2 ds\right), \tag{A7}$$

and the related state-price deflator

$$\overline{\xi}_t = e^{-rt}\overline{Z}_t. \tag{A8}$$

Note that the state-price deflator has the property that

$$\overline{E}_t\left[\overline{\xi}_s S_s\right] = \overline{\xi}_t S_t \;\; \forall s \geq t. \tag{A9}$$

A routine application of the martingale-method machinery yields the optimal consumption level (in terms of $\overline{\xi}_t$):

$$c_t^o = l^{-\frac{1}{\gamma}} e^{-\frac{1}{\gamma}\rho t} \overline{\xi}_t^{-\frac{1}{\gamma}}, \tag{A10}$$

where l is a Lagrange multiplier whose value is chosen so that the budget constraint is exactly satisfied. The martingale method also gives us the optimal level of wealth:

$$\begin{aligned} W_t^o &= \frac{1}{\overline{\xi}_t} \overline{E}_t \left[\int_t^T \overline{\xi}_s c_s^o ds \right] && \text{(A11)} \\ &= l^{-\frac{1}{\gamma}} \frac{1}{\overline{\xi}_t} \overline{E}_t \left[\int_t^T \overline{\xi}_s e^{-\frac{1}{\gamma}\rho s} \overline{\xi}_s^{-\frac{1}{\gamma}} ds \right] && \text{(A12)} \\ &= l^{-\frac{1}{\gamma}} \frac{1}{\overline{\xi}_t} \overline{E}_t \left[\int_t^T e^{-\frac{1}{\gamma}\rho s} \overline{\xi}_s^{\frac{\gamma-1}{\gamma}} ds \right] && \text{(A13)} \\ &= l^{-\frac{1}{\gamma}} \int_t^T e^{-\frac{1}{\gamma}\rho s} \left\{ \frac{1}{\overline{\xi}_t} \overline{E}_t \left[\overline{\xi}_s^{\frac{\gamma-1}{\gamma}} \right] \right\} ds, && \text{(A14)} \end{aligned}$$

where the first equality comes directly from applying the martingale method, the second equality comes from substituting the solution for optimal consumption into the expectation, the third equality aggregates all of the $\overline{\xi}_s$ terms inside the expectation, and the fourth equality interchanges integration and expectation (that is, it applies the Tonelli-Fubini theorem). Rewrite the above in terms of $\overline{Z}_t$:

$$\begin{aligned} W_t^o &= l^{-\frac{1}{\gamma}} \int_t^T e^{-\frac{1}{\gamma}\rho s} e^{rt} \left\{ \frac{1}{\overline{Z}_t} \overline{E}_t \left[e^{-\frac{\gamma-1}{\gamma} rs} \overline{Z}_s^{\frac{\gamma-1}{\gamma}} \right] \right\} ds && \text{(A15)} \\ &= l^{-\frac{1}{\gamma}} \int_t^T e^{-\frac{1}{\gamma}(\rho - r)s} e^{r(t-s)} \left\{ \frac{1}{\overline{Z}_t} \overline{E}_t \left[\overline{Z}_s^{\frac{\gamma-1}{\gamma}} \right] \right\} ds. && \text{(A16)} \end{aligned}$$

Let "$E^*[\cdot]$" represent the expectation with respect to the measure that makes w^* a martingale. Then Bayes' rule implies that (see Cvitanič, Lazrak, Martellini, and Zapatero (2002))

$$E_t^*[Y] = \frac{1}{\overline{Z}_t} \overline{E}_t \left[Y \overline{Z}_s \right] \tag{A17}$$

for any random variable Y that is measurable with respect to the risky-asset price filtration up to time s. Apply this fact to the above, setting $Y = \overline{Z}_s^{-\frac{1}{\gamma}}$. This results in

$$W_t^o = l^{-\frac{1}{\gamma}} \int_t^T e^{-\frac{1}{\gamma}(\rho - r)s} e^{r(t-s)} E_t^* \left[\overline{Z}_s^{-\frac{1}{\gamma}} \right] ds \tag{A18}$$

$$= l^{-\frac{1}{\gamma}} \int_t^T e^{-\frac{1}{\gamma}(\rho - r)s} e^{r(t-s)} \overline{Z}_t^{-\frac{1}{\gamma}} E_t^* \left[\left(\frac{\overline{Z}_s}{\overline{Z}_t} \right)^{-\frac{1}{\gamma}} \right] ds \tag{A19}$$

$$= l^{-\frac{1}{\gamma}} \overline{Z}_t^{-\frac{1}{\gamma}} \int_t^T e^{-\frac{1}{\gamma}(\rho - r)s} e^{r(t-s)} E_t^* \left[\left(\frac{\overline{Z}_s}{\overline{Z}_t} \right)^{-\frac{1}{\gamma}} \right] ds. \tag{A20}$$

We must now solve for the expectation in the integrand above.

To evaluate the expectation, we will first simplify the ratio $\frac{\overline{Z}_s}{\overline{Z}_t}$. First, using the fact that $dw_u^* = \overline{X}_u du + d\overline{w}_u$, we have that

$$\frac{\overline{Z}_s}{\overline{Z}_t} = \exp\left(-\int_t^s \overline{X}_u d\overline{w}_u - \frac{1}{2} \int_t^s \overline{X}_u^2 du \right) \tag{A21}$$

$$= \exp\left(-\int_t^s \overline{X}_u dw_u^* + \frac{1}{2} \int_t^s \overline{X}_u^2 du \right). \tag{A22}$$

An application of Ito's Lemma verifies that

$$\begin{aligned} & \int_t^s \overline{X}_u dw_u^* \\ = \; & \frac{\tau^2}{1+\tau^2 s} \left(\frac{1}{2} w_s^{*2} + \frac{m}{\tau^2} w_s^* \right) \\ & - \frac{\tau^2}{1+\tau^2 t} \left(\frac{1}{2} w_t^{*2} + \frac{m}{\tau^2} w_t^* \right) \\ & + \int_t^s \left[\left(\frac{\tau^2}{1+\tau^2 u} \right)^2 \left(\frac{1}{2} w_u^{*2} + \frac{m}{\tau^2} w_u^* \right) - \frac{1}{2} \frac{\tau^2}{1+\tau^2 u} \right] du. \end{aligned} \tag{A23}$$

Also, we can expand (using the definition of $\overline{X}_u$) to get that

$$\begin{aligned}\frac{1}{2}\overline{X}_u^2 &= \frac{1}{2}\left(\frac{\tau^2}{1+\tau^2 u}w_u^* + \frac{1}{1+\tau^2 u}m\right)^2 && \text{(A24)}\\ &= \frac{1}{2}\left(\frac{\tau^2}{1+\tau^2 u}\right)^2 w_u^{*2} && \text{(A25)}\\ &\quad + \left(\frac{\tau^2}{1+\tau^2 u}\right)^2 \frac{m}{\tau^2}w_u^* \\ &\quad + \frac{1}{2}\left(\frac{\tau^2}{1+\tau^2 u}\right)^2 \left(\frac{m}{\tau^2}\right)^2.\end{aligned}$$

We can now substitute to find that

$$\begin{aligned}&\frac{\overline{Z}_s}{\overline{Z}_t}\\ =\ & \exp\left\{-\int_t^s \overline{X}_u dw_u^* + \frac{1}{2}\int_t^s \overline{X}_u^2 du\right\} && \text{(A26)}\\ =\ & \exp\left\{\begin{array}{c} -\int_t^s \overline{X}_u dw_u^* + \frac{1}{2}\left(\frac{m}{\tau^2}\right)^2 \int_t^s \left(\frac{\tau^2}{1+\tau^2 u}\right)^2 du + \\ \int_t^s \left(\frac{\tau^2}{1+\tau^2 u}\right)^2 \left(\frac{1}{2}w_u^{*2} + \frac{m}{\tau^2}w_u^*\right) du \end{array}\right\} && \text{(A27)}\\ =\ & \exp\left\{\begin{array}{c} -\frac{\tau^2}{1+\tau^2 s}\left(\frac{1}{2}w_s^{*2} + \frac{m}{\tau^2}w_s^*\right) + \frac{\tau^2}{1+\tau^2 t}\left(\frac{1}{2}w_t^{*2} + \frac{m}{\tau^2}w_t^*\right) \\ -\int_t^s \left[\left(\frac{\tau^2}{1+\tau^2 u}\right)^2 \left(\frac{1}{2}w_u^{*2} + \frac{m}{\tau^2}w_u^*\right) - \frac{1}{2}\frac{\tau^2}{1+\tau^2 u}\right] du \\ +\frac{1}{2}\left(\frac{m}{\tau^2}\right)^2 \int_t^s \left(\frac{\tau^2}{1+\tau^2 u}\right)^2 du + \\ \int_t^s \left(\frac{\tau^2}{1+\tau^2 u}\right)^2 \left(\frac{1}{2}w_u^{*2} + \frac{m}{\tau^2}w_u^*\right) du \end{array}\right\} && \text{(A28)}\\ =\ & \exp\left\{\begin{array}{c} -\frac{\tau^2}{1+\tau^2 s}\left(\frac{1}{2}w_s^{*2} + \frac{m}{\tau^2}w_s^*\right) + \frac{\tau^2}{1+\tau^2 t}\left(\frac{1}{2}w_t^{*2} + \frac{m}{\tau^2}w_t^*\right) \\ +\frac{1}{2}\int_t^s \frac{\tau^2}{1+\tau^2 u}du + \frac{1}{2}\left(\frac{m}{\tau^2}\right)^2 \int_t^s \left(\frac{\tau^2}{1+\tau^2 u}\right)^2 du \end{array}\right\} && \text{(A29)}\\ =\ & \exp\left\{\begin{array}{c} -\frac{\tau^2}{1+\tau^2 s}\left(\frac{1}{2}w_s^{*2} + \frac{m}{\tau^2}w_s^*\right) + \frac{\tau^2}{1+\tau^2 t}\left(\frac{1}{2}w_t^{*2} + \frac{m}{\tau^2}w_t^*\right) \\ +\frac{1}{2}\ln\left(1+\tau^2 s\right) - \frac{1}{2}\ln\left(1+\tau^2 t\right) \\ +\frac{1}{2}\left(\frac{m}{\tau^2}\right)^2 \frac{\tau^2}{1+\tau^2 t} - \frac{1}{2}\left(\frac{m}{\tau^2}\right)^2 \frac{\tau^2}{1+\tau^2 s} \end{array}\right\}, && \text{(A30)}\end{aligned}$$

where the first equality was derived above, the second equality follows from substitution of our result on $\overline{X}_u^2$, the third equality follows from substitution of our result on $\int_t^s \overline{X}_u dw_u^*$, the fourth equality follows from simply cancelling the identical term being added and subtracted in the exponent, and the fifth equality follows from integrating the final two terms in the exponent on the lefthand side of the equality.

Noticing that we have two perfect squares in the exponent, and that we are exponentiating two logarithmic terms, we obtain

$$\frac{\overline{Z}_s}{\overline{Z}_t} = \exp\left\{ \begin{array}{c} -\frac{1}{2}\frac{\tau^2}{1+\tau^2 s}\left(w_s^* + \frac{m}{\tau^2}\right)^2 + \frac{1}{2}\frac{\tau^2}{1+\tau^2 t}\left(w_t^* + \frac{m}{\tau^2}\right)^2 \\ +\frac{1}{2}\ln\left(1+\tau^2 s\right) - \frac{1}{2}\ln\left(1+\tau^2 t\right) \end{array} \right\} \tag{A31}$$

$$= \sqrt{\frac{1+\tau^2 s}{1+\tau^2 t}} \exp\left\{ \begin{array}{c} -\frac{1}{2}\frac{\tau^2}{1+\tau^2 s}\left(w_s^* + \frac{m}{\tau^2}\right)^2 + \\ \frac{1}{2}\frac{\tau^2}{1+\tau^2 t}\left(w_t^* + \frac{m}{\tau^2}\right)^2 \end{array} \right\}, \tag{A32}$$

where the first equality follows by grouping the terms in the previous display into squares and the second equality shows the results of canceling the logarithm with the exponentiation to get the leading square-root term.

