CMS-EMS
Center for Mathematical Studies in Economics
And Management Science

Discussion Paper #1566

OPTIMAL REFERENCE POINTS AND ANTICIPATION

Todd Sarver
Northwestern University

June 18, 2012

JEL Classification: D03, D81, G12

Keywords: reference dependence, loss aversion, anticipatory utility, equity premium puzzle, Rabin paradox

NORTHWESTERN
UNIVERSITY

Optimal Reference Points and Anticipation*

Todd Sarver†

June 18, 2012

Abstract

This paper considers a model of reference-dependent utility in which the individual makes a conscious choice of her reference point for future consumption. The model incorporates the combination of loss aversion and anticipatory utility as competing forces in the determination of the optimal reference point: anticipating better outcomes boosts current utility but also raises the reference level for future consumption, making the individual more susceptible to losses. A central focus of the paper is on the implications of this model of *Optimal Anticipation* for attitudes toward risk in dynamic environments. The main representation is formulated in an infinite-horizon framework, and axiomatic foundations are provided. I also describe special cases and show in particular that recursive expected utility in the sense of Epstein and Zin (1989) and Kreps and Porteus (1978) can be reinterpreted in terms of optimal anticipation and loss aversion. Finally, I describe a homogeneous version of the model and apply it to a portfolio choice problem. I show that asset pricing for the Optimal Anticipation model is based on simple modifications of standard Euler equations. While maintaining tractability, this model is rich enough to permit first-order risk aversion and can overcome several deficits of standard expected utility, such as the equity premium puzzle and Rabin's paradox.

*First version September 2011. I thank David Ahn, Eddie Dekel, David Dillenberger, Larry Epstein, Faruk Gul, R. Vijay Krishna, Bart Lipman, Mark Machina, Jonathan Parker, Wolfgang Pesendorfer, Marciano Siniscalchi, Costis Skiadas, and seminar participants at Boston University, Stanford University, Northwestern University, Duke University, Princeton University, University of Pennsylvania, and Caltech for helpful comments and discussions. This paper was previously circulated under the title "Optimal Anticipation."

†Northwestern University, Department of Economics, 2001 Sheridan Road, Evanston, IL 60208. Email: tsarver@northwestern.edu.

1 Introduction

1.1 Motivation

When faced with any uncertain prospect, an individual's satisfaction with the outcome often depends not only on its intrinsic value, but also on how it compares to what was anticipated. For example, a moderate return on an investment may be disappointing if an exceptional return was anticipated; a week-long vacation at the beach may not seem quite so enjoyable to an individual who had anticipated taking a month-long yacht cruise; an individual may be pleased with a low-wage employment opportunity if she had anticipated being unemployed. Insights of this sort are now well-established in economics and are examples of what is referred to as reference-dependent utility.

The incorporation of reference dependence and loss aversion into economic theory dates back to Markowitz (1952) and Kahneman and Tversky (1979). Markowitz first proposed defining utility in terms of gains and losses in order to refine the Friedman-Savage hypothesis.[1] Kahneman and Tversky found that subjects in experiments exhibited a wide variety of violations of expected-utility theory; further, they showed that many of these observations could be explained if individuals evaluate the same outcome differently depending on whether it is a gain or a loss. However, to determine if an outcome will be viewed as a gain or a loss, one must take a stand on what the reference point will be in a particular context. Therefore, an important aspect of any model of reference-dependent utility is the specification of *reference-point formation*. Two popular approaches suggested in recent work are "history-dependent" reference-point formation (Benartzi and Thaler (1995); Barberis and Huang (2001); Barberis, Huang, and Santos (2001)) and forward-looking "equilibrium-based" reference-point formation (Kőszegi and Rabin (2006, 2007)).

In this paper, I develop an alternative approach to reference-point formation based on optimality. I consider an individual who (optimally) engages in psychological planning to better prepare herself mentally for the future outcomes that might arise. I refer to this process of mental preparation as *anticipation*. For example, an individual may choose to look forward to a good outcome or alternatively prepare herself for a bad one. This choice of anticipation by the individual then forms a reference point against which future outcomes are measured. Since the future reference point is based on current anticipation and is chosen optimally, the behavior in this model will be described as *Optimal Anticipation*.

The individual in this model is unconstrained in her choice of reference point, but what

[1]To explain why individuals might purchase both insurance and lotteries at all wealth levels, Markowitz suggested that gains and losses be evaluated relative to current wealth using a gain-loss function that is convex for small-to-moderate gains.

factors should determine her optimal choice? I assume the optimal level of anticipation balances two competing forces. The first is loss aversion—bad outcomes are more painful when a good outcome was anticipated. The second is anticipatory utility—the individual receives a boost in current utility from looking forward to a good outcome in the future. The first feature is common to most models of reference-dependent utility, and provides a benefit to lowering the reference point. In other models of loss aversion, the individual would always like to lower her reference point, but does not have the freedom to do so since the reference point is disciplined by either expectations or history. In contrast, in this model the individual can make a conscious choice of her reference point, and therefore always has the option to "prepare for the worst." However, she might *choose* to anticipate a better outcome in order to enjoy higher anticipatory utility, even though this raises her reference point for future consumption.

The trade-off between loss aversion and anticipatory utility is an important novel aspect of this paper, and is arguably quite intuitive in situations involving risk. For example, an individual may enjoy looking forward to an upcoming vacation. But if for some reason she is ultimately unable to take the vacation, she will experience a loss, presumably a larger one if she spent a great deal of time looking forward to the trip. The extent to which the individual anticipates the vacation therefore naturally depends on how likely it is to be feasible. If it is very unlikely, she would spend little time looking forward to it to lessen her disappointment in the event she cannot go. On the other hand, if the vacation becomes more likely, she might allow herself to enjoy greater anticipation since the risk of disappointment is smaller.

The main focus of this paper is on the effect of optimal anticipation and loss aversion on attitudes toward risk. Using a simple two-period version of the model, I illustrate how several deficits of the standard expected-utility model—the equity premium puzzle and Rabin (2000)'s paradox—can be overcome using the Optimal Anticipation model. The tractability of the model also facilitates more sophisticated applications in general dynamic environments. In particular, it is easily formulated recursively in an infinite-horizon setting. One of the main results is an axiomatic analysis of this recursive model, which characterizes its straightforward behavioral implications. I then describe a homogeneous version of the model and apply it to a portfolio choice problem. I show that asset pricing for the Optimal Anticipation model is based on simple modifications of standard Euler equations. While maintaining tractability, these conditions are rich enough to overcome many limitations of standard expected utility, such as the aforementioned puzzles.

1.2 Model Overview

Before describing the general model, I begin by formalizing the compromise between loss aversion and anticipatory utility using several illustrative examples. To simplify the exposition of ideas in the introduction, I restrict to a two-period version of the model, and defer the infinite-horizon model to Section 3.

In the first example of the Optimal Anticipation representation, anticipatory utility and loss aversion both enter into overall utility in a linear way.

Example 1 (Linear Gain-Loss Function) Consider the following utility representation for consumption c_t in the current period t and consumption in $t+1$ given by the random variable c_{t+1}:

$$u_t(c_t) + \beta \max_{\gamma \in \mathbb{R}} \mathbb{E}_t \Big[\phi\big(u_{t+1}(c_{t+1}) \big| \gamma\big) \Big] \tag{1}$$

where

$$\phi(u|\gamma) = \underbrace{\alpha u}_{\substack{\text{consumption}\\ \text{utility}}} + \underbrace{(1-\alpha)\gamma}_{\substack{\text{anticipatory}\\ \text{utility}}} + \underbrace{\begin{cases} a(u-\gamma) & \text{if } u < \gamma \\ b(u-\gamma) & \text{if } u \geq \gamma \end{cases}}_{\substack{\text{gain-loss}\\ \text{utility}}}$$

for some parameters $\alpha \in [0,1]$ and $a \geq b \geq 0$. The scalar γ in this utility representation can be interpreted as a choice of anticipation by the individual at time t about her utility at time $t+1$. Under this interpretation, the term $(1-\alpha)\gamma$ in the definition of $\phi(\cdot|\gamma)$ is the anticipation utility of the individual—by choosing anticipation level γ, she gets a fixed anticipatory utility of $(1-\alpha)\gamma$, regardless of the realized consumption at $t+1$. The term αu corresponds to standard consumption utility, and the parameter α determines the relative weight of anticipatory utility and consumption utility. The last term in the definition of $\phi(\cdot|\gamma)$ incorporates gain-loss utility relative to the reference point γ. Thus, anticipation at time t forms a reference point for consumption at time $t+1$, and losses relative to this reference level have a greater impact on utility than gains.

In Example 1, consumption utility and gain-loss utility appear as two separate components of overall utility in the function $\phi(\cdot|\gamma)$. While this separation could potentially be appealing from a conceptual point of view, it is in fact without loss of generality to consolidate these two components: The linear gain-loss function in this example can be expressed as

$$\phi(u|\gamma) = \gamma + \begin{cases} \lambda^l(u-\gamma) & \text{if } u < \gamma \\ \lambda^g(u-\gamma) & \text{if } u \geq \gamma, \end{cases} \tag{2}$$

where $\lambda^l \equiv \alpha + a \geq \lambda^g \equiv \alpha + b$. This simplified formulation of the linear gain-loss function makes some of its key properties more transparent. It is immediate from Equation (2) that $\phi(u|\gamma) = u$ if $u = \gamma$. Moreover, the slope of $\phi(\cdot|\gamma)$ to the left of γ is λ^l, and the slope to the right of γ is λ^g. Note that standard additively-separable expected utility corresponds to the special case of $\lambda^l = \lambda^g = 1$.

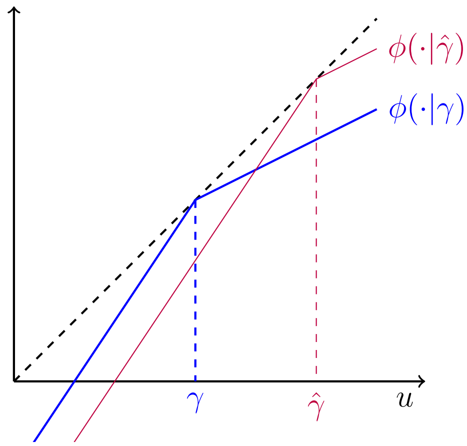

Figure 1: Illustration of linear gain-loss function from Example 1

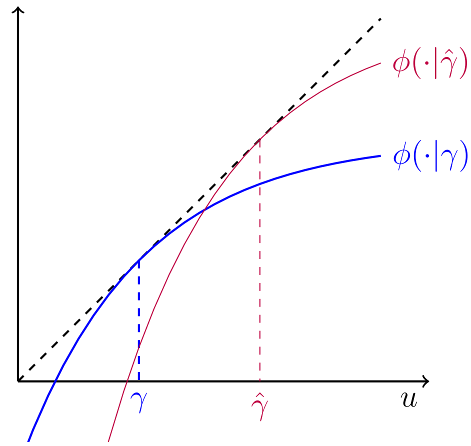

Figure 2: Illustration of non-linear gain-loss function from Example 2

Figure 1 provides an illustration of the linear gain-loss function when $\lambda^l > 1 > \lambda^g > 0$. In this case, $\phi(u|\gamma) < u$ for all $u \neq \gamma$. Thus, for any deterministic future consumption, the individual will optimally choose the anticipation level $\gamma = u_{t+1}(c_{t+1})$. Intuitively, it is desirable to correctly target anticipation if possible—the individual would like to anticipate higher future utility due to the current enjoyment it brings, but does not wish to overshoot in her anticipation due to the loss she will feel from falling short of this benchmark. Of course, for stochastic future consumption, realized future utility will necessarily differ from any choice of γ with positive probability. To show the individual's trade-off when facing random consumption, Figure 1 illustrates that increasing anticipation from γ to $\hat{\gamma}$ has two effects: First, it increases anticipation utility and therefore results in higher overall utility for high realizations of u_{t+1}. Second, for low realizations of u_{t+1}, overall utility is decreased due to loss aversion. The individual chooses γ to balance these two effects. In particular, I show in Section 4.1 that her optimal choice of γ for this linear gain-loss function is a quantile of $u_{t+1}(c_{t+1})$ determined by the parameters λ^l and λ^g; in the case of a continuous distribution, the formula is $\Pr[u_{t+1}(c_{t+1}) \leq \gamma] = \frac{1-\lambda^g}{\lambda^l-\lambda^g}$.

The gain-loss function in Equation (2) will prove useful in several applications described later, but it is just one possible specification. The following example illustrates a non-linear alternative.

Example 2 (Exponential Gain-Loss Function) Consider a utility representation for two-period consumption as in Equation (1), but where the gain-loss function takes the form

$$\phi(u|\gamma) = \gamma + \frac{1}{\theta} - \frac{1}{\theta}\exp(-\theta(u-\gamma))$$

for some parameter $\theta > 0$. For fixed γ, the function $\phi(\cdot|\gamma)$ is a standard exponential utility function with coefficient of absolute risk aversion θ, normalized so that $\phi(u|\gamma) \leq u$, with equality at $u = \gamma$. Since losses relative to the reference level γ have a greater impact on overall utility than gains, higher γ is better for high realizations of u_{t+1} and worse for low realizations of u_{t+1}, similar to the previous example. Figure 2 provides an illustration of this gain-loss function.

In both of the preceding examples, the anticipation choices were utility values γ. However, individuals may be able to engage in psychological preparation that is more sophisticated than simply planning for a particular ex post utility value. For instance, it is conceivable that individuals prepare themselves differently for large-scale risks than small-scale risks. As a concrete example, perhaps in addition to targeting a utility level γ, they are also able to decrease the concavity of the gain-loss function by sacrificing some anticipatory utility. To model such behavior, at least a two-parameter gain-loss function is required. In general, there could be many strategies of mental preparation or anticipation available to an individual. Thus, while the parametric gain-loss functions in these examples are analytically simple and useful in practice, the general Optimal Anticipation model will permit a non-parametric collection of anticipation strategies.

The parametric family of gain-loss functions $\{\phi(\cdot|\gamma) : \gamma \in \mathbb{R}\}$ in Equation (1) can be extended easily to a non-parametric family of functions Φ. Maintaining the two-period setting for simplicity, the general Optimal Anticipation representation takes the form

$$u_t(c_t) + \beta \sup_{\phi\in\Phi} \mathbb{E}_t\Big[\phi\big(u_{t+1}(c_{t+1})\big)\Big], \tag{3}$$

where Φ is any collection of nondecreasing functions $\phi : \mathbb{R} \to \mathbb{R}$ that satisfies

$$\sup_{\phi\in\Phi} \phi(u) = u, \qquad \forall u \in \mathbb{R}. \tag{4}$$

Examples 1 and 2 both had the feature that $\phi(u|\gamma) \leq u$, with equality at $u = \gamma$. Equation (4) is a generalization of this property: it ensures that for deterministic consumption streams, the Optimal Anticipation model reduces to standard additively-separable utility $u_t(c_t) + \beta u_{t+1}(c_{t+1})$. However, overall utility could be lower than in the standard model in the case of stochastic consumption:

$$\sup_{\phi\in\Phi} \mathbb{E}_t\Big[\phi\big(u_{t+1}(c_{t+1})\big)\Big] \leq \mathbb{E}_t\Big[\sup_{\phi\in\Phi} \phi\big(u_{t+1}(c_{t+1})\big)\Big] = \mathbb{E}_t\Big[u_{t+1}(c_{t+1})\Big].$$

Thus, one important implication of anticipatory utility and loss aversion in this model is an increase in risk aversion. Intuitively, if accurate anticipation is needed to obtain the maximum overall utility from a particular outcome, then randomness in the outcome makes it difficult to plan optimally.

In addition to allowing for increased risk aversion, the Optimal Anticipation model also permits risk preferences that are not possible with expected utility. In particular, if the gain-loss function is kinked as in Example 1, then preferences exhibit *first-order risk aversion* (Segal and Spivak (1990)). This is an important feature of the model that will prove useful in applications. Closely related is the model's ability to explain both the equity premium puzzle and Rabin's paradox (see Section 2). And as with many other models of non-expected-utility preferences, it also resolves the classic Allais paradox.