Using the above simplification of $\frac{\overline{Z}_s}{\overline{Z}_t}$, we see that

$$\begin{aligned} & E_t^*\left[\left(\frac{\overline{Z}_s}{\overline{Z}_t}\right)^{-\frac{1}{\gamma}}\right] \\ = & \left(\frac{1+\tau^2 s}{1+\tau^2 t}\right)^{-\frac{1}{2\gamma}} \exp\left\{-\frac{1}{2\gamma}\frac{\tau^2}{1+\tau^2 t}\left(w_t^* + \frac{m}{\tau^2}\right)^2\right\} \\ & E_t^*\left[\exp\left\{\frac{1}{2\gamma}\frac{\tau^2}{1+\tau^2 s}\left(w_s^* + \frac{m}{\tau^2}\right)^2\right\}\right]. \end{aligned} \tag{A33}$$

Noting that conditional on w_t^*, w_s^* is normally distributed with mean w_t^* and variance $(s-t)$, we can rewrite the above display as

$$\begin{aligned} & E_t^*\left[\left(\frac{\overline{Z}_s}{\overline{Z}_t}\right)^{-\frac{1}{\gamma}}\right] \\ = & \left(\frac{1+\tau^2 s}{1+\tau^2 t}\right)^{-\frac{1}{2\gamma}} \exp\left\{-\frac{1}{2\gamma}\frac{\tau^2}{1+\tau^2 t}\left(w_t^* + \frac{m}{\tau^2}\right)^2\right\} \\ & E_t^*\left[\exp\left\{\frac{1}{2\gamma}\frac{\tau^2}{1+\tau^2 s}\left(\epsilon_{s,t} + w_t^* + \frac{m}{\tau^2}\right)^2\right\}\right], \end{aligned} \tag{A34}$$

where $\epsilon_{s,t}$ is distributed normally with mean zero and variance $(s-t)$, conditional on w_t^*. Computing the expectation in the above expression only requires integrating with respect to a mean-zero (and variance $(s-t)$) normal distribution for $\epsilon_{s,t}$. Define $\delta(u) \equiv \frac{\tau^2}{1+\tau^2 u}$ for all $u \in [0,T]$. Performing the integral yields:

$$
\begin{aligned}
& E_t^* \left[\left(\frac{\overline{Z}_s}{\overline{Z}_t} \right)^{-\frac{1}{\gamma}} \right] \\
= \;& \left(\frac{1+\tau^2 s}{1+\tau^2 t} \right)^{-\frac{1}{2\gamma}} \frac{1}{\sqrt{1 - \frac{\tau^2}{\gamma(1+\tau^2 s)} (s-t)}} && \text{(A35)} \\
& \times \exp \left\{ \left[-\frac{\tau^2}{2\gamma (1+\tau^2 t)} + \frac{\frac{\tau^2}{2\gamma(1+\tau^2 s)}}{1 - \frac{\tau^2}{\gamma(1+\tau^2 s)} (s-t)} \right] \left(w_t^* + \frac{m}{\tau^2} \right)^2 \right\} \\
= \;& \left(\frac{\delta(t)}{\delta(s)} \right)^{-\frac{1}{2\gamma}} \frac{1}{\left(1 - \frac{1}{\gamma} \delta(s)(s-t) \right)^{\frac{1}{2}}} && \text{(A36)} \\
& \times \exp \left\{ -\frac{1}{2\gamma} \delta(t) \left(w_t^* + \frac{m}{\tau^2} \right)^2 \left(1 - \frac{\frac{\delta(s)}{\delta(t)}}{1 - \frac{1}{\gamma}\delta(s)(s-t)} \right) \right\} \\
= \;& (1 + \delta(t)(s-t))^{-\frac{1}{2\gamma}} \frac{1}{\left(1 - \frac{1}{\gamma} \delta(s)(s-t) \right)^{\frac{1}{2}}} && \text{(A37)} \\
& \times \exp \left\{ -\frac{1}{2\gamma} \delta(t) \left(w_t^* + \frac{m}{\tau^2} \right)^2 \left(1 - \frac{\frac{\delta(s)}{\delta(t)}}{1 - \frac{1}{\gamma}\delta(s)(s-t)} \right) \right\} \\
= \;& \frac{(1 + \delta(t)(s-t))^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left[(1+\delta(t)(s-t)) \left(1 - \frac{1}{\gamma}\delta(s)(s-t) \right) \right]^{\frac{1}{2}}} && \text{(A38)} \\
& \times \exp \left\{ -\frac{1}{2\gamma} \delta(t) \left(w_t^* + \frac{m}{\tau^2} \right)^2 \left(1 - \frac{\frac{\delta(s)}{\delta(t)}}{1 - \frac{1}{\gamma}\delta(s)(s-t)} \right) \right\} \\
= \;& \frac{(1 + \delta(t)(s-t))^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left[1 + \delta(t)(s-t) - \frac{1}{\gamma}\delta(s)(s-t) - \frac{1}{\gamma}\delta(t)\delta(s)(s-t)^2 \right]^{\frac{1}{2}}} && \text{(A39)} \\
& \times \exp \left\{ -\frac{1}{2\gamma} \delta(t) \left(w_t^* + \frac{m}{\tau^2} \right)^2 \left(1 - \frac{\frac{\delta(s)}{\delta(t)}}{1 - \frac{1}{\gamma}\delta(s)(s-t)} \right) \right\} \\
= \;& \frac{(1 + \delta(t)(s-t))^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left[1 + \delta(t)(s-t) - \frac{1}{\gamma}(s-t)\left(\delta(s) + \delta(t)\delta(s)(s-t)\right) \right]^{\frac{1}{2}}} && \text{(A40)} \\
& \times \exp \left\{ -\frac{1}{2\gamma} \delta(t) \left(w_t^* + \frac{m}{\tau^2} \right)^2 \left(1 - \frac{\frac{\delta(s)}{\delta(t)}}{1 - \frac{1}{\gamma}\delta(s)(s-t)} \right) \right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\left(1+\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left[1+\delta\left(t\right)\left(s-t\right)-\frac{1}{\gamma}\left(s-t\right)\delta\left(t\right)\right]^{\frac{1}{2}}} \qquad \text{(A41)}\\
&\quad \times \exp\left\{-\frac{1}{2\gamma}\delta\left(t\right)\left(w_t^*+\frac{m}{\tau^2}\right)^2\left(1-\frac{\frac{\delta(s)}{\delta(t)}}{1-\frac{1}{\gamma}\delta\left(s\right)\left(s-t\right)}\right)\right\}\\
&= \frac{\left(1+\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}}} \qquad \text{(A42)}\\
&\quad \times \exp\left\{-\frac{1}{2\gamma}\delta\left(t\right)\left(w_t^*+\frac{m}{\tau^2}\right)^2\left(1-\frac{\frac{\delta(s)}{\delta(t)}}{1-\frac{1}{\gamma}\delta\left(s\right)\left(s-t\right)}\right)\right\}\\
&= \frac{\left(1+\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}}} \qquad \text{(A43)}\\
&\quad \times \exp\left\{-\frac{1}{2\gamma}\delta\left(t\right)\left(w_t^*+\frac{m}{\tau^2}\right)^2\frac{1-\frac{1}{\gamma}\delta\left(s\right)\left(s-t\right)-\frac{\delta(s)}{\delta(t)}}{1-\frac{1}{\gamma}\delta\left(s\right)\left(s-t\right)}\right\}\\
&= \frac{\left(1+\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}}} \qquad \text{(A44)}\\
&\quad \times \exp\left\{-\frac{1}{2\gamma}\delta\left(t\right)\left(w_t^*+\frac{m}{\tau^2}\right)^2\frac{1-\frac{1}{\gamma}\frac{\tau^2}{1+\tau^2 s}\left(s-t\right)-\frac{1+\tau^2 t}{1+\tau^2 s}}{1-\frac{1}{\gamma}\frac{\tau^2}{1+\tau^2 s}\left(s-t\right)}\right\}\\
&= \frac{\left(1+\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}}} \qquad \text{(A45)}\\
&\quad \times \exp\left\{-\frac{1}{2\gamma}\delta\left(t\right)\left(w_t^*+\frac{m}{\tau^2}\right)^2\frac{1+\tau^2 s-\frac{1}{\gamma}\tau^2\left(s-t\right)-\left(1+\tau^2 t\right)}{1+\tau^2 s-\frac{1}{\gamma}\tau^2\left(s-t\right)}\right\}\\
&= \frac{\left(1+\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta\left(t\right)\left(s-t\right)\right)^{\frac{1}{2}}} \qquad \text{(A46)}\\
&\quad \times \exp\left\{-\frac{1}{2\gamma}\delta\left(t\right)\left(w_t^*+\frac{m}{\tau^2}\right)^2\frac{\left(1-\frac{1}{\gamma}\right)\tau^2\left(s-t\right)}{1+\tau^2 s-\frac{1}{\gamma}\tau^2\left(s-t\right)}\right\}
\end{aligned}
$$

$$= \frac{(1+\delta(t)(s-t))^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)\right)^{\frac{1}{2}}} \tag{A47}$$

$$\times \exp\left\{-\frac{1}{2\gamma}\delta^2(t)\left(w_t^*+\frac{m}{\tau^2}\right)^2 \frac{\left(\left(1-\frac{1}{\gamma}\right)\tau^2(s-t)\right)/\tau^2}{\left(1+\tau^2 s-\frac{1}{\gamma}\tau^2(s-t)\right)/(1+\tau^2 t)}\right\}$$

$$= \frac{(1+\delta(t)(s-t))^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)\right)^{\frac{1}{2}}} \tag{A48}$$

$$\times \exp\left\{-\frac{1}{2\gamma}\delta^2(t)\left(w_t^*+\frac{m}{\tau^2}\right)^2 \frac{\left(1-\frac{1}{\gamma}\right)(s-t)}{\frac{\delta(t)}{\delta(s)}-\frac{1}{\gamma}\delta(t)(s-t)}\right\}$$

$$= \frac{(1+\delta(t)(s-t))^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)\right)^{\frac{1}{2}}} \tag{A49}$$

$$\times \exp\left\{-\frac{1}{2\gamma}\delta^2(t)\left(w_t^*+\frac{m}{\tau^2}\right)^2 \frac{\left(1-\frac{1}{\gamma}\right)(s-t)}{1+\delta(t)(s-t)-\frac{1}{\gamma}\delta(t)(s-t)}\right\}$$

$$= \frac{(1+\delta(t)(s-t))^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)\right)^{\frac{1}{2}}} \tag{A50}$$

$$\times \exp\left\{-\frac{1}{2\gamma}\delta^2(t)\left(w_t^*+\frac{m}{\tau^2}\right)^2 \frac{\left(1-\frac{1}{\gamma}\right)(s-t)}{1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)}\right\}$$

$$= \frac{(1+\delta(t)(s-t))^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)\right)^{\frac{1}{2}}} \tag{A51}$$

$$\times \exp\left\{-\frac{1}{2\gamma}\left[\frac{\left(1-\frac{1}{\gamma}\right)(s-t)}{1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)}\right]\overline{X}_t^2\right\}.$$

There are seventeen equalities in the above derivation. The first is by computation of the integral. The second is by the definition of the function $\delta(\cdot)$. The third is by the elementary equality $\delta(t)=\delta(s)+\delta(s)\,\delta(t)\,(s-t)$, which is easily checked from the definition of $\delta(\cdot)$. The fourth is by multiplying and dividing the leading fraction by $1+\delta(t)(s-t)$. The fifth is by multiplying out the factors in the denominator of the leading fraction. The sixth is by regrouping the terms in the denominator of the leading fraction. The seventh is, again, an application of the elementary equality $\delta(t) = \delta(s) + \delta(s)\delta(t)(s-t)$, this time in the denominator of the leading fraction. The eighth is a simple regrouping of the terms in the denominator of the leading fraction. The ninth puts both terms in the difference within the exponent over a common denominator. The tenth follows from the definition of the function $\delta(\cdot)$. The eleventh multiplies both the numerator and the denominator of the fraction within the exponent by $(1+\tau^2 s)$. The twelfth cancels some of the resulting terms and regroups the rest. The thirteenth multiplies and divides the term in the exponent by $\delta(t) = \frac{\tau^2}{1+\tau^2 t}$. The fourteenth applies, once again, the definition of $\delta(\cdot)$. The fifteenth applies, once more, the elementary equality $\delta(t) = \delta(s) + \delta(s)\delta(t)(s-t)$. The sixteenth regroups the resulting terms. Finally, the seventeenth follows from the definition of $\overline{X}_t^2$.

Define

$$A(s-t,\delta(t)) \equiv \left[\frac{\left(1-\frac{1}{\gamma}\right)(s-t)}{1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)}\right] \tag{A52}$$

$$\tilde{B}(s-t,\delta(t)) \equiv \frac{(1+\delta(t)(s-t))^{\frac{1}{2}\left(1-\frac{1}{\gamma}\right)}}{\left(1+\left(1-\frac{1}{\gamma}\right)\delta(t)(s-t)\right)^{\frac{1}{2}}}. \tag{A53}$$

The above results imply that

$$E_t^*\left[\left(\frac{\overline{Z}_s}{\overline{Z}_t}\right)^{-\frac{1}{\gamma}}\right] = \tilde{B}(s-t,\delta(t))\exp\left\{-\frac{1}{2\gamma}A(s-t,\delta(t))\,\overline{X}_t^2\right\}. \tag{A54}$$

The optimal value of wealth is then

$$
\begin{aligned}
& W_t^o \\
= \; & l^{-\frac{1}{\gamma}} \overline{Z}_t^{-\frac{1}{\gamma}} \int_t^T e^{-\frac{1}{\gamma}(\rho - r)s} e^{r(t-s)} E_t^* \left[\left(\frac{\overline{Z}_s}{\overline{Z}_t} \right)^{-\frac{1}{\gamma}} \right] ds \qquad \text{(A55)} \\
= \; & l^{-\frac{1}{\gamma}} \overline{Z}_t^{-\frac{1}{\gamma}} \int_t^T e^{-\frac{1}{\gamma}(\rho - r)s} e^{r(t-s)} \qquad \text{(A56)} \\
& \tilde{B}(s-t, \delta(t)) \exp\left\{ -\frac{1}{2\gamma} A(s-t, \delta(t)) \overline{X}_t^2 \right\} ds \\
= \; & \left(\frac{1}{l \overline{Z}_t} \right)^{\frac{1}{\gamma}} e^{-\frac{\rho}{\gamma}t} e^{\frac{r}{\gamma}t} \int_t^T e^{-\frac{1}{\gamma}\rho(s-t)} e^{-\left(1-\frac{1}{\gamma}\right) r(s-t)} \qquad \text{(A57)} \\
& \tilde{B}(s-t, \delta(t)) \exp\left\{ -\frac{1}{2\gamma} A(s-t, \delta(t)) \overline{X}_t^2 \right\} ds \\
= \; & \overline{Y}_t^{\frac{1}{\gamma}} e^{-\frac{\rho}{\gamma}t} \int_t^T e^{-\frac{1}{\gamma}\rho(s-t)} e^{-\left(1-\frac{1}{\gamma}\right) r(s-t)} \qquad \text{(A58)} \\
& \tilde{B}(s-t, \delta(t)) \exp\left\{ -\frac{1}{2\gamma} A(s-t, \delta(t)) \overline{X}_t^2 \right\} ds \\
= \; & \overline{Y}_t^{\frac{1}{\gamma}} e^{-\frac{\rho}{\gamma}t} \int_t^T B(s-t, \delta(t)) \exp\left\{ -\frac{1}{2\gamma} A(s-t, \delta(t)) \overline{X}_t^2 \right\} ds, \qquad \text{(A59)}
\end{aligned}
$$

where the first equality is by the martingale method (as shown above), the second equality is shown in the previous paragraph, the third equality follows from multiplying and dividing the expression in the previous line by $e^{-\frac{\rho}{\gamma}t} e^{\frac{r}{\gamma}t}$, the fourth equality follows from defining $\overline{Y}_t = \frac{1}{l\overline{\xi}_t} = \frac{1}{l\overline{Z}_t e^{-rt}}$ (the second equality being due to the definition of $\overline{\xi}_t$; see above), and the fifth equality follows from defining

$$
B(s-t, \delta(t)) \equiv e^{-\frac{1}{\gamma}\rho(s-t)} e^{-\left(1-\frac{1}{\gamma}\right) r(s-t)} \tilde{B}(s-t, \delta(t)). \qquad \text{(A60)}
$$

If we let $J\left(W, \overline{X}_t, t\right)$ represent the value function, or indirect utility function, for this problem, Cox and Huang (1989) show how to relate the optimal wealth function, as a function of $\overline{Y}_t$, to J. In this setting, the relationship they demonstrate becomes:

$$
\begin{aligned}
& \frac{\partial J}{\partial W} = \qquad \text{(A61)} \\
& W^{-\gamma} e^{-\rho t} \left(\int_t^T B(s-t, \delta(t)) \exp\left\{ -\frac{1}{2\gamma} A(s-t, \delta(t)) \overline{X}_t^2 \right\} ds \right)^{\gamma}.
\end{aligned}
$$

By observing that $\lim_{W\to\infty} J\left(W, \overline{X}_t, t\right) = 0$ (since it is clear that the indirect utility of wealth is nonpositive because $\gamma > 1$, and even a suboptimal investment strategy such as investing and consuming as though the risk premium were known to be zero will deliver a zero indirect utility of wealth in the limit as $W \to \infty$), we can integrate the above display once with respect to W to obtain (since the zero limit implies that the constant of integration must be zero):

$$J\left(W, \overline{X}_t, t\right) = \frac{W^{1-\gamma}}{1-\gamma} e^{-\rho t} \left(\int_t^T B\left(s-t, \delta\left(t\right)\right) \exp\left\{ -\frac{1}{2\gamma} A\left(s-t, \delta\left(t\right)\right) \overline{X}_t^2 \right\} ds \right)^{\gamma}. \tag{A62}$$

We may then calculate the optimal portfolio weights and the optimal consumption-wealth ratio from the indirect utility function, or we may calculate them directly using the martingale method (see Wachter (2002)). They are

$$\frac{c_t^o}{W_t^o} = \frac{1}{\int_t^T B\left(s-t, \delta\left(t\right)\right) \exp\left\{ -\frac{1}{2\gamma} A\left(s-t, \delta\left(t\right)\right) \overline{X}_t^2 \right\} ds} \tag{A63}$$

$$\varphi_t^o = \frac{\overline{X}_t}{\gamma\sigma} \times \left(1 - \delta\left(t\right) \frac{\int_t^T B\left(s-t, \delta\left(t\right)\right) \exp\left\{ -\frac{1}{2\gamma} A\left(s-t, \delta\left(t\right)\right) \overline{X}_t^2 \right\} A\left(s-t, \delta\left(t\right)\right) ds}{\int_t^T B\left(s-t, \delta\left(t\right)\right) \exp\left\{ -\frac{1}{2\gamma} A\left(s-t, \delta\left(t\right)\right) \overline{X}_t^2 \right\} ds} \right). \tag{A64}$$

Since $\overline{X}_t$ is the posterior mean of the Sharpe ratio, which (in our setting of non-stochastic volatility and riskless rate) is identical to the Sharpe ratio constructed using the posterior mean, this proves the proposition. □

Proof of Theorem 5:

We apply the theorem found on page 90 of Ferguson (1967): given a candidate least-favorable prior, we can check that the Bayes decision against it is maxmin by examining the expected utility of that Bayes decision against other parameter distributions. If the Bayes utility of the decision rule (against the candidate least-favorable prior) is equal to its minimal expected utility over all other parameter values in the set, then the decision rule is maxmin and the prior is, in fact, least-favorable.

Our candidate least-favorable prior is the point-mass prior which places probability one on $X = X^{LF}$, as defined in the statement of the theorem. If we can show that, under every X such that $r + \sigma X \in [\underline{\lambda}, \overline{\lambda}]$, the expected utility of the decision that is Bayes against X^{LF} is no higher than the expected utility of this decision against the point-mass prior X^{LF}, then we are done.

First note that the consumption rule, as a fraction of wealth, is certainly feasible; obviously, the portfolio choice is feasible.

Let us calculate the expected utility, under a general X, of the decision that is Bayes against X^{LF}:

$$\begin{aligned}
& E_X\left[\int_0^T e^{-\rho t}\frac{c_t^{1-\gamma}}{1-\gamma}dt\right] \\
= & \int_0^T E_X\left[e^{-\rho t}\frac{c_t^{1-\gamma}}{1-\gamma}\right] dt W_0^{1-\gamma} E_X\left[\left(\frac{W_t}{W_0}\right)^{1-\gamma}\right] dt \qquad \text{(A65)} \\
= & \int_0^T \frac{e^{-\rho t}}{1-\gamma}\left(\frac{a^{LF}}{1-e^{-a^{LF}(T-t)}}\right)^{1-\gamma} \qquad \text{(A66)} \\
& W_0^{1-\gamma}\exp\left\{(1-\gamma)\,t\left(r-\frac{a^{LF}}{1-e^{-a^{LF}(T-t)}}+\varphi_t^{o\,LF}X-\frac{1}{2}\gamma\left(\varphi_t^{o\,LF}\right)^2\sigma^2\right)\right\} dt,
\end{aligned}$$

where the first equality follows from applying the Tonelli-Fubini theorem, the second equality follows from using the structure of optimal consumption in the Bayes decision against X^{LF}, and the third equality is obtained by taking the expectation (since the portfolio weight on the risky asset is constant under the Bayes decision against X^{LF}, this amounts to taking the expectation of the exponential of a Gaussian random variable).

If $\overline{\lambda} - r < 0$, all possible X are negative, and $X^{LF} = \frac{\overline{\lambda}-r}{\sigma}$; thus, the minimum of the above expression occurs at $X = X^{LF}$. If $\underline{\lambda} - r > 0$, then all possible X are positive, and $X^{LF} = \frac{\lambda - r}{\sigma}$. Thus, the minimum of the above expression occurs at $X = X^{LF}$. Finally, if $\underline{\lambda} < 0 < \overline{\lambda}$, then $X^{LF} = 0$ and the above expression does not depend on X (since $\varphi_t^o = 0$ in this case). We have exhausted the possible cases, and shown that the minimum expected utility, over all possible X, of the Bayes decision against X^{LF} is equal to the Bayes utility of the decision. This completes the proof. □

Proof of Theorem 6: We use the same theorem of Ferguson (1967) that we employed in the proof of Theorem 5 above. The parameter here is the prior expected Sharpe ratio, which indexes the normal distributions that are the building blocks of the set of priors. First, we must check that the Bayes decision against the candidate least-favorable prior is, in fact, feasible under any other parameter value (that is, under any other normal distribution in Π). Let "$E^{LF}[\cdot]$" denote expectation under the candidate least-favorable prior, while "$E^*[\cdot]$" denotes expectation under the risk-neutral measure and "$\tilde{E}[\cdot]$" denotes expectation under some arbitrary normal distribution in Π. Then we can see immediately from the martingale-method budget constraints that a consumption plan is feasible for the candidate least-favorable prior if and only if it is feasible for any other prior:

$$E^{LF}\left[\int_0^T c_t \overline{\xi}_t^{LF} dt\right] \leq W_0 \tag{A67}$$

$$\Leftrightarrow \quad E^*\left[\int_0^T c_t e^{-rt} dt\right] \leq W_0 \tag{A68}$$

$$\Leftrightarrow \quad \tilde{E}\left[\int_0^T c_t \tilde{\overline{\xi}}_t dt\right] \leq W_0. \tag{A69}$$

Now we must check that the same portfolio decision finances a given consumption plan under the candidate least-favorable prior and under any other prior:

$$dW_t = (rW_t - c_t)\,dt + \varphi_t W_t \left(\sigma \overline{X}_t^{LF} dt + \sigma d\overline{w}_t^{LF}\right) \tag{A70}$$

$$\Leftrightarrow \quad dW_t = (rW_t - c_t)\,dt + \varphi_t W_t \left(\sigma \tilde{\overline{X}}_t dt + \sigma d\tilde{\overline{w}}_t\right), \tag{A71}$$

where the equivalence follows from the definition of the two different innovation processes; see the proof of Proposition 3.

Finally, we must verify that the candidate maxmin decision rule delivers no less utility for any other normal distribution in Π than it does for the candidate least-favorable distribution (against which it is Bayes). We adopt all of the notation of the proof of Proposition 3. We will show that the instantaneous expected utility of consumption at every moment is minimized by the candidate least-favorable distribution, if the candidate maxmin decision rule is used. This will prove the desired result.

$$
\begin{aligned}
& \tilde{E}\left[e^{-\rho t}\frac{c_t^{o\,(1-\gamma)}}{1-\gamma}\right] \\
= \; & E^*\left[e^{-\rho t}\frac{c_t^{o\,(1-\gamma)}}{1-\gamma}\frac{1}{\tilde{Z}_t}\right] && \text{(A72)} \\
= \; & \frac{e^{-\frac{1}{\gamma}\rho t}}{1-\gamma} l^{\frac{\gamma-1}{\gamma}} e^{\frac{\gamma-1}{\gamma}rt} E^*\left[\overline{Z}_t^{\frac{\gamma-1}{\gamma}}\frac{1}{\tilde{Z}_t}\right], && \text{(A73)}
\end{aligned}
$$

where $\overline{Z}_t$ is Z_t for the candidate least-favorable prior. Now, none of the quantities to the left of the expectations operator in the last line above are affected by which prior we use in evaluating the utility of the consumption and portfolio choice. Thus, we focus on the expectation in the last line above. Since the quantity multiplying it is negative (because $\gamma > 1$), we wish to show that the candidate least-favorable prior *maximizes* this expectation.

$$
\begin{aligned}
& E^*\left[\overline{Z}_t^{\frac{\gamma-1}{\gamma}}\frac{1}{\tilde{Z}_t}\right] \\
= \; & E^*\left[\left(1+\tau^2 t\right)^{-\frac{1}{2\gamma}}\exp\left\{\begin{array}{c}\frac{1}{2\tau^2}\left(\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\tilde{m}^2\right)\\ +\frac{1}{2}\delta\left(t\right)\left(\begin{array}{c}-\frac{\gamma-1}{\gamma}\left(w_t^*+\frac{m^{LF}}{\tau^2}\right)^2\\ +\left(w_t^*+\frac{\tilde{m}}{\tau^2}\right)^2\end{array}\right)\end{array}\right\}\right] && \text{(A74)}\\
= \; & \left(1+\tau^2 t\right)^{-\frac{1}{2\gamma}}\exp\left\{\frac{1}{2\tau^2}\left(\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\tilde{m}^2\right)\right\} && \text{(A75)}\\
& \times E^*\left[\exp\left\{\begin{array}{c}\frac{1}{2\gamma}\delta\left(t\right)w_t^{*\,2}+\frac{\delta(t)}{\tau^2}\left(\tilde{m}-\frac{\gamma-1}{\gamma}m^{LF}\right)w_t^*\\ +\frac{\delta(t)}{2\tau^4}\left(\tilde{m}^2-\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2\right)\end{array}\right\}\right]\\
= \; & \left(1+\tau^2 t\right)^{-\frac{1}{2\gamma}}\exp\left\{\frac{1}{2\tau^2}\left(\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\tilde{m}^2\right)\right\} && \text{(A76)}\\
& \times\exp\left\{\frac{1-\gamma}{2\tau^4}\delta\left(t\right)\left(m^{LF}-\tilde{m}\right)^2\right\}\\
& \times E^*\left[\exp\left\{\frac{1}{2\gamma}\delta\left(t\right)\left(w_t^*+\frac{1}{\tau^2}\left(\left(1-\gamma\right)m^{LF}+\gamma\tilde{m}\right)\right)^2\right\}\right]\\
= \; & \left(1-\frac{1}{\gamma}\delta\left(t\right)t\right)^{-\frac{1}{2}}\left(1+\tau^2 t\right)^{-\frac{1}{2\gamma}} && \text{(A77)}\\
& \times\exp\left\{\frac{1}{2\tau^2}\left(\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\tilde{m}^2\right)\right\}\\
& \times\exp\left\{\frac{1-\gamma}{2\tau^4}\delta\left(t\right)\left(m^{LF}-\tilde{m}\right)^2\right\}\\
& \times\exp\left\{\frac{1}{2\tau^4}\frac{\gamma\delta\left(t\right)}{1-\frac{1}{\gamma}\delta\left(t\right)t}\left(\frac{\gamma-1}{\gamma}m^{LF}-\tilde{m}\right)^2\right\},
\end{aligned}
$$

where the first equality is from our manipulations of the Z_t functions in the proof of Proposition 3, the second equality follows from expanding squares and collecting terms, the third equality completes the square on w_t^*, and the fourth equality follows from applying Lemma 5 to the expectation in the line preceding it.