Although the model permits non-expected-utility preferences, there is an important connection to standard expected utility that enhances the tractability of the model. After choosing her anticipation $\phi \in \Phi$ in period t, the individual evaluates risky consumption c_{t+1} using the composite expected-utility function $\phi \circ u_{t+1}$. Moreover, this same expected-utility function also captures her attitude toward nearby gambles. In particular, if the global utility function is Fréchet differentiable then this is a *local expected-utility function* in the sense of Machina (1982, 1984). That is, the individual's preferences for small changes in risk—such as incremental adjustments to portfolio or insurance decisions—are well approximated using expected-utility analysis on $\phi \circ u_{t+1}$. While Machina's techniques relied on Fréchet differentiability of the global utility function, which is not generally required in this model, many of his insights still apply.[2]

1.3 Outline

The remainder of the paper is organized as follows. Section 2 illustrates the implications of the Optimal Anticipation model for attitudes toward risk by showing that the linear gain-loss function from Example 1 can be used to resolve both the equity premium puzzle and Rabin's paradox. To make the exposition as straightforward as possible, this section again restricts attention to the two-period version of the model.

In Section 3, I describe the infinite-horizon framework and define the Optimal Anticipation representation formally. Following the approach in Epstein and Zin (1989), the recursive utility representation is formulated by applying the certainty equivalent for static risks described above to continuation values in each period. After defining the representation, I also establish the existence of a value function and provide an axiomatic analysis.

[2] For example, Proposition 1 relates nondecreasing local utility functions ϕ to first-order stochastic dominance monotonicity of the overall representation. Also, the optimality conditions for the portfolio choice problems in Sections 2.1 and 6 are based on the local utility function.

A more detailed analysis of several parametric special cases is contained in Section 4. In Section 4.1, the quantile formula for the optimal anticipation for the linear gain-loss function from Example 1 is derived. In Section 4.2, I show that recursive expected utility in the sense of Epstein and Zin (1989) and Kreps and Porteus (1978) (see also Weil (1989, 1990)) can be expressed a special cases of the Optimal Anticipation representation.[3] Example 2 can be used to demonstrate one instance of this equivalence: Theorem 5 (in particular, Corollary 2) shows that for any distribution of future consumption c_{t+1},

$$u_t(c_t) + \beta \max_{\gamma \in \mathbb{R}} \mathbb{E}_t\Big[\gamma + \frac{1}{\theta} - \frac{1}{\theta}\exp\big(-\theta(u_{t+1}(c_{t+1}) - \gamma)\big)\Big] \tag{5a}$$

$$= u_t(c_t) - \beta\frac{1}{\theta}\log \mathbb{E}_t\Big[\exp\big(-\theta u_{t+1}(c_{t+1})\big)\Big]. \tag{5b}$$

The first expression is the exponential gain-loss function from the example, and the second expression is a special case of the Kreps-Porteus expected utility. The connection between these two models provides a useful alternative interpretation to the class of Kreps-Porteus preferences based on reference-dependent utility. Instead of treating the exponential and logarithmic transformations in Equation (5b) as fundamental parameters of the model, Equation (5a) suggests they could be a reduced-form expression for an underlying process of anticipation and loss aversion.[4]

Sections 5 and 6 develop an application to portfolio choice. To facilitate the analysis, I first find conditions on the parameterized Optimal Anticipation representation that generate homothetic preferences in Section 5. In particular, both of the examples of gain-loss functions described in the introduction are shown to lead to such preferences when combined with the appropriate CES utility function u for consumption in each period. Then, in Section 6, I apply the homothetic model to a portfolio choice problem. I find that asset returns satisfy Euler equations that slightly generalize the standard conditions obtained for the additively-separable expected-utility model. Despite their relative simplicity, these conditions have the potential to accommodate a variety of empirically observed pricing phenomenon that cannot be matched with the standard model, such as the combination of a large equity premium, a low and nonvolatile risk free rate, and a counter-cyclical risk premium. Expanding the analysis to obtain more complete qualitative predictions for asset prices and performing a calibration exercise to check the quantitative predictions of the model are the subject of ongoing research.

[3]Note that the Optimal Anticipation model nests Kreps-Porteus expected-utility, but the converse is clearly not true. For instance, any special case of the Optimal Anticipation representation that admits first-order risk aversion (such as Example 1) is necessarily incompatible with expected utility.

[4]Kreps and Porteus (1979) provided a related analysis of preferences induced by some action on the part of the individual, and they determined conditions under which such induced preferences could be represented using Kreps-Porteus utility. Machina (1984) and Ergin and Sarver (2011) extended this idea to broader classes of preferences.

I conclude in Section 7 with a more detailed discussion of related literature. In Section 7.1, I highlight some additional conceptual distinctions between the model of reference-point formation proposed in this paper and the other approaches mentioned earlier in the introduction. In Section 7.2, I describe the connection between the model in this paper and another important class of non-expected-utility preferences: betweenness preferences. In Section 7.3, I discuss several other models related to optimal anticipation, including models of anxiety about future uncertainty, distortion of beliefs about the future, and physical adjustment costs in consumption.

2 The Equity Premium Puzzle and Rabin's Paradox

In this section, I describe two limitations of expected-utility theory and how they can be overcome using the Optimal Anticipation model. I show in Sections 2.1 and 2.2 that both the equity premium puzzle and Rabin's paradox can be resolved using the linear gain-loss function from Example 1. As I discuss in these sections, the two puzzles are both manifestations of a single feature of expected utility—the manner in which it connects attitudes toward gambles of small or moderate size to attitudes for larger gambles—that the Optimal Anticipation representation is able to relax.

This model is certainly not the first to address these puzzles; however, its ability to address them in an intuitive and analytically tractable manner suggest that it may be a useful model for applied work. For simplicity, I will focus on the two-period version of the model throughout this section, but the results easily generalize to the infinite-horizon model that will be presented later.

2.1 The Equity Premium and Risk-Free Rate Puzzles

For the period 1889–1978, the average real return on US stocks and riskless securities were approximately 7% and 1%, respectively. Qualitatively, expected returns on stocks should be higher to compensate for their additional risk (standard deviation of 16% in this sample period); however, there is a quantitative question of precisely how large the difference in expected returns should be. Mehra and Prescott (1985) showed that the additively-separable expected-utility model fails on this quantitative dimension: matching the observed equity premium of 6% given the low historic volatility of consumption growth requires an implausibly high coefficient of relative risk aversion. This observation is referred to as the *equity premium puzzle* (see also Kocherlakota (1996) for a survey and Cochrane (2005, Chapter 21) for a recent textbook treatment).

A convenient way to formulate the equity premium puzzle is using Hansen and Jagannathan (1991) bounds. This approach will also make it transparent why the Optimal

Anticipation model is able to resolve the puzzle. Formally, suppose an individual has additively-separable expected-utility preferences with CES utility $u(c) = \frac{c^\rho}{\rho}$ in each period. This individual consumes c_t in period t, and her investment decisions result in stochastic consumption c_{t+1} in the subsequent period. Consider any asset and denote its stochastic returns by r_{t+1}. The first-order condition for her current level of investment in the asset to be optimal yields the following standard Euler equation:

$$1 = \beta \mathbb{E}_t\Big[\frac{u'(c_{t+1})}{u'(c_t)} r_{t+1}\Big] = \beta \mathbb{E}_t\Big[\Big(\frac{c_{t+1}}{c_t}\Big)^{\rho-1} r_{t+1}\Big]. \tag{6}$$

This pricing equation can be summarized as $\mathbb{E}_t[M_{t+1} r_{t+1}] = 1$, where $M_{t+1} = \beta(\frac{c_{t+1}}{c_t})^{\rho-1}$ is the *stochastic discount factor*. Let r^f_{t+1} denote the risk-free rate, which by the preceding pricing equation satisfies $r^f_{t+1} = \frac{1}{\mathbb{E}_t[M_{t+1}]}$. Then, the excess return returns $r_{t+1} - r^f_{t+1}$ for any asset must satisfy

$$0 = \mathbb{E}_t[M_{t+1}(r_{t+1} - r^f_{t+1})] = \text{cov}(M_{t+1}, r_{t+1} - r^f_{t+1}) + \mathbb{E}_t(M_{t+1})\mathbb{E}_t(r_{t+1} - r^f_{t+1}).$$

Rearranging terms yields the Hansen-Jagannathan bounds:

$$\frac{\mathbb{E}_t(r_{t+1} - r^f_{t+1})}{\sigma_t(r_{t+1})} = -\text{corr}_t(M_{t+1}, r_{t+1} - r^f_{t+1}) \frac{\sigma_t(M_{t+1})}{\mathbb{E}_t(M_{t+1})} \leq \frac{\sigma_t(M_{t+1})}{\mathbb{E}_t(M_{t+1})}. \tag{7}$$

The expression on the left-hand side of Equation (7) is the Sharpe ratio. The equity premium of 6% and standard deviation of 16% from the period 1889–1978 implies a Sharpe ratio of 0.375. Estimates using postwar US data put this number closer to 0.5 (see Campbell (1999)). These values place a lower bound on the volatility of the stochastic discount factor on the right-hand side of Equation (7). These bounds can be related to consumption and risk aversion by observing that for $M_{t+1} = \beta(\frac{c_{t+1}}{c_t})^{\rho-1}$ the term $\frac{\sigma_t(M_{t+1})}{\mathbb{E}_t(M_{t+1})}$ is approximately equal to the product of the coefficient of relative risk aversion $1-\rho$ and standard deviation of log consumption growth.[5] In the original sample period used by Mehra and Prescott, the standard deviation of consumption growth was roughly 3.6%, implying a lower bound for risk aversion of $1 - \rho \geq \frac{0.375}{0.036} = 10.4$. Using postwar data, the volatility of consumption growth is closer to 1%, which together with the larger Sharpe ratio of 0.5, gives a lower bound of 50.

Mehra and Prescott (1985), Lucas (2003), and others have argued that values of the CRAA greater than 10—and certainly greater than 50—are absurdly high and incon-

[5] To illustrate, consider the case of log-normal consumption growth. That is, $\log(\frac{c_{t+1}}{c_t}) = g + \xi_{t+1}$, where $\xi_{t+1} \sim N(0, \sigma_c^2)$. Then, tedious but standard calculations show that $\frac{\sigma_t(M_{t+1})}{\mathbb{E}_t(M_{t+1})} = \sqrt{\exp((1-\rho)^2\sigma_c^2) - 1}$. Taking the first-order approximation of the exponential function simplifies this further to $(1-\rho)\sigma_c$.

sistent with evidence from large-scale risks (e.g., wage premiums for occupations with high earnings risk). Moreover, Weil (1989) observed that even if such a large CRRA were permitted, it would result in a risk-free rate well above 1% for any discount factor $\beta < 1$.[6] This implies that it is impossible to simultaneously match the equity premium and the risk-free rate with the additively-separable model for any CRRA, an observation referred to as the *risk-free rate puzzle* (see also Kocherlakota (1996) and Campbell (1999)).

Turn now to the Optimal Anticipation representation. Consider the parameterized version of the model from Equation (1), and let γ^* denote the optimal anticipation when faced with the random variable c_{t+1}. Assuming that $\phi(\cdot|\gamma^*)$ is differentiable at all but finitely many points and c_{t+1} is continuously distributed, the returns r_{t+1} to any asset must satisfy the following modified Euler equation:[7]

$$1 = \beta \mathbb{E}_t \Big[\phi_1 \big(c_{t+1}^{\rho}/\rho \big| \gamma^* \big) \Big(\frac{c_{t+1}}{c_t} \Big)^{\rho - 1} r_{t+1} \Big],$$

where $\phi_1(\cdot|\cdot)$ denotes the partial derivative of $\phi(\cdot|\cdot)$ with respect to its first argument. For concreteness, focus on the case where $\phi(\cdot|\gamma)$ is the linear gain-loss function from Equation (2). This gain-loss function is differentiable everywhere except at γ (with slope of λ^l below γ and λ^g above γ), and the pricing equation simplifies to[8]

$$1 = \beta \mathbb{E}_t \Big[\Big(\lambda^g + (\lambda^l - \lambda^g) \mathbf{1}_{[c_{t+1}^{\rho}/\rho \leq \gamma^*]} \Big) \Big(\frac{c_{t+1}}{c_t} \Big)^{\rho - 1} r_{t+1} \Big].$$

If $\lambda^l = \lambda^g = 1$, this expression simplifies to the standard condition from Equation (6). To illustrate to new aspects of asset pricing associated with this model of reference-dependent utility, I instead consider the case of $\lambda^l > 1 > \lambda^g$ and $\rho = 1$ (so $u(c) = c$). Then, the Euler equation for asset returns becomes

$$1 = \beta \mathbb{E}_t \Big[\Big(\lambda^g + (\lambda^l - \lambda^g) \mathbf{1}_{[c_{t+1} \leq \gamma^*]} \Big) r_{t+1} \Big].$$

The model's ability to resolve the equity premium puzzle is easily demonstrated by examining the Hansen-Jagannathan bounds in Equation (7) for the stochastic discount factor $M_{t+1} = \beta(\lambda^g + (\lambda^l - \lambda^g)\mathbf{1}_{[c_{t+1} \leq \gamma^*]})$. In Theorem 4, I will show that for any continuous

[6]Intuitively, in the standard model, increasing risk aversion also lowers the elasticity of substitution between periods. Since the average consumption growth rate is between 1% and 2%, individuals would need to be compensated for this nonsmooth consumption by a large risk-free rate.

[7]The derivation of this condition is analogous to that of Equation (6) in the standard model. Suppose the individual considers changing her investment in the asset by α dollars. Her resulting utility is then $u(c_t - \alpha) + \beta \max_{\gamma \in \mathbb{R}} \mathbb{E}_t[\phi(u(c_{t+1} + \alpha r_{t+1})|\gamma)]$. If no additional purchase or short-sale of the asset can improve her utility, then $\alpha = 0$ must be optimal for $\gamma = \gamma^*$. Taking the first-order condition with respect to α yields the above Euler equation.

[8]Here, $\mathbf{1}_{[c_{t+1}^{\rho}/\rho \leq \gamma^*]}$ denotes the indicator function for the event $[c_{t+1}^{\rho}/\rho \leq \gamma^*]$.

distribution, the optimal γ^* for the linear gain-loss function satisfies $\Pr(c_{t+1} \leq \gamma^*) = \frac{1-\lambda^g}{\lambda^l-\lambda^g}$. Therefore,

$$\mathbb{E}_t(M_{t+1}) = \beta\big[\lambda^g + (\lambda^l - \lambda^g)\Pr(c_{t+1} \leq \gamma^*)\big] = \beta,$$

$$\begin{aligned}\mathbb{E}_t(M_{t+1}^2) &= \beta^2\big[(\lambda^g)^2 + 2\lambda^g(\lambda^l - \lambda^g)\Pr(c_{t+1} \leq \gamma^*) + (\lambda^l - \lambda^g)^2\Pr(c_{t+1} \leq \gamma^*)\big] \\ &= \beta^2(\lambda^l + \lambda^g - \lambda^l\lambda^g),\end{aligned}$$

and hence

$$\frac{\sigma_t(M_{t+1})}{\mathbb{E}_t(M_{t+1})} = \frac{\sqrt{\beta^2(\lambda^l + \lambda^g - \lambda^l\lambda^g) - \beta^2}}{\beta} = \sqrt{(\lambda^l - 1)(1 - \lambda^g)}.$$

The parameters λ^l and λ^g can easily be chosen to match the data. For example, if λ^l and λ^g are symmetric about 1, so that $\lambda^l = 1 + \kappa$ and $\lambda^g = 1 - \kappa$, then this expression simplifies to κ. By choosing $\kappa \geq 0.5$, the model delivers Hansen-Jagannathan bounds large enough to be consistent with the Sharpe ratios calculated above.[9]

Notice that the linear gain-loss utility can also resolve the risk-free rate puzzle. The calculations above show that $r_{t+1}^f = \frac{1}{\mathbb{E}_t(M_{t+1})} = \frac{1}{\beta}$ is independent of both the parameters λ^l and λ^g and the distribution of consumption growth. Thus, choosing λ^l and λ^g to give the correct equity premium does not force up the risk-free rate, and the desired rate of 1% is obtained for $\beta = \frac{1}{1.01}$.