Taking the natural logarithm of the expression above leaves us with

$$\begin{aligned} & c + \frac{1}{2\tau^2}\left(\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2 - \tilde{m}^2\right) \\ & + \frac{1-\gamma}{2\tau^4}\delta(t)\left(m^{LF} - \tilde{m}\right)^2 \\ & + \frac{1}{2\tau^4}\frac{\gamma\delta(t)}{1-\frac{1}{\gamma}\delta(t)t}\left(\frac{\gamma-1}{\gamma}m^{LF} - \tilde{m}\right)^2, \end{aligned} \tag{A78}$$

where c is constant with respect to $\tilde{m}$. Take the derivative of this quadratic with respect to $\tilde{m}$ to obtain:

$$\begin{aligned} & -\frac{1}{\tau^2}\tilde{m} - \frac{1-\gamma}{\tau^4}\delta(t)\left(m^{LF} - \tilde{m}\right) - \frac{1}{\tau^4}\frac{\gamma\delta(t)}{1-\frac{1}{\gamma}\delta(t)t}\left(\frac{\gamma-1}{\gamma}m^{LF} - \tilde{m}\right) \\ = & \left(\frac{1-\gamma}{\tau^4}\delta(t) + \frac{\gamma\delta(t)}{\tau^4\left(1-\frac{1}{\gamma}\delta(t)\right)} - \frac{1}{\tau^2}\right)\tilde{m} \\ & - \left(\frac{1-\gamma}{\tau^4}\delta(t) + \frac{(\gamma-1)\delta(t)}{\tau^4\left(1-\frac{1}{\gamma}\delta(t)\right)}\right)m^{LF}. \end{aligned} \tag{A79}$$

To be sure that the quadratic is concave (so that a point satisfying the first-order condition is a maximum rather than a minimum), we show that the coefficient on $\tilde{m}$ in the derivative above (which is twice the coefficient on $\tilde{m}^2$ in the original quadratic) is negative:

$$
\begin{aligned}
& \frac{1-\gamma}{\tau^4}\delta(t)+\frac{\gamma\delta(t)}{\tau^4\left(1-\frac{1}{\gamma}\delta(t)\right)}-\frac{1}{\tau^2} \\
= \; & \frac{\delta(t)}{\tau^4}\left(1-\gamma+\frac{\gamma}{1-\frac{1}{\gamma}\delta(t)}-\tau^2\frac{1}{\delta(t)}\right) && \text{(A80)} \\
= \; & \frac{\delta(t)}{\tau^4}\left(1-\left(1+\tau^2 t\right)+\gamma\left(\frac{1}{1-\frac{1}{\gamma}\delta(t)}-1\right)\right) && \text{(A81)} \\
= \; & -\frac{\delta(t)\,t}{\tau^2}+\frac{\delta(t)}{\tau^4}\frac{\delta(t)\,t}{1-\frac{1}{\gamma}\delta(t)\,t} && \text{(A82)} \\
= \; & \frac{\delta(t)\,t}{\tau^2}\left(\frac{\frac{1}{\tau^2}\delta(t)}{1-\frac{1}{\gamma}\delta(t)\,t}-1\right) && \text{(A83)} \\
= \; & \frac{\delta(t)\,t}{\tau^2}\left(\frac{1}{\tau^2}\delta(t)-1+\frac{1}{\gamma}\delta(t)\,t\right)\frac{1}{1-\frac{1}{\gamma}\delta(t)\,t} && \text{(A84)} \\
< \; & \frac{\delta(t)\,t}{\tau^2}\left(\frac{1}{\tau^2}\delta(t)-1+\delta(t)\,t\right)\frac{1}{1-\frac{1}{\gamma}\delta(t)\,t} && \text{(A85)} \\
= \; & \frac{\delta(t)\,t}{\tau^2}\left(\left(1+\frac{1}{\tau^2}\right)\delta(t)-1\right)\frac{1}{1-\frac{1}{\gamma}\delta(t)\,t} && \text{(A86)} \\
= \; & \frac{\delta(t)\,t}{\tau^2}(1-1)\frac{1}{1-\frac{1}{\gamma}\delta(t)\,t} && \text{(A87)} \\
= \; & 0, && \text{(A88)}
\end{aligned}
$$

in which every equality is by elementary algebraic manipulation, and the inequality is by the fact that $\gamma > 1$, so that $\frac{1}{\gamma}\delta(t) < \delta(t)$, and the facts that $\frac{\delta(t)t}{\tau^2} > 0$ and $\frac{1}{1-\frac{1}{\gamma}\delta(t)t} = \frac{1}{1-\frac{1}{\gamma}\frac{\tau^2 t}{1+\tau^2 t}} > 0$. This proves that the original quadratic was concave, so setting the derivative to zero identifies the global maximum of the quadratic.

Setting the derivative given above to zero, we find that the *unconstrained* maximal $\tilde{m}$, which we denote $\tilde{m}^{*,\,UNC}$, is given by

$$\tilde{m}^{*,\,UNC} = \frac{b}{a}m^{LF}, \tag{A89}$$

where

$$b = \frac{1-\gamma}{\tau^4}\delta(t) + \frac{(\gamma-1)\,\delta(t)}{\tau^4\left(1-\frac{1}{\gamma}\delta(t)\right)} \tag{A90}$$

$$a = \frac{1-\gamma}{\tau^4}\delta(t) + \frac{\gamma\delta(t)}{\tau^4\left(1-\frac{1}{\gamma}\delta(t)\right)} - \frac{1}{\tau^2}. \tag{A91}$$

We have already shown that $a < 0$ above; we now show that $b > 0$:

$$\frac{1-\gamma}{\tau^4}\delta(t) + \frac{(\gamma-1)\,\delta(t)}{\tau^4\left(1-\frac{1}{\gamma}\delta(t)\right)}$$

$$= \frac{\gamma-1}{\tau^4}\delta(t)\left(\frac{1}{1-\frac{1}{\gamma}\delta(t)\,t} - 1\right) \tag{A92}$$

$$= \frac{\gamma-1}{\tau^4}\delta(t)\left(\frac{\frac{1}{\gamma}\delta(t)\,t}{1-\frac{1}{\gamma}\delta(t)\,t}\right) \tag{A93}$$

$$> 0, \tag{A94}$$

in which both equalities follow from elementary algebra and the inequality is due to the facts that $\gamma > 1$, $\delta(t) > 0$, and $\delta(t)\,t = \frac{\tau^2 t}{1+\tau^2 t} \leq 1$, so that $1 - \frac{1}{\gamma}\delta(t)\,t > 0$.

The above implies that the sign of $\tilde{m}^{*,\,UNC}$ is the negative of the sign of m^{LF}. Now consider the three possible cases: if $0 \in \left[\underline{\lambda} - r, \overline{\lambda} - r\right]$ then $m^{LF} = 0$, and the sign of $\tilde{m}^{*,\,UNC}$ is the negative of zero, which is zero, so $\tilde{m}^{*,\,UNC} = 0$ is the constrained maximizer as well as the unconstrained maximizer. If $0 < \underline{\lambda} - r$, then $m^{LF} = \underline{m} = \frac{\underline{\lambda}-r}{\sigma}$ is the smallest possible m, and is positive. Thus, $\tilde{m}^{*,\,UNC}$ is negative, and therefore outside (to the left of, on the real line) the constraint interval. Since the objective function is a concave parabola (so that, on the real line, it is monotonically decreasing to the right of its vertex), this implies that the constrained maximizer $\tilde{m}^* = \underline{m}$, the smallest possible m, so $\tilde{m}^* = m^{LF}$. Finally, suppose that $0 > \overline{\lambda} - r$. Then $m^{LF} = \overline{m} = \frac{\overline{\lambda}-r}{\sigma}$ is the smallest possible m in absolute value, and is negative. Thus, $\tilde{m}^{*,\,UNC}$ is positive, and therefore outside (to the right of, on the real line) the constraint interval. Since the objective function is a concave parabola (so that, on the real line, it is monotonically increasing to the left of its vertex), this implies that the constrained maximizer $\tilde{m}^* = \overline{m}$, the smallest possible m in absolute value, so that $\tilde{m}^* = m^{LF}$.

In each of the three possible cases, then, $\tilde{m}^* = m^{LF}$ maximizes the objective function and hence (due to the negative sign of expected utility, as noted above) minimizes expected utility at the moment t. Since this holds $\forall t \in [0, T]$, the expected utility of the candidate maxmin decision rule is minimized at the candidate least-favorable prior.

Thus, as argued above, the candidate least-favorable prior is, in fact, least-favorable, and the candidate maxmin decision rule is, in fact, maxmin. $\square$

Proof of Theorem 7: We use the same theorem of Ferguson (1967) that we employed in the proof of Theorem 5 above. The parameters here are the prior expected Sharpe ratio and the prior variance of the Sharpe ratio, which index the normal distributions that are the building blocks of the set of priors. First, we must check that the Bayes decision against the candidate least-favorable prior is, in fact, feasible under any other parameter value (that is, under any other normal distribution in Γ). Let "$E^{LF}[\cdot]$" denote expectation under the candidate least-favorable prior, while "$E^*[\cdot]$" denotes expectation under the risk-neutral measure and "$\tilde{E}[\cdot]$" denotes expectation under some arbitrary normal distribution in Γ. Then we can see immediately from the martingale-method budget constraints that a portfolio plan is feasible for the candidate least-favorable prior if and only if it is feasible for any other prior:

$$E^{LF}\left[W_T \overline{\xi}_T^{LF}\right] \leq W_0 \tag{A95}$$

$$\Leftrightarrow \quad E^*\left[W_T e^{-rT}\right] \leq W_0 \tag{A96}$$

$$\Leftrightarrow \quad \tilde{E}\left[W_T \tilde{\overline{\xi}}_T dt\right] \leq W_0. \tag{A97}$$

Now we must check that the same portfolio decision finances a given terminal wealth under the candidate least-favorable prior and under any other prior:

$$dW_t = rW_t dt + \varphi_t W_t \left(\sigma \overline{X}_t^{LF} dt + \sigma d\overline{w}_t^{LF}\right) \tag{A98}$$

$$\Leftrightarrow \quad dW_t = rW_t dt + \varphi_t W_t \left(\sigma \tilde{\overline{X}}_t dt + \sigma d\tilde{\overline{w}}_t\right), \tag{A99}$$

where the equivalence follows from the definition of the two different innovation processes; see the proof of Proposition 3.

Finally, we must verify that the candidate maxmin decision rule delivers no less expected utility for any other normal distribution in Γ than it does for the candidate least-favorable distribution (against which it is Bayes). We adopt all of the notation of the proof of Proposition 3. We begin by calculating the expected utility obtained when the investor makes portfolio and consumption choices that would be optimal under the candidate least-favorable distribution, but the actual prior is not necessarily the candidate least-favorable prior.

$$\begin{aligned} & \tilde{E}\left[\frac{W_T^{o\,(1-\gamma)}}{1-\gamma}\right] \\ = \quad & E^*\left[\frac{W_T^{o\,(1-\gamma)}}{1-\gamma}\frac{1}{\tilde{Z}_T}\right] && \text{(A100)} \\ = \quad & \frac{1}{1-\gamma}k^{\frac{\gamma-1}{\gamma}}e^{\frac{\gamma-1}{\gamma}rT}E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}}\frac{1}{\tilde{Z}_T}\right], && \text{(A101)} \end{aligned}$$

where $\overline{Z}_T$ is Z_T for the candidate least-favorable prior, and k is the Lagrange multiplier delivered by the martingale method in the terminal-wealth investor's Bayesian problem when the prior is the candidate least-favorable prior. Now, none of the quantities to the left of the expectations operator in the last line above are affected by which prior we use in evaluating the utility of the portfolio choice. Thus, we focus on the expectation in the last line above. Since the quantity multiplying it is negative (because $\gamma > 1$), we wish to show that the candidate least-favorable prior *maximizes* this expectation.

$$
\begin{aligned}
& E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}}\frac{1}{\tilde{Z}_T}\right] \\
= \;& E^*\left[\frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\left(1+\tilde{\tau}^2T\right)^{\frac{1}{2}}}\exp\left\{\begin{array}{c}\frac{1}{2\tau^{2,\,LF}}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\frac{1}{2\tilde{\tau}^2}\tilde{m}^2\\ +-\frac{1}{2}\delta^{LF}\left(T\right)\frac{\gamma-1}{\gamma}\left(w_T^*+\frac{m^{LF}}{\tau^{2,\,LF}}\right)^2\\ +\frac{1}{2}\tilde{\delta}\left(T\right)\left(w_T^*+\frac{\tilde{m}}{\tilde{\tau}^2}\right)^2\end{array}\right\}\right] && \text{(A102)}\\
= \;& \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\left(1+\tilde{\tau}^2T\right)^{\frac{1}{2}}}\exp\left\{\frac{1}{2\tau^{2,\,LF}}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\frac{1}{2\tilde{\tau}^2}\tilde{m}^2\right\} && \text{(A103)}\\
& \times E^*\left[\exp\left\{\begin{array}{c}\frac{1}{2}\left(-\frac{\gamma-1}{\gamma}\delta^{LF}\left(T\right)+\tilde{\delta}\left(T\right)\right){w_T^*}^2\\ +\left(\frac{\tilde{\delta}(T)}{\tilde{\tau}^2}\tilde{m}-\frac{\gamma-1}{\gamma}\frac{\delta^{LF}(T)}{\tau^{2,\,LF}}m^{LF}\right)w_T^*\\ +\frac{\tilde{\delta}(T)}{2\tilde{\tau}^4}\tilde{m}^2-\frac{\delta^{LF}(T)}{2\left(\tau^{2,\,LF}\right)^2}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2\end{array}\right\}\right]\\
= \;& \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\left(1+\tilde{\tau}^2T\right)^{\frac{1}{2}}}\exp\left\{\frac{1}{2\tau^{2,\,LF}}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\frac{1}{2\tilde{\tau}^2}\tilde{m}^2\right\} && \text{(A104)}\\
& \times E^*\left[\exp\left\{\begin{array}{c}\frac{1}{2}\left(-\frac{\gamma-1}{\gamma}\delta^{LF}\left(T\right)+\tilde{\delta}\left(T\right)\right)\left(w_T^*+\frac{\frac{\tilde{\delta}(T)}{\tilde{\tau}^2}\tilde{m}-\frac{\gamma-1}{\gamma}\frac{\delta^{LF}(T)}{\tau^{2,\,LF}}m^{LF}}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)}\right)^2\\ -\frac{1}{2}\frac{\left(\frac{\tilde{\delta}(T)}{\tilde{\tau}^2}\tilde{m}-\frac{\gamma-1}{\gamma}\frac{\delta^{LF}(T)}{\tau^{2,\,LF}}m^{LF}\right)^2}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)}\\ +\frac{\tilde{\delta}(T)}{2\tilde{\tau}^4}\tilde{m}^2-\frac{\delta^{LF}(T)}{2\left(\tau^{2,\,LF}\right)^2}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2\end{array}\right\}\right]\\
= \;& \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\left(1+\tilde{\tau}^2T\right)^{\frac{1}{2}}}\exp\left\{\frac{1}{2\tau^{2,\,LF}}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\frac{1}{2\tilde{\tau}^2}\tilde{m}^2\right\} && \text{(A105)}\\
& \times E^*\left[\exp\left\{\begin{array}{c}\frac{1}{2}\left(-\frac{\gamma-1}{\gamma}\delta^{LF}\left(T\right)+\tilde{\delta}\left(T\right)\right)\left(w_T^*+\frac{\frac{\tilde{\delta}(T)}{\tilde{\tau}^2}\tilde{m}-\frac{\gamma-1}{\gamma}\frac{\delta^{LF}(T)}{\tau^{2,\,LF}}m^{LF}}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)}\right)^2\\ +\frac{1}{2}\frac{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\left(\frac{m^{LF}}{\tau^{2,\,LF}}-\frac{\tilde{m}}{\tilde{\tau}^2}\right)^2\end{array}\right\}\right]
\end{aligned}
$$

$$= \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\left(1+\tilde{\tau}^2 T\right)^{\frac{1}{2}}} \exp\left\{ \begin{array}{c} \frac{1}{2\tau^{2,\,LF}}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2 - \frac{1}{2\tilde{\tau}^2}\tilde{m}^2 \\ +\frac{1}{2}\frac{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\left(\frac{m^{LF}}{\tau^{2,\,LF}}-\frac{\tilde{m}}{\tilde{\tau}^2}\right)^2 \end{array} \right\} \tag{A106}$$

$$\times E^*\left[\exp\left\{\frac{1}{2}\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)\left(w_T^* + \frac{\frac{\tilde{\delta}(T)}{\tilde{\tau}^2}\tilde{m} - \frac{\gamma-1}{\gamma}\frac{\delta^{LF}(T)}{\tau^{2,\,LF}}m^{LF}}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)}\right)^2\right\}\right]$$

$$= \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\left(1+\tilde{\tau}^2 T\right)^{\frac{1}{2}}\left(1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T\right)^{\frac{1}{2}}} \tag{A107}$$

$$\exp\left\{ \begin{array}{c} \frac{1}{2\tau^{2,\,LF}}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2 - \frac{1}{2\tilde{\tau}^2}\tilde{m}^2 \\ +\frac{1}{2}\frac{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\left(\frac{m^{LF}}{\tau^{2,\,LF}}-\frac{\tilde{m}}{\tilde{\tau}^2}\right)^2 \end{array} \right\}$$

$$\exp\left\{\frac{1}{2}\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{\left(\frac{\tilde{\delta}(T)}{\tilde{\tau}^2}\tilde{m} - \frac{\gamma-1}{\gamma}\frac{\delta^{LF}(T)}{\tau^{2,\,LF}}m^{LF}\right)^2}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)}\right\}$$

$$= \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\left(1+\tilde{\tau}^2 T\right)^{\frac{1}{2}}\left(1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T\right)^{\frac{1}{2}}} \tag{A108}$$

$$\exp\left\{ \begin{array}{c} \frac{1}{2\tau^{2,\,LF}}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2 - \frac{1}{2\tilde{\tau}^2}\tilde{m}^2 \\ +\frac{1}{2}\frac{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\left(\frac{m^{LF}}{\tau^{2,\,LF}}-\frac{\tilde{m}}{\tilde{\tau}^2}\right)^2 \\ +\frac{1}{2}\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{\left(\frac{\tilde{\delta}(T)}{\tilde{\tau}^2}\tilde{m} - \frac{\gamma-1}{\gamma}\frac{\delta^{LF}(T)}{\tau^{2,\,LF}}m^{LF}\right)^2}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)} \end{array} \right\}.$$

There are seven equalities in the display above. The first is from our manipulations of the Z_T functions in the proof of Proposition 3, the second equality expands squares and collects terms, the third equality completes the square on w_T^*, the fourth equality rearranges the terms separate from the squared w_T^* term into a squared term of their own, the fifth equality removes from the expectation operator the terms not involving w_T^*, the sixth equality applies Lemma 5 to the expectation in the line preceding it, and the seventh equality rearranges the result.

We first show that, regardless of the value of $\tilde{\tau}^2$, the maximizing $\tilde{m}$ is $\tilde{m} = m^{LF}$. To do so, we must show that it maximizes the quantity in the display above.

Noting that $\tilde{m}$ appears only in the exponent above, take the natural logarithm of the above expression to obtain

$$\begin{aligned} & \frac{1}{2\tau^{2,\,LF}}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\frac{1}{2\tilde{\tau}^2}\tilde{m}^2 \\ & +\frac{1}{2}\frac{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\left(\frac{m^{LF}}{\tau^{2,\,LF}}-\frac{\tilde{m}}{\tilde{\tau}^2}\right)^2 \\ & +\frac{1}{2}\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{\left(\frac{\tilde{\delta}(T)}{\tilde{\tau}^2}\tilde{m}-\frac{\gamma-1}{\gamma}\frac{\delta^{LF}(T)}{\tau^{2,\,LF}}m^{LF}\right)^2}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)}+c \\ = {} & a\tilde{m}^2+b\tilde{m}m^{LF}+c, \end{aligned} \tag{A109}$$

where

$$\begin{aligned} a \equiv {} & -\frac{1}{2\tilde{\tau}^2}-\frac{1}{2}\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\frac{1}{\tilde{\tau}^4} \qquad \text{(A110)} \\ & +\frac{1}{2}\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{\tilde{\delta}^2(T)}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)}\frac{1}{\tilde{\tau}^4} \\ b \equiv {} & \frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\frac{1}{\tau^{2,\,LF}\tilde{\tau}^2} \qquad \text{(A111)} \\ & -\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)}\frac{1}{\tau^{2,\,LF}\tilde{\tau}^2}, \end{aligned}$$

and c is constant with respect to $\tilde{m}$. We first show that $a < 0$ (so that the first-order condition is necessary and sufficient for a global maximum) as long as

$\tau^{2,\,LF} > 0$.

$$a = -\frac{1}{2\tilde{\tau}^2} - \frac{1}{2}\frac{\tilde{\delta}(T)\frac{1}{\tilde{\tau}^4}}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T) + \tilde{\delta}(T)} \times \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \frac{\tilde{\delta}(T)}{1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T}\right) \tag{A112}$$

$$= -\frac{1}{2\tilde{\tau}^2} - \frac{1}{2}\frac{\tilde{\delta}(T)\frac{1}{\tilde{\tau}^4}}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T) + \tilde{\delta}(T)} \times \left(\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\left(1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T\right) - \tilde{\delta}(T)}{1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T}\right) \tag{A113}$$

$$= -\frac{1}{2\tilde{\tau}^2} - \frac{1}{2}\frac{\tilde{\delta}(T)\frac{1}{\tilde{\tau}^4}}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T) + \tilde{\delta}(T)} \times \left(\frac{\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)\left(1 + \frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\right)}{1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T}\right) \tag{A114}$$

$$= \frac{1}{2\tilde{\tau}^2}\left(-1 + \frac{\tilde{\delta}(T)\frac{1}{\tilde{\tau}^2}\left(1 + \frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\right)}{1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T}\right) \tag{A115}$$

$$= \frac{1}{2\tilde{\tau}^2\left(1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T\right)} \times \left(\begin{array}{c}\tilde{\delta}(T)\frac{1}{\tilde{\tau}^2}\left(1 + \frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\right) - 1 \\ -\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T\end{array}\right) \tag{A116}$$

$$= \frac{1}{2\tilde{\tau}^2\left(1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T\right)} \times \left(\begin{array}{c}\tilde{\delta}(T)\left(T + \frac{1}{\tilde{\tau}^2}\right) - 1 \\ +\frac{\gamma-1}{\gamma}\tilde{\delta}(T)\,\delta^{LF}(T)\frac{T}{\tilde{\tau}^2} - \frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\end{array}\right) \tag{A117}$$

$$= \frac{1}{2\tilde{\tau}^2\left(1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T\right)} \times \left(\tilde{\delta}(T)\left(T + \frac{1}{\tilde{\tau}^2}\right) - 1 + \frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\left(\tilde{\delta}(T)\frac{1}{\tilde{\tau}^2} - 1\right)\right) \tag{A118}$$

$$= \frac{1}{2\tilde{\tau}^2\left(1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T) - \tilde{\delta}(T)\right)T\right)} \times \left(\tilde{\delta}(T)\left(\frac{\tilde{\tau}^2 T + 1}{\tilde{\tau}^2}\right) - 1 + \frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\left(\frac{1}{1 + \tilde{\tau}^2 T} - 1\right)\right) \tag{A119}$$

$$
\begin{aligned}
&= \frac{1}{2\tilde{\tau}^2\left(1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T\right)}\times \\
&\quad \left(1-1+\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\left(\frac{-\tilde{\tau}^2T}{1+\tilde{\tau}^2T}\right)\right) && \text{(A120)}\\
&= \frac{1}{2\tilde{\tau}^2\left(1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T\right)}\times \\
&\quad \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\left(\frac{-\tilde{\tau}^2T}{1+\tilde{\tau}^2T}\right)\right) && \text{(A121)}\\
&= \frac{1}{2\left(1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T\right)}\times \\
&\quad \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\left(\frac{-T}{1+\tilde{\tau}^2T}\right)\right) && \text{(A122)}\\
&< 0, && \text{(A123)}
\end{aligned}
$$

where each equality is by elementary algebraic manipulation, and the inequality follows from the facts that $T>0$, $1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T>0$, $\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T>0$ if $\tau^{2,\,LF}>0$, and $\frac{-T}{1+\tilde{\tau}^2T}<0$.