It is natural to question whether the Optimal Anticipation model "solves" the equity premium puzzle only because it introduces another free parameter that effectively increases risk aversion. To address this point, it is useful to recall the reasons for criticizing high values of the CRRA within the expected-utility model. The first issue with increasing risk-aversion is the aforementioned risk-free rate puzzle. However, models that separate risk aversion from elasticity of intertemporal substitution (such as the Epstein-Zin-Kreps-Porteus expected-utility model discussed in Sections 4.2 and 5.2.2) are able to match both the equity premium and risk-free rate by increasing the CRRA while at the same time maintaining a high elasticity of intertemporal substitution (see Kocherlakota (1996)). Nonetheless, fitting the data using expected utility with high CRRA still evokes the concerns expressed by Lucas (2003): in environments where the risk to consumption

[9] Note that the Hansen-Jagannathan bounds only provide a lower bound on the volatility of the stochastic discount factor, and a larger value of κ may in fact be needed. If M_{t+1} and $r_{t+1} - r_{t+1}^f$ are not perfectly negatively correlated, then the first equality in Equation (7) requires a larger value of $\sigma_t(M_{t+1})/\mathbb{E}_t(M_{t+1})$. For the expected utility model, it has been shown that the low correlation between consumption growth and excess returns increases the required volatility of the stochastic discount factor by a factor of roughly 5. Thus, in addition to the equity premium puzzle, there is a *correlation puzzle*. However, for the Optimal Anticipation model, the same correlation between consumption growth and excess returns may translate into different correlation between M_{t+1} and excess returns, so addition study is needed to determine the extent to which correlation affects the required value of κ.

is of a larger scale, the implied aversion to risk from a CRRA greater than 50 is completely inconsistent with both introspection and empirical evidence.

While it is the case that taking $\lambda^l > 1 > \lambda^g$ increases risk aversion, it does so in a manner different than simply increasing the CRRA within an expected-utility model. One illustration of the distinction can be found in the expressions for the Hansen-Jagannathan bounds for the two models. For standard expected utility, recall that the bounds are (roughly) proportional to the standard deviation of consumption growth. Intuitively, expected utility becomes approximately linear as variation in consumption becomes small, so either large variation in consumption or large risk aversion is required to generate a large premium for bearing risk. On the other hand, the kink in the linear gain-loss function implies that the actual volatility of consumption growth does not influence the bounds—the value of the stochastic discount factor is λ^l (λ^g, respectively) for any realized consumption below (above, respectively) γ^*, no matter how small the deviation from γ^*. The fact that the Hansen-Jagannathan bounds for the linear gain-loss function do not change as the size of consumption risk increases implies it can generate the aversion to moderate risk in aggregate consumption that is needed to explain the observed equity premium without imposing an excessive aversion to larger idiosyncratic risks to consumption. In the next section, I provide another simple illustration of this feature of the model by showing that it can be used to explain Rabin's paradox.

2.2 Rabin's Paradox

Rabin (2000) performed a calibration exercise for expected utility and showed that attitudes toward gambles on the order of \$100 can have striking implications for attitudes toward risks on the order of \$1,000 to \$10,000. He found that plausible rejection of certain small gambles implied absurdly high aversion to larger gambles.

Suppose an individual has a concave expected-utility function, and at any wealth level below \$300,000, she would reject a 50-50 lose \$100/gain \$110 gamble. Rabin (2000, Table 2) showed this implies that at a wealth level of \$290,000, she would also reject any of the following 50-50 gambles:

lose	gain
\$1,000	\$718,190
\$2,000	\$12,210,880
\$4,000	\$60,528,930
\$10,000	\$1,300,000,000

While the rejection of a lose \$100/gain \$110 gamble seems quite plausible, the rejection of the gambles listed in the table is completely unbelievable; hence the paradox for expected utility.

The intuition for Rabin's result is the following. Suppose an expected-utility maximizing individual would not accept the -\$100/+\$110 gamble at a wealth level of \$290,000. It could be that that her utility function is very concave around this current wealth level (perhaps even kinked), but approximately linear elsewhere. If this is the case, then the individual would reject any gambles proportional to -\$100/+\$110, but could potentially accept a gamble with a proportionally larger gain, such as -\$4,000/+\$5,000. The key to Rabin's conclusion is the stronger assumption that the individual would reject the -\$100/+\$110 bet at *every wealth level* below \$300,000. This implies that not only is her marginal utility dropping rapidly between \$289,900 and \$290,110, but also between $w - \$100$ and $w + \$110$ for any $w \leq \$300,000$. The cumulative effect of these moderate declines in marginal utility over small intervals is an extreme drop in marginal utility over larger intervals, leading to the rejection of the gambles listed above.

To see how Rabin's paradox can be resolved using the Optimal Anticipation model, suppose the individual has linear gain-loss utility as in Equation (2) and take $u_{t+1}(c_{t+1}) = c_{t+1}$. Recall that any choice of anticipation γ^* that satisfies $\Pr(c_{t+1} \leq \gamma^*) = \frac{1-\lambda^g}{\lambda^l - \lambda^g}$ is optimal (Theorem 4). For simplicity, suppose $\lambda^l = 1 + \kappa$ and $\lambda^g = 1 - \kappa$, which implies $\frac{1-\lambda^g}{\lambda^l-\lambda^g} = \frac{1}{2}$. Then, when faced with a 50-50 lose $\$l$/gain $\$g$ gamble at wealth level w, it is optimal to take $\gamma^* = w$,[10] yielding an overall utility of

$$\frac{1}{2}\phi(w - l|w) + \frac{1}{2}\phi(w + g|w) = w - \frac{1}{2}\lambda^l l + \frac{1}{2}\lambda^g g.$$

The individual will therefore accept this gamble if and only if

$$w - \frac{1}{2}\lambda^l l + \frac{1}{2}\lambda^g g \geq w = \max_{\gamma \in \mathbb{R}} \phi(w|\gamma),$$

which simplifies to $\frac{g}{l} \geq \frac{\lambda^l}{\lambda^g}$. For example, for $\lambda^l = 1.05$ and $\lambda^g = .95$, the individual would accept any 50-50 gamble with a gain/loss ratio exceeding $\frac{\lambda^l}{\lambda^g} \approx 1.105$. In this case, she would reject a 50-50 lose \$100/gain \$110 gamble at any wealth level, but would *accept* each of the following 50-50 gambles:

[10] In fact, any γ^* between $w - l$ and $w + g$ gives the same utility, but the calculations are simplest by taking $\gamma^* = w$. Note also that this reference point is consistent with Markowitz (1952)'s suggestion of measuring gains and losses relative to recent wealth. Of course, Markowitz's approach can only be applied to deterministic initial wealth and overlaps with this model only when evaluating 50-50 gambles. For more complex risks, this paper diverges in favor of a more sophisticated approach to reference-point formation.

lose	gain
\$1,000	\$1,200
\$2,000	\$2,400
\$4,000	\$4,800
\$10,000	\$12,000

Thus, the absurd attitudes toward large gambles generated by expected utility are not replicated by the Optimal Anticipation model. To understand how this model solves Rabin's paradox, recall the intuition for his result from above. Knowing that an expected-utility maximizer rejects a -\$100/+\$110 gamble at one particular wealth level w has very weak implications for attitudes toward larger gambles—there could simply be a kink in her utility function at or near w. Knowing that this gamble will be rejected at every wealth level is what implies marginal utility is quickly diminishing everywhere, generating extreme risk aversion for large gambles. However, for the Optimal Anticipation representation with linear gain-loss utility, the kink in the individual's utility function moves with wealth. Hence, to reject the small gamble, the local utility function $\phi(\cdot|\gamma)$ need not be very concave everywhere, only at consumption values near γ.

3 Infinite-Horizon Model

3.1 Framework and Representation

For any topological space X, let $\triangle(X)$ denote the set of all (countably-additive) Borel probability measures on X, endowed with the topology of weak convergence (or weak* topology). This topology is metrizable if X is metrizable. For any $x \in X$, let δ_x denote the Dirac probability measure concentrated at x.

The setting for the axiomatic analysis is the space of infinite-horizon consumption problems. Let C be a compact and connected metrizable space, denoting the consumption space for each period.[11] A 1-period consumption problem is simply a choice from C. Similarly, the space of 2-period consumption problems is $C \times \triangle(C)$, the space of 3-period consumption problems is $C \times \triangle(C \times \triangle(C))$, and so on. The following lemma shows that there is a well-defined space D of infinite-horizon consumption problems. It follows from standard techniques used, for example, by Mertens and Zamir (1985) and Brandenburger

[11] While the axiomatic analysis will be restricted to compact spaces, the applications of the model in Section 6 will allow for the unbounded consumption space $\mathbb{R}_+$. It is possible to generalize the axiomatic analysis to infinite-horizon consumption problems that use a non-compact consumption space by imposing bounded consumption growth rates (see Epstein and Zin (1989) for a formal description of such a framework). However, this would result in additional technical complications and add little to the behavioral insights of the current analysis.

and Dekel (1993) in the context of hierarchies of beliefs and Epstein and Zin (1989) and Gul and Pesendorfer (2004) in recursive consumption problems.

Lemma 1 *If C is a compact and connected metrizable space, then there exists a compact and connected metrizable space D such that D is homeomorphic to $C \times \triangle(D)$.*

Proof: Mertens and Zamir (1985) showed that if C is compact, there exists a compact space E such that E is homeomorphic to $\triangle(C \times E)$. The metrizability of E follows from the same arguments used in Brandenburger and Dekel (1993). Let $D = C \times E$. Then, D is homeomorphic to $C \times \triangle(C \times E) = C \times \triangle(D)$. For any topological space D, the set $\triangle(D)$ is connected. Since C is also connected, it follows that the product $C \times \triangle(D)$ is connected (Theorem 23.6 in Munkres (2000)). Thus, D is connected. ■

Since D is homeomorphic to $C \times \triangle(D)$, elements of D will typically be denoted by (c, m), where $c \in C$ and $m \in \triangle(D)$. The primitive of the axiomatic model is a binary relation $\succsim$ on the set of infinite-horizon consumption problems D.

Definition 1 An *Optimal Anticipation* representation is a tuple (V, u, Φ, β) consisting of a continuous function $V : D \to \mathbb{R}$ that represents $\succsim$, a continuous and nonconstant function $u : C \to \mathbb{R}$, a collection Φ of continuous and nondecreasing functions $\phi : V(D) \to \mathbb{R}$,[12] and a scalar $\beta \in (0, 1)$ such that

$$V(c, m) = u(c) + \beta \sup_{\phi \in \Phi} \int_D \phi\big(V(\bar{c}, \bar{m})\big)\, dm(\bar{c}, \bar{m}), \tag{8}$$

for all $(c, m) \in D$, and

$$\sup_{\phi \in \Phi} \phi(v) = v, \quad \forall v \in V(D). \tag{9}$$

The interpretation of the Optimal Anticipation representation is the same as the intuition provided in the introduction for the 2-period model—the choice of ϕ corresponds to anticipation or psychological preparation about future utility. In addition, the parameterized examples described in the introduction in which the individual anticipates an actual utility value γ can be applied to the infinite-horizon model (see Section 4). The only distinction is that in the 2-period model the individual anticipated her utility in the subsequent period in isolation, whereas in the Optimal Anticipation representation the individual anticipates the entire continuation value $V(\bar{c}, \bar{m})$ for the subsequent period onward. This approach allows the representation to be formulated recursively, making the model more parsimonious and tractable.

[12] The set $V(D)$ denotes the range of V.

To better understand the connection between the Optimal Anticipation representation and related recursive models, it is useful to reformulate the representation in slightly different terms. Fix any Optimal Anticipation representation (V, u, Φ, β), and let $a = \min V$ and $b = \max V$. Define a function $W : \triangle([a, b]) \to \mathbb{R}$ by

$$W(\mu) = \sup_{\phi \in \Phi} \int_a^b \phi(v)\, d\mu(v). \tag{10}$$

Let $m \circ V^{-1}$ denote the distribution (on $[a, b]$) of continuation values induced by the measure $m \in \triangle(D)$.[13] Then, by the change of variables formula,

$$W(m \circ V^{-1}) = \sup_{\phi \in \Phi} \int_a^b \phi(v)\, d(m \circ V^{-1})(v) = \sup_{\phi \in \Phi} \int_D \phi(V(\bar{c}, \bar{m}))\, dm(\bar{c}, \bar{m}), \tag{11}$$

which implies Equation (8) can be expressed as

$$V(c, m) = u(c) + \beta W(m \circ V^{-1}). \tag{12}$$

Also, Equation (9) implies

$$W(\delta_v) = v, \qquad \forall v \in [a, b]. \tag{13}$$

In other words, W maps any deterministic continuation value back to that same value, and hence, W is a certainty equivalent function for lotteries $\mu = m \circ V^{-1}$ over continuation values. In addition, since it is the supremum of a collection of expected-utility functions by Equation (10), the certainty equivalent function W is convex; in fact, convexity turns out to be the characterizing property of the Optimal Anticipation certainty equivalent. The representation described in Equations (12) and (13) corresponds to the general representation from Epstein and Zin (1989) (but without the assumption of CES utility). The Optimal Anticipation representation is the special case of this general representation where the certainty equivalent function W takes the particular form described in Equation (10).

In contrast, existing applications of the Epstein and Zin (1989) model of recursive utility have used alternative specifications of the certainty equivalent function. The most prevalent special case is that of expected utility in the sense of Kreps and Porteus (1978), and I discuss in Sections 4.2 and 5.2.2 how this special case of Epstein-Zin preferences can be formulated as a special case of the Optimal Anticipation representation. Examples of non-expected-utility certainty equivalents in the literature include the following: various specializations of the betweenness preferences of Chew (1983) and Dekel (1986) were used

[13] This is standard notation for the distribution of a random variable. Intuitively, the probability that m yields a continuation value in a set $E \subset [a, b]$ is the probability that $V(\bar{c}, \bar{m}) \in E$, which is $m \circ V^{-1}(E)$.

by Epstein and Zin (1991) and Epstein and Zin (2001); a special case of rank-dependent utility was used by Epstein and Zin (1990); the disappointment aversion model of Gul (1991) was used by Bekaert, Hodrick, and Marshall (1997); and a generalization of the disappointment aversion model (that still falls within the betweenness class of preferences) was used by Routledge and Zin (2010).

3.2 Existence of a Value Function

Note that the value function V is included explicitly in the definition of the Optimal Anticipation representation. However, it may be desirable to obtain such a value function from the other parameters (u, Φ, β) of the representation. Building on techniques from Epstein and Zin (1989), the following result shows that this is possible.

Theorem 1 *Suppose $\beta \in (0,1)$ and $u : C \to \mathbb{R}$ is a continuous and nonconstant function. Let $[a,b] = u(C)$, where $-\infty < a < b < +\infty$,*[14] *and let $I = [\frac{a}{1-\beta}, \frac{b}{1-\beta}]$. Let Φ be any collection of continuous and nondecreasing functions $\phi : I \to \mathbb{R}$ that satisfies $\sup_{\phi \in \Phi} \phi(v) = v$ for all $v \in I$. Then, there exists a bounded and lower semicontinuous function $V : D \to I$ that satisfies Equation (8).*

For the Optimal Anticipation representation to be well-defined, the functions $\phi \in \Phi$ must be defined everywhere on the set $V(D)$. However, if V is not known and needs to be determined from the other parameters of the representation (u, Φ, β), then the relevant domain of the functions $\phi \in \Phi$ is not known *a priori*. Nonetheless, Theorem 1 shows that the range of V can be determined from the range of u, and hence it suffices to consider functions ϕ defined on this interval I. In particular, if $u \geq 0$ ($u \leq 0$, respectively) then it suffices to define ϕ on $\mathbb{R}_+$ ($\mathbb{R}_-$, respectively).