The display above also shows that $a=0$ if $\tau^{2,\,LF}=0$.

We now demonstrate that $b < 0$.

$$\begin{aligned}
b &= \frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\frac{1}{\tau^{2,\,LF}\tilde{\tau}^{2}} \qquad \text{(A124)} \\
&\quad -\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\times \\
&\quad \frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)}{\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)}\frac{1}{\tau^{2,\,LF}\tilde{\tau}^{2}} \\
&= \frac{1}{\tau^{2,\,LF}\tilde{\tau}^{2}}\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\times \qquad \text{(A125)} \\
&\quad \left(1-\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\right) \\
&= \frac{1}{\tau^{2,\,LF}\tilde{\tau}^{2}}\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\times \qquad \text{(A126)} \\
&\quad \left(\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\right) \\
&= -\frac{1}{\tau^{2,\,LF}\tilde{\tau}^{2}}\left(\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)\,T}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\right) \qquad \text{(A127)} \\
&= -\frac{\frac{\gamma-1}{\gamma}\frac{1}{1+\tau^{2,\,LF}T}\frac{1}{1+\tilde{\tau}^{2}T}T}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T} \qquad \text{(A128)} \\
&< 0, \qquad \text{(A129)}
\end{aligned}$$

where each equality is by elementary algebraic manipulation, and the inequality follows from the facts that $\frac{1}{1+\tau^{2,\,LF}T} > 0$, $\frac{1}{1+\tilde{\tau}^{2}T} > 0$, $T > 0$, and $1 + \left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T > 0$.

Suppose that $\tau^{2,\,LF} = 0$. The facts we have shown above regarding the signs of a and b imply that $a = 0$ and $b < 0$ in this case, so the maximizing $\tilde{m}$ is the one of largest absolute value which is of sign opposite that of m^{LF} or, if all m in the constraint set have the same sign, it is the m in the constraint set of smallest absolute value. Thus, if the interval $[\underline{m}, \overline{m}]$ (that is, the interval $\left[\frac{\underline{\lambda}-r}{\sigma}, \frac{\overline{\lambda}-r}{\sigma}\right]$) does not include 0, then $\tilde{m}$ is the smallest (in absolute value) element of the interval, but so is m^{LF} (by definition), so $\tilde{m} = m^{LF}$. Now, if $0 \in [\underline{m}, \overline{m}]$, then $m^{LF} = 0$ by definition. Inspecting the quadratic that we seek to maximize, and noting that $a = 0$ because we are considering the case $\tau^{2,\,LF} = 0$, we see that any choice of $\tilde{m}$ delivers the same value of the objective function, which is the c of the quadratic objective function. Thus, $\tilde{m} = m^{LF}$ is a maximizer of the quadratic (this is sufficient for our purposes; we need not show that it is the unique maximizer). Intuitively, if $\tau^{2,\,LF} = 0$ and $m^{LF} = 0$ then the investor holds only the riskless asset for the entire investment horizon (since the investor's problem then reduces to the original Merton problem with Sharpe ratio zero), so the actual expected return on the risky asset has no impact on the investor's expected utility.

Now suppose that $\tau^{2,\,LF} > 0$. Then, as we demonstrated above, $a < 0$ and $b < 0$, so that the unconstrained maximizer of the quadratic objective function (which is $-\frac{b}{2a} m^{LF}$) satisfies

$$\text{sign}\left(\tilde{m}^{*,\,UNC}\right) \quad = \quad -\text{sign}\left(m^{LF}\right). \tag{A130}$$

Recall that a concave quadratic is monotone increasing for arguments less than its maximizer, and monotone decreasing for arguments greater than its maximizer. Now, if $[\underline{m}, \overline{m}]$ does not include zero, then (by definition) m^{LF} has the smallest absolute value of any $m \in [\underline{m}, \overline{m}]$. But, since the unconstrained maximizer of the quadratic has the opposite sign of that of m^{LF}, the constrained maximizer of the quadratic must be $\tilde{m} = m^{LF}$. If $0 \in [\underline{m}, \overline{m}]$, on the other hand, then $m^{LF} = 0$ and the above shows that $\tilde{m} = m^{LF} = 0$.

The preceding logic shows that $\tilde{m} = m^{LF}$ maximizes the quadratic objective function, and thus minimizes expected utility, without regard to the value of $\tilde{\tau}^2$.

Define

$$f\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \equiv \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\left(1+\frac{\gamma-1}{\gamma}\frac{\tau^{2,\,LF}T}{1+\tau^{2,\,LF}T}\left(1+\tilde{\tau}^2 T\right)\right)^{\frac{1}{2}}} \tag{A131}$$

and

$$\begin{aligned} g\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \equiv\ & \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2} - \frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)} \times \\ & \left(\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2}\right)^2 + \frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T} \times \\ & \frac{1}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\left(\frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}}\delta^{LF}(T) - \frac{1}{\tilde{\tau}^2}\tilde{\delta}(T)\right)^2. \end{aligned} \tag{A132}$$

Then substituting $\tilde{m} = m^{LF}$ into the above expression for $E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}}\frac{1}{\tilde{Z}_T}\right]$ yields

$$\begin{aligned} & E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}}\frac{1}{\tilde{Z}_T}\right] \\ = \ & f\left(\tilde{\tau}^2, \tau^{2,\,LF}\right)\exp\left\{\frac{1}{2}\left(m^{LF}\right)^2 g\left(\tilde{\tau}^2, \tau^{2,\,LF}\right)\right\}. \end{aligned} \tag{A133}$$

We now simplify the expressions for f and for g.

$$
\begin{aligned}
& g\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \\
= \; & \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2} - \frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)\,\frac{1}{\left(\tau^{2,\,LF}\right)^2}}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)} \qquad \text{(A134)} \\
& +2\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)\,\frac{1}{\tau^{2,\,LF}\tilde{\tau}^2}}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)} - \frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)\,\frac{1}{\tilde{\tau}^4}}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)} \\
& +\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{1}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\frac{\left(\frac{\gamma-1}{\gamma}\right)^2\left(\delta^{LF}(T)\right)^2}{\left(\tau^{2,\,LF}\right)^2} \\
& -2\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{1}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\tilde{\delta}(T)}{\tau^{2,\,LF}\tilde{\tau}^2} \\
& +\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{1}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)}\frac{\tilde{\delta}^2(T)}{\tilde{\tau}^4} \\
= \; & \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2} \qquad \text{(A135)} \\
& -\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{1}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)} \\
& \left(\begin{array}{c}
\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\frac{1}{\left(\tau^{2,\,LF}\right)^2}\left(\tilde{\delta}(T)\left(1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T\right)-\frac{\gamma-1}{\gamma}\delta^{LF}(T)\right) \\
-2\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\tilde{\delta}(T)}{\tau^{2,\,LF}\tilde{\tau}^2}\left(1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T-1\right) \\
+\tilde{\delta}(T)\,\frac{1}{\tilde{\tau}^4}\left(\delta^{LF}(T)\left(1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T\right)-\tilde{\delta}(T)\right)
\end{array}\right) \\
= \; & \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2} \qquad \text{(A136)} \\
& -\frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\frac{1}{-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)} \\
& \left(\begin{array}{c}
\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,\frac{1}{\left(\tau^{2,\,LF}\right)^2}\left(\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)\left(1-\tilde{\delta}(T)\,T\right)\right) \\
+2\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\tilde{\delta}(T)}{\tau^{2,\,LF}\tilde{\tau}^2}\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)T \\
+\tilde{\delta}(T)\,\frac{1}{\tilde{\tau}^4}\left(\left(-\frac{\gamma-1}{\gamma}\delta^{LF}(T)+\tilde{\delta}(T)\right)\left(-1-\frac{\gamma-1}{\gamma}\delta^{LF}(T)\,T\right)\right)
\end{array}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2} \\
&\quad - \frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T}\left(\begin{array}{c} \frac{\gamma-1}{\gamma}\delta^{LF}(T)\frac{1}{(\tau^{2,\,LF})^2}\left(1-\tilde{\delta}(T)T\right) \\ +2\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\tilde{\delta}(T)}{\tau^{2,\,LF}\tilde{\tau}^2}T \\ +\tilde{\delta}(T)\frac{1}{\tilde{\tau}^4}\left(-1-\frac{\gamma-1}{\gamma}\delta^{LF}(T)T\right)\end{array}\right)
\end{aligned}
\tag{A137}
$$

$$
\begin{aligned}
&= \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2} \\
&\quad - \frac{1}{1+\left(\frac{\gamma-1}{\gamma}\delta^{LF}(T)-\tilde{\delta}(T)\right)T} \\
&\quad \left(\begin{array}{c} \frac{\gamma-1}{\gamma}\delta^{LF}(T)\frac{1}{(\tau^{2,\,LF})^2} - \frac{\gamma-1}{\gamma}\delta^{LF}(T)\frac{1}{(\tau^{2,\,LF})^2}\tilde{\delta}(T)T \\ +2\frac{\frac{\gamma-1}{\gamma}\delta^{LF}(T)\tilde{\delta}(T)}{\tau^{2,\,LF}\tilde{\tau}^2}T \\ -\tilde{\delta}(T)\frac{1}{\tilde{\tau}^4} - \tilde{\delta}(T)\frac{1}{\tilde{\tau}^4}\frac{\gamma-1}{\gamma}\delta^{LF}(T)T\end{array}\right)
\end{aligned}
\tag{A138}
$$

$$
\begin{aligned}
&= \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2} \\
&\quad - \frac{\frac{\gamma-1}{\gamma}\frac{1+\tilde{\tau}^2T}{\tau^{2,\,LF}} - \frac{\gamma-1}{\gamma}\frac{\tilde{\tau}^2}{\tau^{2,\,LF}}T + 2\frac{\gamma-1}{\gamma}T - \frac{1+\tau^{2,\,LF}T}{\tilde{\tau}^2} - \frac{\gamma-1}{\gamma}\frac{\tau^{2,\,LF}}{\tilde{\tau}^2}T}{(1+\tilde{\tau}^2T)(1+\tau^{2,\,LF}T) + \left(\frac{\gamma-1}{\gamma}\tau^{2,\,LF}(1+\tilde{\tau}^2T) - \tilde{\tau}^2(1+\tau^{2,\,LF}T)\right)T}
\end{aligned}
\tag{A139}
$$

$$
\begin{aligned}
&= \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2} \\
&\quad - \frac{\frac{\gamma-1}{\gamma}\frac{1+\tilde{\tau}^2T}{\tau^{2,\,LF}} - \frac{\gamma-1}{\gamma}\frac{\tilde{\tau}^2}{\tau^{2,\,LF}}T + 2\frac{\gamma-1}{\gamma}T - \frac{1+\tau^{2,\,LF}T}{\tilde{\tau}^2} - \frac{\gamma-1}{\gamma}\frac{\tau^{2,\,LF}}{\tilde{\tau}^2}T}{1+\tau^{2,\,LF}T + \frac{\gamma-1}{\gamma}(1+\tilde{\tau}^2T)\tau^{2,\,LF}T}
\end{aligned}
\tag{A140}
$$

$$
\begin{aligned}
&= \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{1}{\tilde{\tau}^2} \\
&\quad - \frac{\frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} + 2\frac{\gamma-1}{\gamma}T - \frac{1}{\tilde{\tau}^2} - \frac{2\gamma-1}{\gamma}\frac{\tau^{2,\,LF}}{\tilde{\tau}^2}T}{1+\tau^{2,\,LF}T + \frac{\gamma-1}{\gamma}(1+\tilde{\tau}^2T)\tau^{2,\,LF}T}
\end{aligned}
\tag{A141}
$$

$$
= \frac{\left(\begin{array}{c} \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - \frac{\gamma-1}{\gamma}\frac{1}{\tau^{2,\,LF}} - 2\frac{\gamma-1}{\gamma}T \\ +\frac{1}{\tilde{\tau}^2} - \frac{1}{\tilde{\tau}^2} + \frac{\gamma-1}{\gamma}T + \left(\frac{\gamma-1}{\gamma}\right)^2(1+\tilde{\tau}^2T)T \\ -\frac{\tau^{2,\,LF}}{\tilde{\tau}^2}T - \frac{\gamma-1}{\gamma}(1+\tilde{\tau}^2)\frac{\tau^{2,\,LF}}{\tilde{\tau}^2}T \\ +\frac{2\gamma-1}{\gamma}\frac{\tau^{2,\,LF}}{\tilde{\tau}^2}T \end{array}\right)}{1+\tau^{2,\,LF}T + \frac{\gamma-1}{\gamma}(1+\tilde{\tau}^2T)\tau^{2,\,LF}T}
\tag{A142}
$$

$$
= \frac{\left(\begin{array}{c} \left[\frac{\gamma-1}{\gamma} + \left(\frac{\gamma-1}{\gamma}\right)^2 - 2\frac{\gamma-1}{\gamma} + \frac{2\gamma-1}{\gamma}\frac{\tau^{2,\,LF}}{\tilde{\tau}^2} - \frac{2\gamma-1}{\gamma}\frac{\tau^{2,\,LF}}{\tilde{\tau}^2}\right]T \\ +\left[\left(\frac{\gamma-1}{\gamma}\right)^2\tilde{\tau}^2 - \frac{\gamma-1}{\gamma}\tau^{2,\,LF}\right]T^2 \end{array}\right)}{1+\tau^{2,\,LF}T + \frac{\gamma-1}{\gamma}(1+\tilde{\tau}^2T)\tau^{2,\,LF}T}
\tag{A143}
$$

$$= \frac{\left[\left(\frac{\gamma-1}{\gamma}\right)^2 - \frac{\gamma-1}{\gamma}\right] T + \frac{\gamma-1}{\gamma}\left[\frac{\gamma-1}{\gamma}\tilde{\tau}^2 - \tau^{2,\,LF}\right] T^2}{1+\tau^{2,\,LF}T + \frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T} \quad \text{(A144)}$$

$$= \frac{\left[\left(\frac{\gamma-1}{\gamma}\right)^2 - \frac{\gamma-1}{\gamma}\right] T + \frac{\gamma-1}{\gamma}\left[\frac{\gamma-1}{\gamma}\tilde{\tau}^2 - \tau^{2,\,LF}\right] T^2}{1+\tau^{2,\,LF}T + \frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T} \quad \text{(A145)}$$

$$= \frac{-\frac{\gamma-1}{\gamma^2}T\left(1-\left(\left(\gamma-1\right)\tilde{\tau}^2 - \gamma\tau^{2,\,LF}\right)T\right)}{1+\tau^{2,\,LF}T + \frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T}. \quad \text{(A146)}$$

Turning to f, we have

$$f\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) = \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\left(1+\frac{\gamma-1}{\gamma}\frac{\tau^{2,\,LF}T}{1+\tau^{2,\,LF}T}\left(1+\tilde{\tau}^2 T\right)\right)^{\frac{1}{2}}} \quad \text{(A147)}$$

$$= \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{\gamma-1}{2\gamma}}}{\frac{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right)^{\frac{1}{2}}}{\left(1+\tau^{2,\,LF}T\right)^{\frac{1}{2}}}} \quad \text{(A148)}$$

$$= \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{2\gamma-1}{2\gamma}}}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right)^{\frac{1}{2}}}. \quad \text{(A149)}$$

We now compute the partial derivative with respect to $\tilde{\tau}^2$ of g.

$$
\begin{aligned}
& \frac{\partial}{\partial \tilde{\tau}^2} g\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \\
= \; & \frac{\left[\begin{array}{c} \left(\frac{\gamma-1}{\gamma}\right)^2 T^2 \left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right) \\ +\frac{1}{\gamma}\left(\frac{\gamma-1}{\gamma}\right)^2 \tau^{2,\,LF}T^3\left(1-\left(\left(\gamma-1\right)\tilde{\tau}^2-\gamma\tau^{2,\,LF}\right)T\right) \end{array}\right]}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right)^2} && \text{(A150)} \\
= \; & \left(\frac{\gamma-1}{\gamma}\right)^2 T^2 \frac{\left[\begin{array}{c} \left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right) \\ +\frac{1}{\gamma}\tau^{2,\,LF}T\left(1-\left(\left(\gamma-1\right)\tilde{\tau}^2-\gamma\tau^{2,\,LF}\right)T\right) \end{array}\right]}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right)^2} && \text{(A151)} \\
= \; & \left(\frac{\gamma-1}{\gamma}\right)^2 T^2 \frac{\left[\begin{array}{c} 1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\tau^{2,\,LF}T \\ +\frac{\gamma-1}{\gamma}\tilde{\tau}^2\tau^{2,\,LF}T^2+\frac{1}{\gamma}\tau^{2,\,LF}T \\ -\frac{\gamma-1}{\gamma}\tau^{2,\,LF}\tilde{\tau}^2T^2+\left(\tau^{2,\,LF}\right)^2T^2 \end{array}\right]}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right)^2} && \text{(A152)} \\
= \; & \left(\frac{\gamma-1}{\gamma}\right)^2 T^2 \frac{\left[1+\tau^{2,\,LF}T+\tau^{2,\,LF}T+\left(\tau^{2,\,LF}\right)^2T^2\right]}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right)^2} && \text{(A153)} \\
= \; & \left(\frac{\gamma-1}{\gamma}\right)^2 T^2 \frac{\left(1+\tau^{2,\,LF}T\right)^2}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right)^2}. && \text{(A154)}
\end{aligned}
$$

We will also require the partial derivative of f with respect to $\tilde{\tau^2}$.

$$
\begin{aligned}
& \frac{\partial}{\partial \tilde{\tau}^2} f\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \\
= \; & -\frac{1}{2}\frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{2\gamma-1}{2\gamma}}}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right)^{\frac{3}{2}}}\frac{\gamma-1}{\gamma}\tau^{2,\,LF}T^2. && \text{(A155)}
\end{aligned}
$$

Now observe that

$$
\begin{aligned}
& \frac{\partial}{\partial \tilde{\tau}^2} E^* \left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}} \frac{1}{\tilde{Z}_T} \right] \\
= \; & \frac{\partial}{\partial \tilde{\tau}^2} f\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \exp\left\{ \frac{1}{2} \left(m^{LF}\right)^2 g\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \right\} && \text{(A156)} \\
= \; & \left[\frac{\partial}{\partial \tilde{\tau}^2} f\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) + f\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \frac{1}{2} \left(m^{LF}\right)^2 \frac{\partial}{\partial \tilde{\tau}^2} g\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \right] && \text{(A157)} \\
& \exp\left\{ \frac{1}{2} \left(m^{LF}\right)^2 g\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \right\} \\
= \; & \left[\begin{array}{c} -\frac{1}{2} \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{2\gamma-1}{2\gamma}}}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}(1+\tilde{\tau}^2 T)\tau^{2,\,LF}T\right)^{\frac{3}{2}}} \frac{\gamma-1}{\gamma} \tau^{2,\,LF} T^2 \\ + \left(m^{LF}\right)^2 \left(\frac{\gamma-1}{\gamma}\right)^2 T^2 \frac{\left(1+\tau^{2,\,LF}T\right)^2}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}(1+\tilde{\tau}^2 T)\tau^{2,\,LF}T\right)^2} \\ \times \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{2\gamma-1}{2\gamma}}}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}(1+\tilde{\tau}^2 T)\tau^{2,\,LF}T\right)^{\frac{1}{2}}} \end{array} \right] && \text{(A158)} \\
& \exp\left\{ \frac{1}{2} \left(m^{LF}\right)^2 g\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \right\} \\
= \; & \left[\begin{array}{c} -\frac{1}{2} \frac{\gamma-1}{\gamma} \tau^{2,\,LF} T^2 \left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right) \\ + \left(m^{LF}\right)^2 \left(\frac{\gamma-1}{\gamma}\right)^2 T^2 \left(1+\tau^{2,\,LF}T\right)^2 \end{array} \right] && \text{(A159)} \\
& \frac{\left(1+\tau^{2,\,LF}T\right)^{\frac{2\gamma-1}{2\gamma}}}{\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2 T\right)\tau^{2,\,LF}T\right)^{\frac{5}{2}}} \exp\left\{ \frac{1}{2} \left(m^{LF}\right)^2 g\left(\tilde{\tau}^2, \tau^{2,\,LF}\right) \right\}.
\end{aligned}
$$

The equalities follow from substituting the forms derived above for f and for the partials of f and g with respect to $\tilde{\tau}^2$ into the expression for the partial derivative, and then performing elementary algebra. Noting that the quantity multiplying the term in square brackets is always positive, we focus on the term in square brackets.

$$
\begin{aligned}
&-\frac{1}{2}\frac{\gamma-1}{\gamma}\tau^{2,\,LF}T^2\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tilde{\tau}^2T\right)\tau^{2,\,LF}T\right)\\
&+\left(m^{LF}\right)^2\left(\frac{\gamma-1}{\gamma}\right)^2T^2\left(1+\tau^{2,\,LF}T\right)^2\\
= \;&-\frac{1}{2}\frac{\gamma-1}{\gamma}\tau^{2,\,LF}T^2\left(\begin{array}{c}1+\tau^{2,\,LF}T\\ +\frac{\gamma-1}{\gamma}\left(1+\tau^{2,\,LF}T-\tau^{2,\,LF}T+\tilde{\tau}^2T\right)\\ \times\tau^{2,\,LF}T\end{array}\right) \quad \text{(A160)}\\
&+\left(m^{LF}\right)^2\left(\frac{\gamma-1}{\gamma}\right)^2T^2\left(1+\tau^{2,\,LF}T\right)^2\\
= \;&\frac{1}{2}\left(\frac{\gamma-1}{\gamma}\right)^2\left(\tau^{2,\,LF}\right)^2T^4\left(\tau^{2,\,LF}-\tilde{\tau}^2\right) \quad \text{(A161)}\\
&-\frac{1}{2}\frac{\gamma-1}{\gamma}\tau^{2,\,LF}T^2\left(1+\tau^{2,\,LF}T+\frac{\gamma-1}{\gamma}\left(1+\tau^{2,\,LF}T\right)\tau^{2,\,LF}T\right)\\
&+\left(m^{LF}\right)^2\left(\frac{\gamma-1}{\gamma}\right)^2T^2\left(1+\tau^{2,\,LF}T\right)^2\\
= \;&\frac{1}{2}\left(\frac{\gamma-1}{\gamma}\right)^2\left(\tau^{2,\,LF}\right)^2T^4\left(\tau^{2,\,LF}-\tilde{\tau}^2\right) \quad \text{(A162)}\\
&+\frac{1}{2}\frac{\gamma-1}{\gamma}T^2\left(1+\tau^{2,\,LF}T\right)\left\{\begin{array}{c}-\tau^{2,\,LF}\left(1+\frac{\gamma-1}{\gamma}\tau^{2,\,LF}T\right)\\ +\left(m^{LF}\right)^2\frac{\gamma-1}{\gamma}\left(1+\tau^{2,\,LF}T\right)\end{array}\right\}.
\end{aligned}
$$

The term in braces on the last line of the display above is a quadratic in $\tau^{2,\,LF}$. Let us analyze this quadratic. First, note that the coefficient on $\left(\tau^{2,\,LF}\right)^2$ is $-\frac{\gamma-1}{\gamma}T<0$, so the quadratic is concave in $\tau^{2,\,LF}$. Rewrite the quadratic as

$$
\begin{aligned}
&\left(-\frac{\gamma-1}{\gamma}T\right)\left(\tau^{2,\,LF}\right)^2 \quad \text{(A163)}\\
&+\left(\left(m^{LF}\right)^2\frac{\gamma-1}{\gamma}T-1\right)\tau^{2,\,LF}\\
&+\left(\left(m^{LF}\right)^2\frac{\gamma-1}{\gamma}\right).
\end{aligned}
$$

The quadratic formula now yields the roots of this quadratic, which are

$$\tau_1^2 = \tag{A164}$$
$$\frac{1}{2}\left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma}T} + \sqrt{\left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma}T}\right)^2 + 4\frac{\left(m^{LF}\right)^2}{T}}\right)$$

$$\tau_2^2 = \tag{A165}$$
$$\frac{1}{2}\left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma}T} - \sqrt{\left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma}T}\right)^2 + 4\frac{\left(m^{LF}\right)^2}{T}}\right).$$

Since $\left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma}T}\right)^2 < \left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma}T}\right)^2 + 4\frac{\left(m^{LF}\right)^2}{T}$ for all $m^{LF} \neq 0$, the second root, τ_2^2, is negative for all $m^{LF} \neq 0$. If $m^{LF} = 0$, then $\tau_2^2 = -\frac{1}{\frac{\gamma-1}{\gamma}T} < 0$. Thus, $\tau_2^2 < 0$ for any value of m^{LF}.

By definition, a variance is nonnegative. Thus, we may restrict attention to the root τ_1^2. Observe that $\tau_1^2 = \tau^{2,*}$ by the definition of $\tau^{2,*}$ in the statement of Theorem 7. We have shown that the other root of the quadratic, τ_2^2, is negative, and that the quadratic is concave; together, these facts imply that the quadratic is positive for any nonnegative $\tau^2 < \tau^{2,*}$ and negative for any $\tau^2 > \tau^{2,*}$. Finally, it is obvious (by the very definition of a root) that the quadratic is zero for $\tau^2 = \tau^{2,*}$.