There are two noticeable gaps in Theorem 1. First, it does not ensure the uniqueness of the function V. Second, it does not ensure that the function V is continuous, only lower semicontinuous. Since resolving these issues is not central to the analysis in this paper, obtaining a stronger version of this result is left as an interesting open question for future research. However, it is worth noting that in the case of homothetic preferences, it is often possible to ensure both uniqueness and continuity of the value function (see Theorem 7 as well as related results in Marinacci and Montrucchio (2010)).

3.3 Axioms and Representation Result

In this section, I provide an axiomatic foundation for the Optimal Anticipation representation. The first three axioms are entirely standard.

[14] Since C is compact and connected and u is continuous, $u(C)$ is a closed and bounded interval in $\mathbb{R}$.

Axiom 1 (Weak Order) *The relation $\succsim$ is complete and transitive.*

Axiom 2 (Nontriviality) *There exist $c, c' \in C$ and $m \in \triangle(D)$ such that $(c, m) \succ (c', m)$.*

Axiom 3 (Continuity) *The sets $\{(c, m) \in D : (c, m) \succ (c', m')\}$ and $\{(c, m) \in D : (c, m) \prec (c', m')\}$ are open for all $(c', m') \in D$.*

The following stationarity axiom is also standard for recursive utility models.

Axiom 4 (Stationarity) *For any $c, \bar{c}, \bar{c}' \in C$ and $\bar{m}, \bar{m}' \in \triangle(D)$,*

$$(c, \delta_{(\bar{c},\bar{m})}) \succsim (c, \delta_{(\bar{c}',\bar{m}')}) \iff (\bar{c}, \bar{m}) \succsim (\bar{c}', \bar{m}')$$

The following axiom applies the separability condition of Debreu (1960) to all triples of consumption today, consumption tomorrow, and the lottery following tomorrow's consumption.

Axiom 5 (Separability) *For any $c, c', \bar{c}, \bar{c}' \in C$ and $\bar{m}, \bar{m}' \in \triangle(D)$,*

1. *$(c, \delta_{(\bar{c},\bar{m})}) \succsim (c', \delta_{(\bar{c}',\bar{m})})$ if and only if $(c, \delta_{(\bar{c},\bar{m}')}) \succsim (c', \delta_{(\bar{c}',\bar{m}')})$.*
2. *$(c, \delta_{(\bar{c},\bar{m})}) \succsim (c', \delta_{(\bar{c},\bar{m}')})$ if and only if $(c, \delta_{(\bar{c}',\bar{m})}) \succsim (c', \delta_{(\bar{c}',\bar{m}')})$.*

Condition 1 in Axiom 5 says that the comparison of c today and $\bar{c}$ tomorrow versus c' today and $\bar{c}'$ tomorrow is the same regardless of the lottery ($\bar{m}$ or $\bar{m}'$) following tomorrow's consumption. Likewise, condition 2 says that comparison of c today and lottery $\bar{m}$ following tomorrow versus c' today and $\bar{m}'$ following tomorrow is the same for any consumption tomorrow ($\bar{c}$ or $\bar{c}'$). Note that Axiom 5 only applies to consumption problems in which the one-step-ahead continuation problem is deterministic. Intuitively, in the case of deterministic consumption problems, the optimal anticipation preferences reduce to standard additively separable preferences.

The next axiom ensures that preferences respect the first-order stochastic dominance order on $\triangle(D)$. Recall that in the case of monetary gambles, FOSD roughly corresponds to increasing the probability of better monetary outcomes. The same is true is this setting, with $(\bar{c}, \bar{m})$ being a "better" continuation path than $(\bar{c}', \bar{m}')$ if and only if $(c, \delta_{(\bar{c},\bar{m})}) \succsim (c, \delta_{(\bar{c}',\bar{m}')})$ for some $c \in C$.

Axiom 6 (FOSD) *For any $c \in C$ and $m, m' \in \triangle(D)$, if*

$$m\Big(\Big\{(\bar{c},\bar{m}) : (c,\delta_{(\bar{c},\bar{m})}) \succsim (c,\delta_{(\bar{c}',\bar{m}')})\Big\}\Big) \geq m'\Big(\Big\{(\bar{c},\bar{m}) : (c,\delta_{(\bar{c},\bar{m})}) \succsim (c,\delta_{(\bar{c}',\bar{m}')})\Big\}\Big)$$

for all $(\bar{c}',\bar{m}') \in D$, then $(c,m) \succsim (c,m')$.[15]

The final axiom is key to the optimal anticipation interpretation of the preferences.

Axiom 7 (Convexity) *For any $c, c' \in C$ and $m, m' \in \triangle(D)$,*

$$(c, \tfrac{1}{2}m + \tfrac{1}{2}m') \succsim (c', m) \implies (c, m') \succsim (c', \tfrac{1}{2}m + \tfrac{1}{2}m')$$

To interpret Axiom 7 in the simplest case, let $m = \delta_{(\bar{c},\bar{m})}$ and $m' = \delta_{(\bar{c}',\bar{m}')}$ for $(\bar{c}',\bar{m}') \succ (\bar{c},\bar{m})$. Suppose also that there exist $c, c' \in C$ such that $(c, \frac{1}{2}m + \frac{1}{2}m') \sim (c', m)$; that is, the utility gain from switching from m to $\frac{1}{2}m + \frac{1}{2}m'$ is the same as the utility gain from switching from c to c'. Note that changing from m to $\frac{1}{2}m + \frac{1}{2}m'$ increases the probability of the better outcome by $\frac{1}{2}$. However, when faced with $\frac{1}{2}m + \frac{1}{2}m'$, the individual still may not want to anticipate the better outcome $(\bar{c}',\bar{m}')$ since it is not certain. In contrast, moving from $\frac{1}{2}m + \frac{1}{2}m'$ to m' again increases the probability of the better outcome by $\frac{1}{2}$, but also makes it certain. The individual can therefore better target her anticipation, and hence should have a (weakly) bigger utility gain from this change. Axiom 7 formalizes this intuition by requiring that $(c, m') \succsim (c', \frac{1}{2}m + \frac{1}{2}m')$.

It is worth pointing out that since the Optimal Anticipation representation is convex in m, preferences over continuation paths will necessarily satisfy quasiconvexity:

$$(c, m) \succsim (c, m') \implies (c, m) \succsim (c, \alpha m + (1-\alpha)m').$$

However, it is well-known that quasiconvexity of a preference is not sufficient to guarantee it admits a convex utility representation. Therefore, Axiom 7 instead leverages the additive separability afforded by Axiom 5 to obtain convexity.

The following result provides a behavioral characterization of the Optimal Anticipation representation.

Theorem 2 *The relation $\succsim$ satisfies Axioms 1–7 if and only if it has an Optimal Anticipation representation (V, u, Φ, β).*

The next result describes the uniqueness properties of the representation.

[15] Implicit in this axiom is the assumption that the set $\{(\bar{c},\bar{m}) : (c,\delta_{(\bar{c},\bar{m})}) \succsim (c,\delta_{(\bar{c}',\bar{m}')})\}$ is Borel measurable for each $(\bar{c}',\bar{m}') \in D$. However, if the continuity axiom is imposed, then each of these sets is closed and hence measurable.

Theorem 3 *Two Optimal Anticipation representations* $(V_1, u_1, \Phi_1, \beta_1)$ *and* $(V_2, u_2, \Phi_2, \beta_2)$ *represent the same preference if and only if* $\beta_1 = \beta_2$ *and there exist scalars* $\alpha > 0$ *and* $\lambda \in \mathbb{R}$ *such that* $V_2 = \alpha V_1 + \lambda$*,* $u_2 = \alpha u_1 + \lambda(1 - \beta_1)$*, and*

$$\sup_{\phi_2 \in \Phi_2} \int \phi_2\big(\alpha v + \lambda\big)\, d\mu(v) = \alpha \sup_{\phi_1 \in \Phi_1} \int \phi_1(v)\, d\mu(v) + \lambda, \qquad \forall \mu \in \triangle(V_1(D)).$$

3.4 Proof Sketch and Discussion

The proof of Theorem 2 is contained in Appendix B.2. In this section, I outline the key steps in the construction of the representation.

Axioms 1 and 3 imply that $\succsim$ has a continuous representation $V : D \to \mathbb{R}$. Moreover, using standard techniques together with results from Debreu (1960) on additively separable utility, it can be shown that Axioms 1–6 imply that this representation V can be chosen so as to satisfy Equation (12) for some scalar $\beta \in (0,1)$, continuous and nonconstant function $u : C \to \mathbb{R}$, and continuous function $W : \triangle([a,b]) \to \mathbb{R}$ (where $a = \min V$ and $b = \max V$) that satisfies Equation (13).

As noted above, the Optimal Anticipation representation corresponds to the particular specification of the representation in Equations (12) and (13) in which the certainty equivalent function W is given by Equation (10). To obtain this particular form, Axiom 7 is first used to establish the convexity of W. Then, standard duality results can be applied to show that W is the supremum of some collection of affine functions. Since any affine function on $\triangle([a,b])$ can be given an expected-utility representation, this implies there exists a collection Φ of continuous functions $\phi : [a,b] \to \mathbb{R}$ such that W is given by Equation (10):

$$W(\mu) = \sup_{\phi \in \Phi} \int_a^b \phi(v)\, d\mu(v).$$

Then, the change of variables arguments from Equation (11) can be applied:

$$\begin{aligned} V(c,m) &= u(c) + \beta W(m \circ V^{-1}) \\ &= u(c) + \beta \sup_{\phi \in \Phi} \int_a^b \phi(v)\, d(m \circ V^{-1})(v) \\ &= u(c) + \beta \sup_{\phi \in \Phi} \int_D \phi(V(\bar{c}, \bar{m}))\, dm(\bar{c}, \bar{m}). \end{aligned}$$

Moreover, since $W(\delta_v) = v$, we have $\sup_{\phi \in \Phi} \phi(v) = v$ for all $v \in [a,b]$.

The only remaining step is to show there exists a collection Φ satisfying Equation (10) such that each $\phi \in \Phi$ is a nondecreasing function. It can be shown that Axiom 6 implies that W is monotone with respect to FOSD. Then, the following result completes the proof.

Proposition 1 *Suppose $W : \triangle([a,b]) \to \mathbb{R}$ for $-\infty < a < b < +\infty$, and suppose W is lower semicontinuous in the topology of weak convergence and convex. Then, the following are equivalent:*

1. *W is monotone with respect to FOSD: For $\mu, \eta \in \triangle([a,b])$, if $\mu([v,b]) \geq \eta([v,b])$ for all $v \in [a,b]$, then $W(\mu) \geq W(\eta)$.*

2. *There exists a collection Φ of continuous and nondecreasing functions $\phi : [a,b] \to \mathbb{R}$ such that*
$$W(\mu) = \sup_{\phi \in \Phi} \int \phi(v)\, d\mu(v).$$

Proposition 1 is interesting in its own right, as it is a variation of one of the local expected-utility results of Machina (1982). Machina's approach was to assume Fréchet differentiability of the function W and relate the global properties of W to the local properties of its derivative. In contrast, Proposition 1 assumes W is convex and lower semicontinuous and relates a property of W (FOSD monotonicity) to the properties of its supporting linear (expected-utility) functions. Local expected-utility results for convex functions have also been obtained elsewhere, but under the assumption of differentiability or else stronger forms of continuity: For example, Machina (1984) considered convex and Fréchet differentiable functions and therefore was able to apply many results from his prior work (Machina (1982)). Chatterjee and Krishna (2010) relaxed the assumption of differentiability and obtained local expected-utility results for concave and Lipschitz continuous functions.

4 Parametric Special Cases

As in the leading examples from the introduction, it is often useful to work with the special case of the general Optimal Anticipation model in which the collection Φ is parameterized by some real number γ. Formally, let the index set be some interval I of real numbers, and suppose for each $\gamma \in I$ there is a corresponding function $\phi(\cdot|\gamma)$. Then, the set of gain-loss functions takes the form $\Phi = \{\phi(\cdot|\gamma) : \gamma \in I\}$, and Equation (8) simplifies to

$$V(c,m) = u(c) + \beta \sup_{\gamma \in I} \int_D \phi\big(V(\bar{c},\bar{m})\big|\gamma\big)\, dm(\bar{c},\bar{m}).$$

To simplify notation below, this representation will be expressed more compactly using expected values:

$$V(c,m) = u(c) + \beta \sup_{\gamma \in I} \mathbb{E}_m\big[\phi(V|\gamma)\big]. \tag{14}$$

Equation (9) in the definition of the Optimal Anticipation representation also requires that $\sup_{\gamma \in I} \phi(v|\gamma) = v$ for all $v \in V(D)$. To ensure this condition is satisfied, all of the examples below normalize the parameterization to satisfy $I = V(D)$ and $\phi(v|\gamma) \leq v$ for all $v \in I$, with equality at $v = \gamma$. In other words, the set of continuation utility values is itself the index set. Anticipating the correct continuation value ($\gamma = v$) returns that same value, and anticipating incorrectly ($\gamma \neq v$) yields potentially lower values. Also, note that by Theorem 1, even if the value function V is not known a priori, its range can be inferred from u and β, and it suffices to take $I = [\frac{a}{1-\beta}, \frac{b}{1-\beta}]$, where $[a, b] = u(C)$.

The follow sections consider some important examples of parameterized representations. In Section 4.1, I analyze the linear gain-loss utility introduced in Example 1 in greater detail. In Section 4.2, I show that Epstein-Zin-Kreps-Porteus expected utility can be formulated as a special case of the parameterized Optimal Anticipation model.

4.1 Linear Gain-Loss Utility

In this section, I apply the linear gain-loss function from Example 1 to the recursive Optimal Anticipation representation, and I show that there is a maximizing anticipation level γ given by a quantile-based formula. Suppose the value function V is given by Equation (14) and the gain-loss function is as in Equation (2):

$$\phi(v|\gamma) = \gamma + \begin{cases} \lambda^l(v - \gamma) & \text{if } v < \gamma \\ \lambda^g(v - \gamma) & \text{if } v \geq \gamma, \end{cases} \tag{15}$$

where $\lambda^l \geq 1 \geq \lambda^g \geq 0$. The following result characterizes the optimal γ for any distribution over continuation paths.

Theorem 4 *Suppose I is a closed and bounded interval in $\mathbb{R}$, and for each $\gamma \in I$ define $\phi(\cdot|\gamma) : I \to \mathbb{R}$ by Equation (15), where $\lambda^l \geq 1 \geq \lambda^g \geq 0$. Fix any measurable function $V : D \to I$ and $m \in \triangle(D)$. Then, γ is a maximizer of $\mathbb{E}_m[\phi(V|\gamma)]$ if and only if it satisfies*

$$m(V < \gamma) \leq \frac{1 - \lambda^g}{\lambda^l - \lambda^g} \leq m(V \leq \gamma).$$[16]

While Theorem 4 focuses on the application of the linear gain-loss function to a value function V and a probability measure m on continuation paths, it is easy to see that the same optimality conditions for γ are necessary and sufficient for any random variable

[16] I am adopting the typical abbreviated notation: $m(V < \gamma)$ means $m(\{(c, m) \in D : V(c, m) < \gamma\})$. Also, note that if $\lambda^l = \lambda^g = 1$, then this expression is indeterminate. However, it is easy to see that in this case, $\mathbb{E}_m[\phi(V|\gamma)] = \mathbb{E}_m[V]$ for all $\gamma \in I$. Hence, any choice of γ is optimal.

defined on any probability space. Hence, the same quantile condition applies in the two-period linear gain-loss model, as well as in later applications to portfolio choice (where the value function depends on wealth and a state variable).