We consider three exhaustive and mutually exclusive cases. First, it may be that $\tau^{2,LF} < \tau^{2,*}$. By the definition of $\tau^{2,LF}$ in Theorem 7, this implies that $\tau^{2,LF} = \frac{\overline{\nu}^2}{\sigma^2} < \tau^{2,*}$, so that $\tau^{2,LF}$ is a boundary solution that takes on the largest value of τ^2 permitted by the interval $\left[\frac{\underline{\nu}^2}{\sigma^2}, \frac{\overline{\nu}^2}{\sigma^2}\right]$. By the logic above, the quadratic in braces in equation (A163) is positive in this case. Further, for any $\tilde{\tau}^2 \in \left[\frac{\underline{\nu}^2}{\sigma^2}, \frac{\overline{\nu}^2}{\sigma^2}\right]$, we have $\tau^{2,LF} - \tilde{\tau}^2 = \frac{\overline{\nu}^2}{\sigma^2} - \tilde{\tau}^2 \geq \frac{\overline{\nu}^2}{\sigma^2} - \frac{\overline{\nu}^2}{\sigma^2} = 0$. Thus, the partial derivative of $E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}} \frac{1}{\tilde{Z}_T}\right]$ with respect to $\tilde{\tau}^2$ is always positive in this case. $E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}} \frac{1}{\tilde{Z}_T}\right]$ is maximized, then, at the largest possible value of $\tilde{\tau}^2$, that is, at $\tilde{\tau}^2 = \frac{\overline{\nu}^2}{\sigma^2} = \tau^{2,LF}$.

In the first of the three possible cases, we have shown that $\tilde{\tau}^2 = \tau^{2,\,LF}$ maximizes the objective function.

Second, it may be that $\tau^{2,\,LF} > \tau^{2,\,*}$. By the definition of $\tau^{2,\,LF}$ in Theorem 7, this implies that $\tau^{2,\,LF} = \frac{\underline{\nu}^2}{\sigma^2} > \tau^{2,\,*}$ so that $\tau^{2,\,LF}$ is a boundary solution that takes on the smallest value of τ^2 permitted by the interval $\left[\frac{\underline{\nu}^2}{\sigma^2}, \frac{\overline{\nu}^2}{\sigma^2}\right]$. By the logic above, the quadratic in braces in equation (A163) is negative in this case. Further, for any $\tilde{\tau}^2 \in \left[\frac{\underline{\nu}^2}{\sigma^2}, \frac{\overline{\nu}^2}{\sigma^2}\right]$ we have $\tau^{2,\,LF} - \tilde{\tau}^2 = \frac{\underline{\nu}^2}{\sigma^2} - \tilde{\tau}^2 \leq \frac{\underline{\nu}^2}{\sigma^2} - \frac{\underline{\nu}^2}{\sigma^2} = 0$. Thus, the partial derivative of $E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}} \frac{1}{\tilde{Z}_T}\right]$ with respect to $\tilde{\tau}^2$ is always negative in this case. $E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}} \frac{1}{\tilde{Z}_T}\right]$ is maximized, then, at the smallest possible value of $\tilde{\tau}^2$, that is, at $\tilde{\tau}^2 = \frac{\underline{\nu}^2}{\sigma^2} = \tau^{2,\,LF}$. In the second of the three possible cases, we have shown that $\tilde{\tau}^2 = \tau^{2,\,LF}$ maximizes the objective function.

Third (and finally), it may be that $\tau^{2,\,LF} = \tau^{2,\,*}$. Then, as shown above, the quadratic in braces in equation (A163) is zero. Further, $\tau^{2,\,LF} - \tilde{\tau}^2$ is (obviously) zero for $\tilde{\tau}^2 = \tau^{2,\,LF}$, positive for $\tilde{\tau}^2 < \tau^{2,\,LF}$, and negative for $\tilde{\tau}^2 > \tau^{2,\,LF}$. Thus, the partial derivative of $E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}} \frac{1}{\tilde{Z}_T}\right]$ with respect to $\tilde{\tau}^2$ is zero for $\tilde{\tau}^2 = \tau^{2,\,LF}$, positive for $\tilde{\tau}^2 < \tau^{2,\,LF}$, and negative for $\tilde{\tau}^2 > \tau^{2,\,LF}$. $E^*\left[\overline{Z}_T^{\frac{\gamma-1}{\gamma}} \frac{1}{\tilde{Z}_T}\right]$ is maximized, then, at $\tilde{\tau}^2 = \tau^{2,\,LF}$. In the third and final of the three possible cases, we have shown that $\tilde{\tau}^2 = \tau^{2,\,LF}$ maximizes the objective function.

We have now completed the proof that, if an investor with power utility over terminal wealth behaves optimally with respect to a normal prior distribution with mean λ^{LF} and variance $\nu^{2,\,LF}$, then the expected-utility-minimizing prior in the set Γ given the investor's portfolio choice rule is, in fact, the normal distribution with mean λ^{LF} and variance $\nu^{2,\,LF}$. This implies that the normal distribution with mean λ^{LF} and variance $\nu^{2,\,LF}$, which was our candidate least-favorable prior distribution, is indeed the least-favorable prior distribution. □

Proof of Proposition 5: We begin by reproducing the definition of $\tau^{2,*}$:

$$\tau^{2,*} = \frac{1}{2}\left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma}T} + \sqrt{\left(\left(m^{LF}\right)^2 - \frac{1}{\frac{\gamma-1}{\gamma}T}\right)^2 + 4\frac{\left(m^{LF}\right)^2}{T}}\right). \tag{A166}$$

We wish to show that

$$\tau^{2,*} \in \left[\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2, \left(m^{LF}\right)^2\right]. \tag{A167}$$

In the proof of Theorem 7, it was shown that $\tau^{2,*}$ is a positive root of the quadratic

$$\begin{aligned} &\left(-\frac{\gamma-1}{\gamma}T\right)\left(\tau^2\right)^2 + \left(\left(m^{LF}\right)^2\frac{\gamma-1}{\gamma}T - 1\right)\tau^2 \\ &+ \left(\left(m^{LF}\right)^2\frac{\gamma-1}{\gamma}\right) \\ = &\left(-\frac{\gamma-1}{\gamma}T\right)\tau^2\left(\tau^2 - \left(m^{LF}\right)^2\right) + \left(\left(m^{LF}\right)^2\frac{\gamma-1}{\gamma} - \tau^2\right). \end{aligned} \tag{A168}$$

If $\tau^2 < \frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2$, both of the terms to the right of the equality sign in the equation above must be positive, so the whole quadratic must be positive, and $\tau^{2,*} \geq \frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2$ since $\tau^{2,*}$ is a root of the quadratic. On the other hand, if $\tau^2 > \left(m^{LF}\right)^2$, both of the terms to the right of the equality sign in the equation above must be negative, so the whole quadratic must be negative, and $\tau^{2,*} \leq \left(m^{LF}\right)^2$ since $\tau^{2,*}$ is a root of the quadratic.

Together, these inequalities restrict $\tau^{2,*}$ to the desired interval. Now, by the definition of $\tau^{2,LF}$ in terms of $\tau^{2,*}$, we have that $\tau^{2,LF} \in \left[\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2, \left(m^{LF}\right)^2\right] \cap \left[\frac{\underline{\nu}^2}{\sigma^2}, \frac{\overline{\nu}^2}{\sigma^2}\right] \subset \left[\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2, \left(m^{LF}\right)^2\right]$, which proves the first assertion of the proposition.

We now evaluate the limits appearing in the statement of the proposition. First,

$$\lim_{T\to\infty}\left\{\sigma^2\tau^{2,*}\right\} = \sigma^2\frac{1}{2}\left(\left(m^{LF}\right)^2+\sqrt{\left(\left(m^{LF}\right)^2\right)^2}\right) \tag{A169}$$

$$= \sigma^2\frac{1}{2}\left(2\left(m^{LF}\right)^2\right) \tag{A170}$$

$$= \sigma^2\left(m^{LF}\right)^2. \tag{A171}$$

Next,

$$\lim_{T\to 0}\left\{\sigma^2\tau^{2,*}\right\}$$

$$= \sigma^2\lim_{T\to 0}\left\{\frac{1}{2}\frac{\left(\begin{array}{c}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2T-1\\ +\left(\begin{array}{c}\left(\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2T-1\right)^2\\ +4\left(\frac{\gamma-1}{\gamma}\right)^2\left(m^{LF}\right)^2T\end{array}\right)^{\frac{1}{2}}\end{array}\right)}{\frac{\gamma-1}{\gamma}T}\right\} \tag{A172}$$

$$= \sigma^2\frac{1}{2}\frac{\lim_{T\to 0}\left\{\begin{array}{c}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2\\ +\frac{1}{2}\frac{2\left(\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2T-1\right)\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2+4\left(\frac{\gamma-1}{\gamma}\right)^2\left(m^{LF}\right)^2}{\sqrt{\left(\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2T-1\right)^2+4\left(\frac{\gamma-1}{\gamma}\right)^2\left(m^{LF}\right)^2T}}\end{array}\right\}}{\frac{\gamma-1}{\gamma}} \tag{A173}$$

$$= \sigma^2\frac{1}{2}\frac{\lim_{T\to 0}\left\{\begin{array}{c}\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2-\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2\\ +2\left(\frac{\gamma-1}{\gamma}\right)^2\left(m^{LF}\right)^2\end{array}\right\}}{\frac{\gamma-1}{\gamma}} \tag{A174}$$

$$= \sigma^2\frac{1}{2}\left[\frac{2\left(\frac{\gamma-1}{\gamma}\right)^2\left(m^{LF}\right)^2}{\frac{\gamma-1}{\gamma}}\right] \tag{A175}$$

$$= \sigma^2\frac{\left(\frac{\gamma-1}{\gamma}\right)^2\left(m^{LF}\right)^2}{\frac{\gamma-1}{\gamma}} \tag{A176}$$

$$= \sigma^2\frac{\gamma-1}{\gamma}\left(m^{LF}\right)^2, \tag{A177}$$

where the second equality follows from L'Hopital's rule and the fact that both the numerator and the denominator of the previous expression converge to zero as $T \to 0$.

The final limit result is

$$\begin{aligned}
& \lim_{\gamma\to\infty} \left\{\sigma^2\tau^{2,*}\right\} \\
= & \ \sigma^2 \frac{1}{2}\left(\left(m^{LF}\right)^2 - \frac{1}{T} + \sqrt{\left(\left(m^{LF}\right)^2 - \frac{1}{T}\right)^2 + 4\frac{\left(m^{LF}\right)^2}{T}}\right) \quad \text{(A178)} \\
= & \ \sigma^2 \frac{1}{2}\left(\begin{array}{c} \left(m^{LF}\right)^2 - \frac{1}{T} \\ + \left(\left(m^{LF}\right)^4 - 2\left(m^{LF}\right)^2 \frac{1}{T} + \frac{1}{T^2} + 4\frac{\left(m^{LF}\right)^2}{T}\right)^{\frac{1}{2}} \end{array}\right) \quad \text{(A179)} \\
= & \ \sigma^2 \frac{1}{2}\left(\left(m^{LF}\right)^2 - \frac{1}{T} + \sqrt{\left(m^{LF}\right)^4 + 2\left(m^{LF}\right)^2 \frac{1}{T} + \frac{1}{T^2}}\right) \quad \text{(A180)} \\
= & \ \sigma^2 \frac{1}{2}\left(\left(m^{LF}\right)^2 - \frac{1}{T} + \sqrt{\left(\left(m^{LF}\right)^2 + \frac{1}{T}\right)^2}\right) \quad \text{(A181)} \\
= & \ \sigma^2 \frac{1}{2}\left(\left(m^{LF}\right)^2 - \frac{1}{T} + \left(m^{LF}\right)^2 + \frac{1}{T}\right) \quad \text{(A182)} \\
= & \ \sigma^2 \frac{1}{2}\left(2\left(m^{LF}\right)^2\right) \quad \text{(A183)} \\
= & \ \sigma^2 \left(m^{LF}\right)^2. \quad \text{(A184)}
\end{aligned}$$

It remains only to prove that $m^{LF} = 0 \Rightarrow \sigma^2\tau^{2,*} = 0$. If $m^{LF} = 0$, then

$$\begin{aligned}
\tau^{2,*} &= \frac{1}{2}\left(-\frac{1}{\frac{\gamma-1}{\gamma}T} + \sqrt{\left(-\frac{1}{\frac{\gamma-1}{\gamma}T}\right)^2}\right) \quad \text{(A185)} \\
&= \frac{1}{2}\left(-\frac{1}{\frac{\gamma-1}{\gamma}T} + \frac{1}{\frac{\gamma-1}{\gamma}T}\right) \quad \text{(A186)} \\
&= 0, \quad \text{(A187)}
\end{aligned}$$

so the desired implication holds and the proof of the proposition is complete. □

References

Cox, J. C. and C. Huang (1989). Optimum consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory 49*, 33–83.

Cox, J. C. and C. Huang (1991). A variational problem arising in financial economics. *Journal of Mathematical Economics 20*, 465–487.

Cvitanič, J., A. Lazrak, L. Martellini, and F. Zapatero (2002). Revisiting Treynor and Black (1973): An intertemporal model of active portfolio management. University of Southern California working paper.

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* San Diego, California: Academic Press.

Fishburn, P. C. (1979). *Utility Theory for Decision Making, 2nd edition* (Second ed.). Huntington, New York: Robert E. Krieger Publishing Co., Inc.

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

Liptser, R. S. and A. N. Shiryayev (2001). *Statistics of Random Processes I.* New York, New York: Springer-Verlag.

Wachter, J. A. (2002). Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets. *Journal of Financial and Quantitative Analysis 37*, 63–91.