To better understand the condition in this result, consider first the case where m induces a continuous distribution of the value function V, so $m(V = \gamma) = 0$ for any γ. In this case, the optimality condition in Theorem 4 simplifies to $m(V \leq \gamma) = \frac{1-\lambda^g}{\lambda^l-\lambda^g}$. The inequality in the theorem allows for the possibility that the distribution of V could have atoms. For example, suppose $m(V = a) = m(V = b) = \frac{1}{2}$ for $a < b$, and suppose $\frac{1-\lambda^g}{\lambda^l-\lambda^g} = \frac{1}{2}$. Then, any γ satisfying $a \leq \gamma \leq b$ maximizes $\mathbb{E}_m[\phi(V|\gamma)]$.

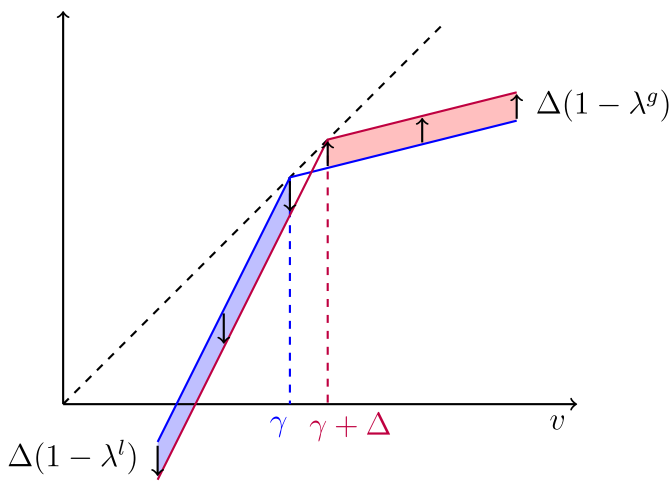

Figure 3: Effect of Increasing γ by Δ

Figure 3 illustrates the intuition behind Theorem 4 for parameter values $\lambda^l > 1 > \lambda^g$. Fixing any initial value γ, it is easy to see that increasing anticipation slightly to $\gamma+\Delta$ has the following effects: (i) for realizations $v \leq \gamma$, utility $\phi(v|\gamma)$ changes by $\Delta(1-\lambda^l) < 0$; (ii) for realizations $v \geq \gamma + \Delta$, utility changes by $\Delta(1-\lambda^g) > 0$. In the limit as $\Delta \to 0$, the area between γ and $\gamma + \Delta$ vanishes and can be ignored. Thus, the rate of change of $\mathbb{E}_m[\phi(V|\gamma)]$ with respect to an infinitesimal increase in γ is

$$m(V \leq \gamma)(1-\lambda^l) + m(V > \gamma)(1-\lambda^g).$$

A necessary condition for γ to be optimal is that this value be weakly less than 0. Rearranging terms gives $\frac{1-\lambda^g}{\lambda^l-\lambda^g} \leq m(V \leq \gamma)$. A similar analysis for the case of a decrease in γ gives the other inequality in the optimality condition of Theorem 4. Intuitively, since $\phi(\cdot|\gamma)$ is piecewise linear, the exact realized value of V is not import for the choice of γ, only the probability that V is to above or below γ. Hence, the optimal level of anticipation can be determined from a simple quantile formula, making the linear gain-loss function particularly amenable to applications.

4.2 Epstein-Zin-Kreps-Porteus Expected Utility

While the general formulation in Epstein and Zin (1989) permits a large variety of non-expected-utility certainty equivalents, the most widely applied special case is expected utility in the sense of Kreps and Porteus (1978). In this case the value function V is given by

$$V(c,m) = u(c) + \beta h^{-1}\left(\mathbb{E}_m\big[h(V)\big]\right) \tag{16}$$

for some strictly increasing and continuous function $h : V(D) \to \mathbb{R}$. Such a value function is typically referred to as *Epstein-Zin-Kreps-Porteus (EZKP) expected utility.* The usual specification of EZKP utility gives both u and h homogeneous (CES) form; this special case will be described in detail in Section 5.2.2.

Although preferences over lotteries $m \in \triangle(D)$ satisfy the expected-utility axioms, it is important to note that EZKP utility does not correspond to additively-separable expected utility unless $h(v) = v$ for all $v \in V(D)$. As emphasized by Epstein and Zin (1989), incorporating the function h allows for a separation between risk aversion and intertemporal substitution that is not possible in the additively-separable model. Since the h^{-1} and h transformations cancel out for deterministic consumption streams, intertemporal substitution is determined entirely by u. However, when facing risk about future consumption, the transformation h alters risk aversion. The commonly used (and empirically more relevant) case is where h is a concave transformation, and therefore risk aversion is increased.

In this section, I show that the EZKP representation with concave h corresponds to a special case of the Optimal Anticipation model. This connection will be important for two reasons. First, it shows that the Optimal Anticipation representation is general enough to incorporate a class of preferences that has proved quite powerful in applied work. Second, it provides a simple and concrete interpretation for the separation of risk aversion and intertemporal substitution based on reference-dependent utility.

The simplest approach to establishing this connection is to utilize the role of convexity in the Optimal Anticipation representation. Specifically, if h is concave then h^{-1} is convex; thus the value function V in Equation (16) is convex in m. The construction described in Section 3.4 can therefore be applied to obtain a collection of functions Φ such that for all $m \in \triangle(D)$,

$$h^{-1}\left(\mathbb{E}_m\big[h(V)\big]\right) = \sup_{\phi\in\Phi} \mathbb{E}_m\big[\phi(V)\big].$$

While these observations are sufficient to show that EZKP utility is a special case of the Optimal Anticipation representation, the following result goes a step further and describes the specific functional form of the corresponding representation. Having such a closed-form solution will be useful, for example, when applying the asset pricing equations derived in Section 6 to various transformations h.

Theorem 5 *Suppose $h : I \to \mathbb{R}$ is differentiable,*[17] *concave, and $h' > 0$, where I is an interval in $\mathbb{R}$ (not necessarily bounded). For any measurable function $V : D \to I$, if $h \circ V$ is integrable with respect to $m \in \triangle(D)$, then*

$$h^{-1}\left(\mathbb{E}_m\left[h(V)\right]\right) = \max_{\gamma \in I} \mathbb{E}_m\left[\frac{h(V)}{h'(\gamma)} - \frac{h(\gamma)}{h'(\gamma)} + \gamma\right].$$

Morevoer, the right-hand side is maximized by $\gamma = h^{-1}\left(\mathbb{E}_m[h(V)]\right)$.

Proof: The concavity of h implies that $h(v) - h(\gamma) \leq h'(\gamma)(v - \gamma)$ for any $\gamma, v \in I$. Rearranging terms yields

$$\frac{h(v)}{h'(\gamma)} - \frac{h(\gamma)}{h'(\gamma)} + \gamma \leq v,$$

with equality if $\gamma = v$. Taking $v = h^{-1}\left(\mathbb{E}_m[h(V)]\right)$, this implies

$$\mathbb{E}_m\left[\frac{h(V)}{h'(\gamma)} - \frac{h(\gamma)}{h'(\gamma)} + \gamma\right] = \frac{\mathbb{E}_m[h(V)]}{h'(\gamma)} - \frac{h(\gamma)}{h'(\gamma)} + \gamma \leq h^{-1}\left(\mathbb{E}_m\left[h(V)\right]\right),$$

with equality if $\gamma = h^{-1}\left(\mathbb{E}_m[h(V)]\right)$. $\blacksquare$

The connection between EZKP utility and the Optimal Anticipation representation follows directly from this result. Suppose (V, u, h, β) is an EZKP representation as in Equation (16). If h satisfies the conditions in Theorem 5 on $I = V(D)$, then V corresponds to the parameterized Optimal Anticipation representation in Equation (14) for

$$\phi(v|\gamma) = \frac{h(v)}{h'(\gamma)} - \frac{h(\gamma)}{h'(\gamma)} + \gamma.$$

Note also that since h is increasing and concave, this parameterization satisfies $\phi(v|\gamma) \leq v$, with equality at $v = \gamma$.

[17] Differentiability is only assumed for expositional simplicity. As is clear from the proof, if h is not differentiable at a point γ, then $h'(\gamma)$ in Theorem 5 can be replaced by any scalar α in the superdifferential of h at γ, that is, any α greater than the right derivative of h and less than the left derivative.

5 Homothetic Preferences

In applications to finance and macroeconomics, it is natural to focus on the case where the consumption space in each period is the non-negative real numbers, $C = \mathbb{R}_+$. To improve tractability and facilitate analysis using a representative agent, it is also useful to work with preferences that are homothetic: the ranking of any pair of stochastic consumption streams remain unchanged if both streams are scaled by the same multiplicative factor. This assumption simplifies portfolio choice problems (see Section 6), since it implies the fraction of current wealth consumed in any period and the portfolio decision will both be independent of the actual amount of current wealth.

As in Epstein and Zin (1989), it is possible to obtain a homothetic utility representation by assuming u takes a CES form and imposing appropriate homogeneity or shift-invariance assumptions on the certainty equivalent function. Specifically, if $u(c) = \frac{c^\rho}{\rho}$ for some $0 \neq \rho < 1$ and the certainty equivalent is homogeneous of degree 1, then preferences will be homothetic. The case of $\rho = 0$ corresponds to log utility, and when $u(c) = \log(c)$, a shift-invariant certainty equivalent delivers homothetic preferences.

Homogeneous and shift-invariant versions of the Optimal Anticipation certainty equivalent are described formally in Section 5.1. In Section 5.2, the linear gain-loss utility and EZKP utility from Section 4 are revisited and homothetic special cases are characterized. Section 5.3 then describes some additional continuity properties that are obtained for homogeneous and shift-invariant certainty equivalents.

5.1 Homogeneity and Shift Invariance

The results in this section are developed to facilitate the applications in Section 6. It is therefore useful to simplify the analysis by focusing on the parameterized version of the model from Equation (14). As in the examples in Section 4, suppose the set of values I for the parameter γ is the same as the set of continuation utility values, and impose the following assumption to ensure that $\sup_{\gamma \in I} \phi(v|\gamma) = v$ for every $v \in I$.

Assumption 1 *For each $\gamma \in I$, the function $\phi(\cdot|\gamma) : I \to [-\infty, \infty)$ is nondecreasing, concave, and satisfies $\phi(v|\gamma) \leq v$ for all $v \in I$, with equality at $v = \gamma$.*

Note that Assumption 1 allows the gain-loss function to take the value $-\infty$. In order to incorporate CES utility specifications, it is necessary to relax some of the continuity assumptions imposed in the axiomatic analysis of Section 3 and allow functions to take values in the extended real numbers. For example, if $\rho < 0$, then the utility function $u(c) = \frac{c^\rho}{\rho}$ takes the value $-\infty$ at $c = 0$. Similarly, several examples of gain-loss functions used later will take $\phi(0|\gamma) = -\infty$ for $\gamma > 0$.

As in Theorem 1, if the range of u includes only positive (negative, respectively) numbers, then the range of the value function will also include only positive (negative, respectively) numbers. Therefore, the domain I of the gain-loss functions $\phi(\cdot|\gamma)$ will depend on whether $\rho > 0$, $\rho < 0$, or $\rho = 0$ (log utility) in the CES specification of u. The relevant cases are summarized as follows:

- *Case 1:* If $u(c) = \frac{c^\rho}{\rho}$ for $\rho \in (0, 1)$, then $u(\mathbb{R}_+) = \mathbb{R}_+$. Let $I = \mathbb{R}_+$.
- *Case 2:* If $u(c) = \frac{c^\rho}{\rho}$ for $\rho < 0$, then $u(\mathbb{R}_+) = [-\infty, 0)$. Let $I = (-\infty, 0)$.[18]
- *Case 3:* If $u(c) = \log(c)$, then $u(\mathbb{R}_+) = [-\infty, \infty)$. Let $I = \mathbb{R}$.

In each of the three cases described above, the parameterized gain loss function gives a certainty equivalent as in Equation (10):

$$W(\mu) = \sup_{\gamma \in I} \int \phi(v|\gamma)\, d\mu(v). \tag{17}$$

The measure μ is a distribution over continuation values I. In this section, I analyze the properties of the certainty equivalent in isolation from the rest of the Optimal Anticipation representation. However, it is easy to apply this certainty equivalent in a recursive representation with value function V. For any measure $m \in \triangle(D)$ over continuation paths, one can simply take the induced distribution over continuation values $\mu = m \circ V^{-1}$ and apply the change of variables from Equation (11).

As a brief technical aside, the integral in Equation (17) may not be defined for a measure μ with unbounded support. However, for any measure with bounded support, Assumption 1 ensures this expression gives a well-defined certainty equivalent. To see this, fix any bounded interval $[a, b] \subset I$ and $\mu \in \triangle([a, b])$. Note that $\phi(v|\gamma) \leq v \leq b$ for any $\gamma \in I$ and $v \in [a, b]$. Therefore, the integral $\int_a^b \phi(v|\gamma)\, d\mu(v)$ is defined and bounded above by b for every $\mu \in \triangle([a, b])$, although it make take the value $-\infty$. Then, since $\phi(v|a) \geq \phi(a|a) = a$ for all $v \in [a, b]$, it follows that $a \leq W(\mu) \leq b$ for all $\mu \in \triangle([a, b])$. Since W is well-defined in this case, the results in this section will assume explicitly that measures have bounded support. In the portfolio choice problem considered in Section 6, bounds on returns will play the same role.

With the framework and technical assumptions in place, I now return to the desired properties of the certainty equivalent: homogeneity and shift invariance. First, define these properties formally for the underlying gain-loss function.

[18] It will be notationally more convenient in the sequel to exclude extended real numbers from the set of parameter values I in cases 2 and 3. This is possible by adopting the convention that $\phi(-\infty|\gamma) = -\infty$ for all $\gamma \in I$.

Definition 2 Consider a parameterized function $\phi(\cdot|\gamma) : I \to [-\infty, \infty)$ for $\gamma \in I$.

- In the case of $I = \mathbb{R}_+$ or $I = (-\infty, 0)$, say $\phi(\cdot|\cdot)$ is *homogeneous* (of degree 1) if $\phi(\alpha v|\alpha\gamma) = \alpha\phi(v|\gamma)$ for all $\alpha > 0$ and $v, \gamma \in I$.
- In the case of $I = \mathbb{R}$, say $\phi(\cdot|\cdot)$ is *shift invariant* if $\phi(\alpha + v|\alpha + \gamma) = \alpha + \phi(v|\gamma)$ for all $\alpha \in \mathbb{R}$ and $v, \gamma \in I$.

The following result shows that homogeneity of $\phi(\cdot|\cdot)$ yields homogeneity of the certainty equivalent. Similarly, shift invariance carries over from the gain-loss function to the certainty equivalent.

Lemma 2 *Suppose $\phi(\cdot|\cdot)$ satisfies Assumption 1 on I.*

1. *In the case of $I = \mathbb{R}_+$ or $I = (-\infty, 0)$, if $\phi(\cdot|\cdot)$ is homogeneous, then for any $\mu \in \triangle(I)$ with bounded support and $\alpha > 0$,*

$$\sup_{\gamma \in I} \int \phi(\alpha v|\gamma)\, d\mu(v) = \alpha \sup_{\gamma \in I} \int \phi(v|\gamma)\, d\mu(v).$$

2. *In the case of $I = \mathbb{R}$, if $\phi(\cdot|\cdot)$ is shift invariant, then for any $\mu \in \triangle(I)$ with bounded support and $\alpha \in \mathbb{R}$,*

$$\sup_{\gamma \in I} \int \phi(\alpha + v|\gamma)\, d\mu(v) = \alpha + \sup_{\gamma \in I} \int \phi(v|\gamma)\, d\mu(v).$$

Proof: For part 1, homogeneity implies that $\alpha \int \phi(v|\gamma)\, d\mu(v) = \int \phi(\alpha v|\alpha\gamma)\, d\mu(v)$ for any $\gamma \in I$. Therefore,

$$\alpha \sup_{\gamma \in I} \int \phi(v|\gamma)\, d\mu(v) = \sup_{\gamma \in I} \int \phi(\alpha v|\alpha\gamma)\, d\mu(v) = \sup_{\gamma \in I} \int \phi(\alpha v|\gamma)\, d\mu(v),$$

as desired. The proof of part 2 is similar. ■

Lemma 2 implies that homogeneous and shift-invariant certainty equivalents are easily characterized within the parameterized Optimal Anticipation model—simply impose the same properties on the gain-loss function. Moreover, homogeneity or shift invariance of the parameterized family $\phi(\cdot|\gamma)$ implies that it is in fact determined by a *single* function. For instance, consider the case of homogeneous $\phi(\cdot|\cdot)$ on $I = \mathbb{R}_+$. Then, $\phi(v|\gamma) = \gamma\phi(\frac{v}{\gamma}|1)$ for any $\gamma > 0$, so the parameterized gain-loss function is determined for each $\gamma > 0$ by specifying $\phi(\cdot|1)$. Similarly, a homogeneous gain-loss function on $I = (-\infty, 0)$ satisfies

$\phi(v|\gamma) = -\gamma\phi(\frac{v}{-\gamma}|-1)$ and is therefore determined for each $\gamma < 0$ by specifying $\phi(\cdot|-1)$. Finally, a shift-invariant gain-loss function on $I = \mathbb{R}$ satisfies $\phi(v|\gamma) = \gamma + \phi(v - \gamma|0)$, and hence it suffices to specify $\phi(\cdot|0)$.

This property of homogeneous and shift-invariant gain-loss functions permits examples to be easily constructed. To illustrate for the case of $I = \mathbb{R}_+$, any function $\varphi : I \to [-\infty, \infty)$ can be chosen as a candidate for $\phi(\cdot|1)$, so long as it is nondecreasing, concave, and satisfies $\varphi(v) \leq v$ for all $v \geq 0$, with equality at $v = 1$. For any such φ, define the function $\phi(\cdot|\gamma) : I \to [-\infty, \infty)$ for $\gamma \geq 0$ as follows:[19]

$$\phi(v|\gamma) = \begin{cases} \gamma\varphi(\frac{v}{\gamma}) & \text{if } \gamma > 0, \\ 0 & \text{if } \gamma = 0. \end{cases}$$

Although φ need not be homogeneous, this construction guarantees that the function $\phi(\cdot|\cdot)$ and the resulting certainty equivalent will be. Analogous constructions are possible for homogeneous gain-loss functions on $I = (-\infty, 0)$ and shift invariant gain-loss functions on $I = \mathbb{R}$.

5.2 Examples

5.2.1 Homothetic Linear Gain-Loss Utility

Consider again the linear gain-loss function analyzed in Section 4.1. In this section, I make the simple observation that it is both homogeneous and shift invariant, and therefore can be applied to each of the three cases described in the previous section to obtain homothetic preferences. For any $v, \gamma \in \mathbb{R}$ and $\alpha > 0$,

$$\begin{aligned} \phi(\alpha v|\alpha\gamma) &= \alpha\gamma + \begin{cases} \lambda^l(\alpha v - \alpha\gamma) & \text{if } \alpha v < \alpha\gamma \\ \lambda^g(\alpha v - \alpha\gamma) & \text{if } \alpha v \geq \alpha\gamma \end{cases} \\ &= \alpha\phi(v|\gamma). \end{aligned}$$

In addition, for any $v, \gamma \in \mathbb{R}$ and $\alpha \in \mathbb{R}$,

$$\begin{aligned} \phi(\alpha + v|\alpha + \gamma) &= \alpha + \gamma + \begin{cases} \lambda^l((\alpha + v) - (\alpha + \gamma)) & \text{if } \alpha + v < \alpha + \gamma \\ \lambda^g((\alpha + v) - (\alpha + \gamma)) & \text{if } \alpha + v \geq \alpha + \gamma \end{cases} \\ &= \alpha + \phi(v|\gamma). \end{aligned}$$

[19] In the case of $I = \mathbb{R}_+$, $\phi(\cdot|0)$ is not pinned down from φ, and homogeneity only requires that $\phi(v|0) = \alpha v$ for some $\alpha \in [0, 1]$. Taking $\alpha = 0$ will prove convenient for subsequent results. For example, this ensures that the optimal γ is greater than 0 for any μ with support in $\mathbb{R}_{++}$ (see Theorem 6).

Thus, the linear gain-loss function is homogeneous on both $I = \mathbb{R}_+$ and $I = (-\infty, 0)$, and it is shift invariant on $I = \mathbb{R}$. This implies it can be used in conjunction with the utility function $u(c) = \frac{c^\rho}{\rho}$ for any $0 \neq \rho < 1$ or with log utility $u(c) = \log(c)$ to deliver a homothetic value function.

5.2.2 Homothetic Epstein-Zin-Kreps-Porteus Expected Utility

It was shown in Section 4.2 that Epstein-Zin-Kreps-Porteus expected utility can be expressed as a special case of the Optimal Anticipation representation. As noted previously, the most common specification of EZKP utility in applications takes u and h to be homogeneous functions. For example, let $u(c) = \frac{c^\rho}{\rho}$ and $h(v) = v^{\alpha/\rho}$ for $\rho, \alpha \in (0, 1)$. Then, Equation (16) simplifies to

$$V(c,m) = \frac{c^\rho}{\rho} + \beta \left(\mathbb{E}_m\left[V^{\alpha/\rho}\right]\right)^{\rho/\alpha}.$$

In this representation, the elasticity of intertemporal substitution is $\frac{1}{1-\rho}$ and the coefficient of relative risk aversion is $1 - \alpha$ (see Epstein and Zin (1989) for discussion of this interpretation of the parameters). More generally, to consider any $0 \neq \rho < 1$ and $0 \neq \alpha < 1$, define $u(c) = \frac{c^\rho}{\rho}$ and $h(v) = \alpha\big(\rho v\big)^{\alpha/\rho}$. These definitions ensure that u and h are well-defined and increasing even when ρ and α take negative values. In this case, the EZKP value function takes the form:[20]

$$V(c,m) = \frac{c^\rho}{\rho} + \beta \frac{1}{\rho} \left(\mathbb{E}_m\Big[\big(\rho V\big)^{\alpha/\rho}\Big]\right)^{\rho/\alpha}. \tag{18}$$

The following corollary of Theorem 5 connects the homogeneous EZKP and Optimal Anticipation representations for two cases: $\rho \in (0, 1)$ and $\rho < 0$. The parameter restriction $\alpha \leq \rho$ implies that the transformation h in the EZKP representation is concave.

Corollary 1 *Fix any $\rho, \alpha \neq 0$ such that $\alpha \leq \rho < 1$, and define $\phi(\cdot|\gamma)$ as follows:*[21]

$$\phi(v|\gamma) = \begin{cases} \gamma\left[\frac{\rho}{\alpha}\left(\frac{v}{\gamma}\right)^{\alpha/\rho} + 1 - \frac{\rho}{\alpha}\right] & \text{if } \gamma \neq 0 \\ 0 & \text{if } \gamma = 0. \end{cases}$$

1. *Suppose $\rho \in (0, 1)$ and $V : D \to \mathbb{R}_+$ satisfies Equation (18). Then,*

$$V(c,m) = \frac{c^\rho}{\rho} + \beta \max_{\gamma \geq 0} \mathbb{E}_m\big[\phi(V|\gamma)\big].$$

[20]This is a simple transformation of the value function considered by Epstein and Zin (1989). Letting $\hat{V}(c,m) = (\rho V(c,m))^{1/\rho}$ gives their recursive equation: $\hat{V}(c,m) = (c^\rho + \beta(\mathbb{E}_m[\hat{V}^\alpha])^{\rho/\alpha})^{1/\rho}$.

[21]To ensure $\phi(\cdot|\gamma)$ is well-defined, this expression is applied only to $v \geq 0$ if $\gamma > 0$ and to $v \leq 0$ if $\gamma < 0$.

2. Suppose $\rho < 0$ and $V : D \to [-\infty, 0)$ satisfies Equation (18). Then,

$$V(c,m) = \frac{c^\rho}{\rho} + \beta \max_{\gamma<0} \mathbb{E}_m\big[\phi(V|\gamma)\big].$$

Proof: Let $h(v) = \alpha(\rho v)^{\alpha/\rho}$, defined on $\mathbb{R}_+$ and $[-\infty, 0)$ for case 1 and 2, respectively. Then, $h^{-1}(\mathbb{E}_m[h(V)]) = \frac{1}{\rho}(\mathbb{E}_m[(\rho V)^{\alpha/\rho}])^{\rho/\alpha}$. Adopting the usual conventions for multiplying or dividing by $\pm\infty$, the arguments in Theorem 5 are easily extended to functions mapping into the extended real numbers. Note that

$$\frac{h(v)}{h'(\gamma)} - \frac{h(\gamma)}{h'(\gamma)} + \gamma = \gamma\Big[\frac{\rho}{\alpha}\Big(\frac{v}{\gamma}\Big)^{\alpha/\rho} + 1 - \frac{\rho}{\alpha}\Big].$$

These terms are well-defined in the relevant ranges except when $\alpha < 0 < \rho$, in which case $\frac{h(0)}{h'(0)} = \frac{-\infty}{+\infty}$. For this case, setting $\phi(v|\gamma) = 0$ for $\gamma = 0$ resolves the indeterminacy and gives the desired equivalence. ■

Corollary 1 does not consider the case of $\rho = 0$. In this case, CES utility takes the form $u(c) = \log(c)$, and EZKP preferences are homothetic if the value function satisfies

$$V(c,m) = \log(c) - \beta\frac{1}{\theta}\log \mathbb{E}_m\big[\exp(-\theta V)\big] \tag{19}$$

for some $\theta > 0$.[22] As emphasized by Tallarini (2000), after taking appropriate transformations of the value function, Equation (19) with $\theta = -\alpha(1-\beta)$ corresponds to the limiting case of Equation (18) for $\rho \to 0$ (see also Hansen, Heaton, Lee, and Roussanov (2007)). In this case, the elasticity of intertemporal substitution is 1 and the coefficient of relative risk aversion is again $1 - \alpha$. The specification if Equation (19) has been used in a number of macroeconomic applications (e.g., Hansen, Sargent, and Tallarini (1999) and Tallarini (2000)), and has also been reinterpreted in terms of robustness to model uncertainty.[23]

Corollary 2 *Suppose $\theta > 0$ and $V : D \to [-\infty, \infty)$ satisfies Equation (19). Then,*

$$V(c,m) = \log(c) + \beta \max_{\gamma\in\mathbb{R}} \mathbb{E}_m\big[\phi(V|\gamma)\big],$$

where $\phi(\cdot|\gamma)$ is defined by $\phi(v|\gamma) = \gamma + \frac{1}{\theta} - \frac{1}{\theta}\exp(-\theta(v-\gamma))$ for $\gamma \in \mathbb{R}$.

[22] The limiting case of $\theta = 0$ corresponds to additively-separable utility $V(c,m) = \log(c) + \beta\mathbb{E}_m[V]$.

[23] The connection between Equation (19) and the multiplier preferences of Hansen and Sarget (2001) has been established in a variety of settings. See, for example, Skiadas (2003), Maenhout (2004), and Strzalecki (2011). For a detailed discussion of this reinterpretation in the context of the equity premium puzzle, see Barillas, Hansen, and Sargent (2009).

Proof: Again adopt the usual conventions for $\pm\infty$, and apply the formula from Theorem 5 to $h(v) = -\exp(-\theta v)$. ■

5.3 Continuity of Homothetic Representations

Since the utility functions $u(c) = \frac{c^\rho}{\rho}$ for $\rho < 0$ and $u(c) = \log(c)$ take the value $-\infty$ at $c = 0$, the homothetic Optimal Anticipation representation will necessarily have points of discontinuity. However, it will be important for later results that this value function is continuous when consumption is bounded away from zero. While the continuity of any CES utility function u is immediate in this case, the following result shows that this continuity is preserved by the certainty equivalent.

Theorem 6 *Suppose I, $\phi(\cdot|\cdot)$, and the scalars $a, b \in I$ satisfy one of the following sets of assumptions:*

1. *$I = \mathbb{R}_+$, $\phi(\cdot|\cdot)$ satisfies Assumption 1 and homogeneity, $\phi(v|0) = 0$, and $0 < a < b$;*
2. *$I = (-\infty, 0)$, $\phi(\cdot|\cdot)$ satisfies Assumption 1 and homogeneity, and $a < b < 0$; or*
3. *$I = \mathbb{R}$, $\phi(\cdot|\cdot)$ satisfies Assumption 1 and is shift invariant, and $a < b$.*

Suppose also that $\phi(v|\gamma) > -\infty$ for all v, γ in some open set containing $[a, b]$.[24] *Then, for any $\mu \in \triangle([a, b])$, there exists $\gamma \in [a, b]$ that solves the maximization problem in Equation (17), and the function $W : \triangle([a, b]) \to [a, b]$ defined by*

$$W(\mu) = \max_{\gamma \in I} \int \phi(v|\gamma)\, d\mu(v)$$

is continuous in the topology of weak convergence.

Theorem 6 is based on relatively standard arguments. Using concavity of each $\phi(\cdot|\gamma)$ together with either homogeneity or shift invariance, it can be shown that $\phi(v|\gamma) \le \phi(v|a)$ for any $\gamma < a$ and $v \in [a, b]$. Likewise, $\phi(v|\gamma) \le \phi(v|b)$ for any $\gamma > b$ and $v \in [a, b]$. This implies that one can restrict attention to $\gamma \in [a, b]$ when solving the maximization problem in Equation (17). The assumed properties of $\phi(\cdot|\cdot)$ can also be used to show that $\int \phi(v|\gamma)\, d\mu(v)$ is continuous in both $\gamma \in [a, b]$ and $\mu \in \triangle([a, b])$. The existence of a maximizing γ then follows directly, and the continuity of the certainty equivalent function follows from Berge's Maximum theorem.

[24] If $\phi(\cdot|\gamma)$ is real valued then this restriction is trivially satisfied. For another example, if $I = \mathbb{R}_+$ and $\phi(v|\gamma) > -\infty$ for all $v > 0$, then any $0 < a < b$ will satisfy this restriction.

6 Application to Portfolio Choice

In this section, I determine the equilibrium asset pricing conditions for a representative agent economy. Let $a_t \geq 0$ denote the agent's wealth at the start of period t and $c_t \geq 0$ denote consumption in period t. Let Z be a finite set, representing the state of the economy. Assume the state $z \in Z$ evolves according to a stationary Markov process with transition probabilities $P(z_{t+1}|z_t)$. The state z_t is perfectly observed by the agent at the start of period t.

Suppose there are K assets, and the gross return to holding asset k between t and $t+1$ is $r_k(z_{t+1})$. Let w_{kt} denote the fraction of post-consumption wealth invested in asset k in period t. For $c_t \in [0, a_t]$ and $w_t \in \triangle(K) \equiv \{w \in \mathbb{R}_+^K : \sum_k w_k = 1\}$, the wealth in period $t+1$ is given by the following budget constraint: $a_{t+1} = (a_t - c_t)(w_t \cdot r(z_{t+1})) = (a_t - c_t)(\sum_k w_{kt} r_k(z_{t+1}))$.

6.1 Existence of a Value Function

In this section, I show that for the homothetic parameterized representations described in Section 5, there exists a unique homothetic value function. These results currently focus on the case of $u(c) = \frac{c^\rho}{\rho}$ for $\rho \in (0,1)$, but can also be established for $\rho < 0$ and log utility ($\rho = 0$) (this will be done in subsequent versions of this paper).

Since the environment is stationary, it is possible to drop the period t and $t+1$ subscripts and obtain a value function that depends solely on the current state and wealth.

Theorem 7 *Suppose $\rho \in (0,1)$, $\beta \in (0,1)$, and for every $z \in Z$ and $1 \leq k \leq K$,*

$$r_k(z) > 0 \quad \textit{and} \quad \beta^{1/\rho} \int r_k(z')\, dP(z'|z) < 1.$$

Suppose $\phi(\cdot|\cdot)$ satisfies Assumption 1 and homogeneity on $I = \mathbb{R}_+$, $\phi(v|0) = 0$, and $\phi(v|\gamma) > -\infty$ for all $v > 0$. Then, there exists a unique homogeneous (of degree ρ) value function $J : \mathbb{R}_+ \times Z \to \mathbb{R}_+$ satisfying

$$J(a,z) = \max_{\substack{c \in [0,a] \\ w \in \triangle(K)}} \left[\frac{c^\rho}{\rho} + \beta \max_{\gamma \geq 0} \int_Z \phi\Big(J\big(w \cdot r(z')(a-c),\, z'\big) \Big| \gamma \Big)\, dP(z'|z) \right]. \qquad (20)$$

Except in a very restricted set of cases, the usual techniques from Blackwell (1965) cannot be applied to obtain a value function in this environment. Instead, the existence of the function J follows from an alternative fixed point theorem from Marinacci and

Montrucchio (2010) (see Appendices A and B.4). Once the existence of a value function is established, the existence of maximizing c, w, and γ can be obtained used the continuity properties from Theorem 6.

6.2 Asset Returns

Since preferences are homothetic, it is possible to represent the entire economy using a representative agent and hence obtain an Euler equation for asset returns based on aggregate variables. Using techniques from Epstein and Zin (1989) (see also Backus, Routledge, and Zin (2004) for a survey), the following result gives equilibrium conditions for aggregate consumption and asset returns. To simplify notation, denote the expectation conditional on the information z_t held by the agent at time t by $\mathbb{E}_t(\cdot)$, and write r_{t+1} for $r(z_{t+1})$, $r_{k,t+1}$ for $r_k(z_{t+1})$, and so on.

Theorem 8 *Suppose ρ, β, r, and $\phi(\cdot|\cdot)$ are as in Theorem 7. If $\phi(\cdot|\gamma)$ is differentiable at all $v > 0$, then in equilibrium the following first-order condition is satisfied for assets $1 \le k \le K$:*[25]

$$\beta\mathbb{E}_t\left[\phi_1\left(\left(\frac{c_{t+1}}{c_t}\right)^{\rho-1}(w_t \cdot r_{t+1})\middle|\gamma_t\right)\left(\frac{c_{t+1}}{c_t}\right)^{\rho-1} r_{k,t+1}\right] = 1, \tag{21}$$

where

$$\gamma_t \in \operatorname*{argmax}_{\gamma \ge 0} \mathbb{E}_t\left[\phi\left(\left(\frac{c_{t+1}}{c_t}\right)^{\rho-1}(w_t \cdot r_{t+1})\middle|\gamma\right)\right]. \tag{22}$$

7 Discussion of Related Literature

7.1 Reference-Dependent Utility

In this section, I discuss in more detail the conceptual differences between the model of reference-point formation proposed in this paper and the other approaches mentioned in the introduction. One common alternative approach is to assume that reference points are based on past experience. For example, Benartzi and Thaler (1995) (and later Barberis and Huang (2001) and Barberis, Huang, and Santos (2001)) took a history-dependent approach to apply reference-dependent utility in finance. They assumed that past asset returns determine the reference point against which changes in the current value of the portfolio are measured.[26] Although it is quite plausible that past returns influence indi-

[25]Here, $\phi_1(\cdot|\cdot)$ denotes the partial derivative of $\phi(\cdot|\cdot)$ with respect to its first argument.

[26]These models concern not only reference-dependence, but also narrow framing—they assume individuals are not loss averse with respect to consumption directly, but with respect to fluctuations in their financial wealth.

viduals' attitudes toward current returns, in many situations it is natural to expect that there is also a forward-looking component to reference point formation. For instance, suppose an individual who has been earning a consistent 8% rate of return on her portfolio receives a tip about a new investment that she believes will earn a return of 25% in the coming year. Suppose the realized return from this investment is ultimately 9%. The history-dependent approach would suggest that this outcome should be viewed as a gain relative to past returns of 8%, whereas it is quite conceivable that the individual will actually view this as a loss relative the expected returns of 25%.

As an alternative to history-dependent reference points, Kőszegi and Rabin (2006, 2007) adopted a forward-looking equilibrium-based approach. In their model, the alternative the individual plans to choose becomes her reference point. The concept of "personal equilibrium" is used to ensure no incentive to deviate from this plan, holding fixed the reference point: The alternative will actually be chosen only if it is better than every other alternative with itself taken as the reference point. For choice in stochastic environments, the same solution concept applies, with utility determined by taking expectations with respect to both the lottery being evaluated and the reference lottery. Once consequence of their model is that reference points are forced upon the individual by the lottery she faces, and therefore improvements in outcomes necessarily result in more demanding reference points. For example, suppose an individual is told that with some small probability $\varepsilon > 0$ an anonymous benefactor will give her \$1,000,000. Holding fixed her reference point, this would be unambiguously good news. However, in the Kőszegi-Rabin model, this possibility raises her reference point as well, in some cases to the extent that the individual is actually worse off.[27]

This paper maintains a forward-looking approach to reference dependence, but allows the individual more freedom in the determination of her reference point. This permits the model to capture some intuitive aspects of individuals' responses to risky environments that do not fit with previous models. For instance, if an individual is told that she may receive a large monetary payment with small probability as in the previous example, she would likely view this as good news. Still, she might describe her attitude toward the possibility with phrases like "not getting my hopes up," "not getting too excited," or "being cautiously optimistic." Although informal statements, these are suggestive that anticipating better outcomes (i.e., "getting one's hopes up" or "getting excited") makes the individual more sensitive to bad outcomes and, moreover, that the individual can exert some control over her anticipation. The model in this paper captures these intuitions by tying current anticipation to future reference points and allowing the individual to

[27]Kőszegi and Rabin (2007, Section 4) showed that an individual with their utility specification may prefer to commit at an early stage to have a first-order stochastically dominating lottery made unavailable. In the above example, this constitutes refusing early on to accept any possible future payments from a benefactor.

choose her anticipation optimally.

7.2 Betweenness Preferences

This model also has a connection to the "betweenness" class of static risk preferences considered by Chew (1983) and Dekel (1986), which have also been used successfully to address a number of issues in macroeconomics and finance (e.g., Epstein and Zin (1991), Bekaert, Hodrick, and Marshall (1997), Routledge and Zin (2010)). In the utility representation for betweenness preferences, there is a local utility function similar to the transformation ϕ used in the Optimal Anticipation representation. In particular, the local utility function for the disappointment aversion representation of Gul (1991) (which is a special case of the betweenness class) bears some resemblance to the specification of $\phi(\cdot|\gamma)$ in Example 1. Although they were not originally interpreted in such a way, these models could also be thought of as models of reference dependence.

However, the key distinction is that the local utility function for these representations is determined by a fixed point, rather than by optimization. Therefore, the conceptual difference is again that the individual is not able to exert control over her own reference point. Note that unlike the Kőszegi-Rabin model, betweenness preferences can never violate first-order stochastic dominance as in the benefactor example from Section 7.1. However, like the Kőszegi-Rabin model, betweenness preferences require that the reference point increases with respect to first-order stochastic dominance improvements to the chosen lottery, even if the individual would benefit from maintaining a lower reference point. From a practical standpoint, exploring the different implications of the Optimal Anticipation model in applications is an important question for future research.

7.3 Other Related Models

The term anticipation has been used in a number of contexts and assigned a number of meanings. For example, Caplin and Leahy (2001) and Epstein (2008) used the term anticipation to refer to anxiety about future uncertainty. In these models, this anticipation or anxiety is not modeled as a choice variable of the individual. More importantly, anticipatory utility distorts the preferences of the individual in one period relative to the next, leading (in general) to dynamically-inconsistent behavior.

In this paper, I have assumed that beliefs are fixed and the individual is free to choose her anticipation. There is a related literature that instead considers optimal beliefs. These papers assume that anticipatory utility is a fixed function of future expected utility (e.g., a scalar multiple) but the individual is able to distort her beliefs to increase her expectations and hence her anticipatory utility. For example, Brunnermeier and Parker

(2005) considered one such model of "optimal expectations." The individual in their model has the freedom to choose her beliefs in an optimistic way; however, instead of using loss aversion to discipline the choice of beliefs, they assumed that the individual's future decisions will be based on her chosen beliefs. Hence the "cost" of distorted beliefs is the poor outcomes that may result from suboptimal decisions. Similarly, Bénabou and Tirole (2011) modeled an individual who may wish to distort her beliefs about her own type to increase her self-esteem or anticipatory utility. The individual does not directly choose her beliefs, but has bounded memory of her type and engages in self-signaling using her past actions. Thus, more favorable beliefs again come at the expense of worse actions.

Finally, there is a connection between the Optimal Anticipation model and macroeconomic models that incorporate costs of adjusting consumption relative to planned or past consumption (see, e.g., Grossman and Laroque (1990), Gabaix and Laibson (2001), and the references therein). These models typically involve some consumption pre-commitment and a physical cost of changing consumption relative to these commitments. The process of anticipation and loss aversion in this paper could be thought of as the psychological version of such adjustment costs—the individual pre-commits to an anticipated utility and experiences a psychological cost if realized future utility differs from this benchmark.

A Fixed Points of Homogeneous Functions

Theorem 9 (Marinacci and Montrucchio (2010)) *Let Z be any nonempty set, and let $B(Z)$ denote the space of all bounded real-valued functions on Z, endowed with the supremum norm. Let X be a closed linear subspace of $B(Z)$ that contains the constant functions, and let $A = \{f \in X : 0 \leq f \leq b\}$ for some $b > 0$. Suppose $T : A \to A$ satisfies:*

1. *T is monotone, i.e., $f \leq g$ implies $Tf \leq Tg$.*
2. *There exists $\delta > 0$ such that $T(\alpha f) \geq \alpha T f + (1-\alpha)\delta$ for all $f \in A$ and $\alpha \in [0,1]$.*

Then, T has a unique fixed point $f \in A$. Moreover, this fixed point is globally attracting, i.e., $\|T^n g - f\|_\infty \to 0$ for any $g \in A$.

Theorem 9 is a very slight adaptation of Theorems 10 and 11 in Marinacci and Montrucchio (2010). The only substantive difference is that condition 2 in Theorem 9 uses a fixed $\delta > 0$, whereas Marinacci and Montrucchio (2010) used $T(0)$ in the place of δ. Their version is not sufficient for my applications of this result. However, it is clear that an immediate adaptation of their proof would yield my result. I still include a proof because the use of a fixed δ in condition 2 allows for a different approach to proving the result, one that may be somewhat illuminating: The conditions in Theorem 9 imply that T can be transformed using the exponential and logarithmic functions to obtain a contraction mapping (see details below).[28]

Lemma 3 *Let $\{f_n\}$ be a sequence in $B(Z)$, and suppose $f \in B(Z)$. If $\|f_n - f\|_\infty \to 0$, then $\|\exp \circ f_n - \exp \circ f\|_\infty \to 0$.*

Proof: Fix any $\varepsilon > 0$. Since f is bounded and $\|f_n - f\|_\infty \to 0$, there exists $a, b \in \mathbb{R}$ such that $a \leq f \leq b$ and $a \leq f_n \leq b$ for all n. Since the function $x \mapsto \exp(x)$ is Lipschitz continuous on the bounded interval $[a,b]$, there exists $K > 0$ such that $|\exp(x) - \exp(x')| \leq K|x - x'|$ for any $x, x' \in [a,b]$. Thus, $|\exp(f_n(z)) - \exp(f(z))| \leq K|f_n(z) - f(z)|$ for all $z \in Z$, and hence $\|\exp \circ f_n - \exp \circ f\|_\infty \leq K\|f_n - f\|_\infty \to 0$. ■

Proof of Theorem 9: First, consider the case of $\delta \geq b$. Condition 2 with $\alpha = 0$ implies $T(0) \geq \delta$. Since T is monotone and maps A into A, this implies $b \leq \delta \leq T(0) \leq Tf \leq b$, for all $f \in A$. Thus, $\delta = b$ and $Tf = b$ for all $f \in A$, implying $f = b$ is the unique fixed point of T.

Next, consider the case of $0 < \delta < b$. Since $T(0) \geq \delta$ by condition 2, monotonicity of T implies $\delta \leq Tf \leq b$ for all $f \in A$. In particular, any fixed point f of T must satisfy $\delta \leq f \leq b$.

[28] Marinacci and Montrucchio (2010) show that a different metric can be applied to A, one for which T is a contraction. I take the (mathematically equivalent) approach of maintaining the supremum norm, but transforming the function T in order to obtain a contraction.

Since $\frac{\delta}{b} \in (0,1)$, note that $\alpha + (1-\alpha)\frac{\delta}{b} \geq \alpha^{1-\frac{\delta}{b}}$ for any $\alpha \geq 0$.[29] Therefore, for any $f \in A$ and $\alpha \in [0,1]$, condition 2 implies

$$\begin{aligned} T(\alpha f) &\geq \alpha Tf + (1-\alpha)\delta \\ &= \alpha Tf + (1-\alpha)\left(\tfrac{\delta}{b}\right) b \\ &\geq \left(\alpha + (1-\alpha)\tfrac{\delta}{b}\right) Tf \geq \left(\alpha^{1-\frac{\delta}{b}}\right) Tf. \end{aligned} \tag{23}$$

Now, let $A' = \{f \in B(Z) : \exp \circ f \in A\}$. Note that A' is nonempty since it contains all constant functions less than $\log(b)$, and it is a closed subset of $B(Z)$ by Lemma 3. Define $T' : A' \to A'$ by $T'f = \log \circ (T(\exp \circ f))$. To see that T' maps into A', recall that $0 < \delta \leq T(\exp \circ f) \leq b$ for any $f \in A'$. Therefore, T' is well-defined and $\log(\delta) \leq T'f \leq \log(b)$, implying $T'f \in B(Z)$. Since, in addition, $\exp \circ (T'f) = T(\exp \circ f) \in A$, this implies $T'f \in A'$.

Fix any $f \in A'$ and $k > 0$, and let $\alpha = \exp(-k) < 1$. Then, by Equation (23),

$$\begin{aligned} T'(f-k) &= \log \circ \Big(T(\alpha \exp \circ f)\Big) \\ &\geq \log \circ \Big(\left(\alpha^{1-\frac{\delta}{b}}\right) T(\exp \circ f)\Big) \\ &= \left(1 - \tfrac{\delta}{b}\right)(-k) + T'f. \end{aligned} \tag{24}$$

Therefore, by the theorem of Blackwell (1965), T' is a contraction. Specifically, for any $f, g \in A'$, let $k = \|f - g\|_\infty$. Then, $f - k \leq g$, and by the monotonicity of T and Equation (24),

$$\left(1 - \tfrac{\delta}{b}\right)(-k) + T'f \leq T'(f-k) \leq T'g.$$

This implies $T'f - T'g \leq (1 - \frac{\delta}{b})\|f - g\|_\infty$, and reversing the roles of f and g gives $|T'f - T'g| \leq (1 - \frac{\delta}{b})\|f - g\|_\infty$. Since $1 - \frac{\delta}{b} < 1$, the Banach Contraction Mapping Theorem implies there is a unique $f \in A'$ such that $T'f = f$. Moreover, f is globally attracting in A', i.e., $\|T'^n g - f\|_\infty \to 0$ for any $g \in A'$.

Let $\hat{f} = \exp \circ f \in A$ for the fixed point f of T'. Then, $f = T'f = \log \circ (T\hat{f})$, which implies $T\hat{f} = \hat{f}$. Thus, $\hat{f}$ is a fixed point of T. To see that $\hat{f}$ is the unique and globally attracting fixed point of T in A, fix any $g \in A$. Then, $Tg \in A$ and $Tg \geq \delta$ as argued above, and hence $\log \circ (Tg) \in A'$. Therefore, $T'^n(\log \circ (Tg)) \to f$ in the supremum norm by the preceding arguments. However, since $T'^n(\log \circ (Tg)) = \log \circ (T^{n+1} g)$, this implies $\|\log \circ (T^n g) - f\|_\infty \to 0$. By Lemma 3, conclude that $\|T^n g - \hat{f}\|_\infty \to 0$. ■

Corollary 3 *Let X be as in Theorem 9. Let $A = \{f \in X : a \leq f \leq 0\}$ for some $a < 0$. Suppose $T : A \to A$ satisfies:*

[29] This inequality was also used by Marinacci and Montrucchio (2010) in their proof and is straightforward to verify: The derivative of the concave function $\alpha \mapsto \alpha^{1-\frac{\delta}{b}}$ at $\alpha = 1$ is $1 - \frac{\delta}{b}$. Likewise, the derivative of the affine function $\alpha \mapsto \alpha + (1-\alpha)\frac{\delta}{b}$ at $\alpha = 1$ is $1 - \frac{\delta}{b}$. Since both functions take the value 1 at $\alpha = 1$, the desired inequality follows.

1. *T is monotone, i.e., $f \le g$ implies $Tf \le Tg$.*

2. *There exists $\delta < 0$ such that $T(\alpha f) \le \alpha Tf + (1-\alpha)\delta$ for all $f \in A$ and $\alpha \in [0,1]$.*

Then, T has a unique fixed point $f \in A$. Moreover, this fixed point is globally attracting, i.e., $\|T^n g - f\|_\infty \to 0$ for any $g \in A$.

Proof: Let $A' = -A = \{f \in X : 0 \le f \le -a\}$, and define $T' : A' \to A'$ by $T'f = -T(-f)$ for $f \in A'$. Then,

$$f \le g \implies -f \ge -g \implies T(-f) \ge T(-g) \implies T'f \le T'g.$$

Thus, T' is also monotone. For any $f \in A'$ and $\alpha \in [0,1]$,

$$T'(\alpha f) = -T(\alpha(-f)) \ge -\Big[\alpha T(-f) + (1-\alpha)\delta\Big] = \alpha T'f + (1-\alpha)(-\delta).$$

Apply Theorem 9 to conclude that T' has a unique and globally attracting fixed point $f \in A'$. Then, $T(-f) = -T'f = -f$, so $-f \in A$ is the unique and globally attracting fixed point of T in A. ■

B Proofs

B.1 Proof of Proposition 1

Proposition 1 will be proved by means of a separation argument. Let $C([a,b])$ denote the set of all continuous functions on $[a,b]$ and let $ca([a,b])$ denote the set of all signed (countably-additive) Borel measures of bounded variation on the interval $[a,b]$. Consider the following subset of $ca([a,b])$:

$$F \equiv \left\{\mu \in ca([a,b]) : \int f(x)\,d\mu(x) \ge 0 \text{ for every nondecreasing } f \in C([a,b])\right\}. \tag{25}$$

Note that F is a cone in $ca([a,b])$. In addition, since the constant functions identically equal to 1 and -1 are both nondecreasing, $\mu([a,b]) = 0$ for all $\mu \in F$. The following lemma makes some other simple observations about F that will be central to the proof of the proposition.

Lemma 4 *The set F defined in Equation (25) is a weak* closed and convex subset of $ca([a,b])$. Moreover, for any $\mu, \eta \in \triangle([a,b])$, the following are equivalent:*

1. *μ first order stochastically dominates η, i.e., $\mu([x,b]) \ge \eta([x,b])$ for all $x \in [a,b]$.*

2. *$\int f(x)\,d\mu(x) \ge \int f(x)\,d\eta(x)$ for every nondecreasing $f \in C([a,b])$.*

3. $\mu - \eta \in F$.

Proof: For any $f \in C([a,b])$, the set

$$F_f \equiv \left\{\mu \in ca([a,b]) : \int f(x)\, d\mu(x) \geq 0\right\}$$

is weak* closed and convex. The set F is the intersection of such sets F_f, with the intersection being taken over all nondecreasing $f \in C([a,b])$. As the intersection of closed and convex sets, F is also closed and convex. To complete the proof, note that the equivalence of 1 and 2 is a standard result, and the equivalence of 2 and 3 is immediate. ■

Proof of Proposition 1: To see that 2 implies 1, suppose Φ is a collection of continuous and nondecreasing functions. If μ first order stochastically dominates η, then

$$\int \phi(x)\, d\mu(x) \geq \int \phi(x)\, d\eta(x), \quad \forall \phi \in \Phi.$$

Therefore,

$$\sup_{\phi\in\Phi} \int \phi(x)\, d\mu(x) \geq \sup_{\phi\in\Phi} \int \phi(x)\, d\eta(x).$$

To prove that 1 implies 2, it suffices to show that for any $\mu \in \triangle([a,b])$ and any $\alpha < W(\mu)$, there exists a nondecreasing $\phi_{\mu,\alpha} \in C([a,b])$ such that $\alpha \leq \int \phi_{\mu,\alpha}(x)\, d\mu(x)$ and $\int \phi_{\mu,\alpha}(x)\, d\eta(x) \leq W(\eta)$ for all $\eta \in \triangle([a,b])$. Then, letting

$$\Phi = \{\phi_{\mu,\alpha} : \mu \in \triangle([a,b]), \alpha < W(\mu)\},$$

it follows directly that

$$W(\mu) = \sup_{\phi\in\Phi} \int \phi(x)\, d\mu(x).$$

Fix any $\mu \in \triangle([a,b])$ and any $\alpha < W(\mu)$. The proof is completed by showing the existence of a function $\phi_{\mu,\alpha}$ as described above. This is accomplished using a separation argument similar to standard duality results for convex functions (see, e.g., Ekeland and Turnbull (1983) or Phelps (1993)). The epigraph of W is defined as follows:

$$\text{epi}(W) = \{(\eta, t) \in \triangle([a,b]) \times \mathbb{R} : t \geq W(\eta)\}$$

Since W is convex with a convex domain $\triangle([a,b])$, epi(W) is a convex subset of $ca([a,b]) \times \mathbb{R}$. Moreover, as a weak* lower semicontinuous function with a weak* closed domain, it is a standard result that epi(W) is a closed subset of $ca([a,b]) \times \mathbb{R}$.[30] Now, define a set $F_{\mu,\alpha}$ as follows:

$$F_{\mu,\alpha} \equiv (\{\mu\} + F) \times \{\alpha\} = \{\mu + \nu : \nu \in F\} \times \{\alpha\}$$

[30]The set $ca([a,b]) \times \mathbb{R}$ is endowed with the product topology generated by the weak* topology on $ca([a,b])$ and the Euclidean topology on $\mathbb{R}$.

By Lemma 4, $F_{\mu,\alpha}$ is a closed and convex subset of $ca([a,b]) \times \mathbb{R}$. Establishing the following claim allows the separating hyperplane theorem to be applied.[31]

Claim 1 *For $\alpha < W(\mu)$, the set $\text{epi}(W) - F_{\mu,\alpha}$ is convex, and $(0,0) \notin \text{cl}(\text{epi}(W) - F_{\mu,\alpha})$.*

Proof of claim: First, note that $F_{\mu,\alpha} \cap \text{epi}(W) = \emptyset$. To see this, take any $(\eta, t) \in F_{\mu,\alpha}$. Then, clearly $t = \alpha$, and by definition, $\eta - \mu \in F$. If $\eta \notin \triangle([a,b])$, then it is trivial that $(\eta, t) \notin \text{epi}(W)$. Alternatively, if $\eta \in \triangle([a,b])$, then Lemma 4 implies η first order stochastically dominates μ. In this case $W(\eta) \geq W(\mu) > \alpha = t$, so again $(\eta, t) \notin \text{epi}(W)$. Thus, $F_{\mu,\alpha}$ and $\text{epi}(W)$ are disjoint, closed, and convex sets.

Since $F_{\mu,\alpha}$ and $\text{epi}(W)$ are convex and disjoint, $\text{epi}(W) - F_{\mu,\alpha}$ is convex and $(0,0) \notin \text{epi}(W) - F_{\mu,\alpha}$. Since W is weak* lower semicontinuous and has a weak* compact domain $\triangle([a,b])$, it attains a minimum value $\underline{W}$. Therefore, $\text{epi}(W)$ can be written as the union of the following two sets:

$$B_1 \equiv \text{epi}(W) \cap \left[\triangle([a,b]) \times [\underline{W}, W(\mu)]\right] = \{(\eta, t) \in \triangle([a,b]) \times \mathbb{R} : W(\mu) \geq t \geq W(\eta)\}$$
$$B_2 \equiv \text{epi}(W) \cap \left[\triangle([a,b]) \times [W(\mu), +\infty)\right] = \{(\eta, t) \in \triangle([a,b]) \times \mathbb{R} : t \geq \max\{W(\eta), W(\mu)\}\}$$

As the intersection of a closed set and a compact set, B_1 is compact, and as the intersection of two closed sets, B_2 is closed. Since the difference of a compact set and a closed set is closed, $B_1 - F_{\mu,\alpha}$ is closed. Since $B_1 - F_{\mu,\alpha} \subset \text{epi}(W) - F_{\mu,\alpha}$, this set does not contain $(0,0)$. Also note that for every $(\nu, t) \in B_2 - F_{\mu,\alpha}$, it must be that $t \geq W(\mu) - \alpha > 0$. Therefore, $B_2 - F_{\mu,\alpha} \subset ca([a,b]) \times [W(\mu) - \alpha, +\infty)$, a closed set not containing $(0,0)$. Thus, $\text{epi}(W) - F_{\mu,\alpha}$ is contained in the union of the closed sets $B_1 - F_{\mu,\alpha}$ and $ca([a,b]) \times [W(\mu) - \alpha, +\infty)$, each of which does not contain $(0,0)$. ■

Continuing the proof of Proposition 1, note that $ca([a,b]) \times \mathbb{R}$ is a locally convex Hausdorff space (Theorem 5.73 in Aliprantis and Border (2006)). Therefore, the separating hyperplane theorem (Theorem 5.79 in Aliprantis and Border (2006)) implies there exists a weak* continuous linear functional $f : ca([a,b]) \to \mathbb{R}$ and $\lambda \in \mathbb{R}$ such that

$$f(\nu) + \lambda t < f(0) + \lambda 0 = 0, \quad \forall (\nu, t) \in \text{epi}(W) - F_{\mu,\alpha}.$$

Since $(\eta - \mu - \nu, t - \alpha) \in \text{epi}(W) - F_{\mu,\alpha}$ for any $(\eta, t) \in \text{epi}(W)$ and $\nu \in F$, this implies

$$f(\eta) + \lambda t < f(\mu) + f(\nu) + \lambda \alpha, \quad \forall (\eta, t) \in \text{epi}(W),\ \forall \nu \in F. \tag{26}$$

Taking $(\eta, t) = (\mu, W(\mu))$ and $\nu = 0$, it follows that $\lambda W(\mu) < \lambda \alpha$. Since $\alpha < W(\mu)$, this implies $\lambda < 0$. Therefore, setting $\nu = 0$ in the Equation (26), conclude that for all $\eta \in \triangle([a,b])$,

$$f(\eta) + \lambda W(\eta) < f(\mu) + \lambda \alpha \implies W(\eta) > -\tfrac{f(\eta)}{\lambda} + \tfrac{f(\mu)}{\lambda} + \alpha$$

[31]Although $\text{epi}(W)$ and $F_{\mu,\alpha}$ are disjoint, closed, and convex sets, standard separation theorems require that at least one of the sets either be compact or have a nonempty interior. Therefore, a slightly more involved argument is required here.

Consider the weak* continuous linear functional $\eta \mapsto -\frac{f(\eta)}{\lambda}$ defined on $ca([a,b])$. Since the weak* topology on $ca([a,b])$ is generated by $C([a,b])$, every weak* continuous linear functional on $ca([a,b])$ corresponds to some $\psi \in C([a,b])$ (Theorem 5.93 in Aliprantis and Border (2006)). In particular, there exists $\psi \in C([a,b])$ such that $-\frac{f(\eta)}{\lambda} = \int \psi(x)\, d\eta(x)$ for all $\eta \in ca([a,b])$. Define $\phi_{\mu,\alpha} \in C([a,b])$ by $\phi_{\mu,\alpha}(x) = \psi(x) + \frac{f(\mu)}{\lambda} + \alpha$ for $x \in [a,b]$. Then, for every $\eta \in \triangle([a,b])$,

$$W(\eta) > \int \psi(x)\, d\eta(x) + \tfrac{f(\mu)}{\lambda} + \alpha = \int \left[\psi(x) + \tfrac{f(\mu)}{\lambda} + \alpha\right] d\eta(x) = \int \phi_{\mu,\alpha}(x)\, d\eta(x).$$

In addition,

$$\int \phi_{\mu,\alpha}(x)\, d\mu(x) = -\tfrac{f(\mu)}{\lambda} + \tfrac{f(\mu)}{\lambda} + \alpha = \alpha.$$

The proof is completed by showing that $\phi_{\mu,\alpha}$ is nondecreasing. Fix any $x, y \in [a,b]$ with $x < y$. Note that $r(\delta_y - \delta_x) \in F$ for all $r > 0$. Therefore, Equation (26) implies

$$\begin{aligned} f(\mu) + \lambda W(\mu) &< f(\mu) + f(r(\delta_y - \delta_x)) + \lambda\alpha \\ &= f(\mu) + r[f(\delta_y) - f(\delta_x)] + \lambda\alpha \end{aligned}$$

For this to be true for every $r > 0$, it must be that $f(\delta_y) \geq f(\delta_x)$. This implies $-\frac{f(\delta_y)}{\lambda} \geq -\frac{f(\delta_x)}{\lambda}$, which in turn implies $\phi_{\mu,\alpha}(y) \geq \phi_{\mu,\alpha}(x)$. ■

B.2 Proof of Theorem 2

Lemma 5 *The relation $\succsim$ satisfies weak order, nontriviality, continuity, stationarity, and separability (Axiom 1–5) if and only if there exist continuous and nonconstant functions $u_1 : C \to \mathbb{R}$ and $u_2 : \triangle(D) \to \mathbb{R}$ and a scalar $\beta \in (0,1)$ such that the following hold:*

1. *The function $V : D \to \mathbb{R}$ defined by $V(c,m) = u_1(c) + u_2(m)$ represents $\succsim$.*
2. *For every $(\bar{c}, \bar{m}) \in D$, $u_2(\delta_{(\bar{c},\bar{m})}) = \beta(u_1(\bar{c}) + u_2(\bar{m}))$.*

Proof: The necessity of weak order, nontriviality, and continuity are immediate. It follows from condition 2 that for any $c, \bar{c} \in C$ and $\bar{m} \in \triangle(D)$,

$$V(c, \delta_{(\bar{c},\bar{m})}) = u_1(c) + \beta u_1(\bar{c}) + \beta u_2(\bar{m}) = u_1(c) + \beta V(\bar{c}, \bar{m}).$$

The necessity of stationarity and separability follow directly from this expression.

For sufficiency, the first step is obtain an additively separable representation on a restricted domain. Note that in addition to the separability conditions listed in Axiom 5, stationarity (Axiom 4) implies that $(c, \delta_{(\bar{c},\bar{m})}) \succsim (c, \delta_{(\bar{c}',\bar{m}')})$ if and only if $(c', \delta_{(\bar{c},\bar{m})}) \succsim (c', \delta_{(\bar{c}',\bar{m}')})$. Therefore, the assumed axioms are sufficient to apply Theorem 3 of Debreu (1960) to obtain continuous functions $f : C \to \mathbb{R}$, $g : C \to \mathbb{R}$, and $h : \triangle(D) \to \mathbb{R}$ such that

$$(c, \delta_{(\bar{c},\bar{m})}) \succsim (c', \delta_{(\bar{c}',\bar{m}')}) \iff f(c) + g(\bar{c}) + h(\bar{m}) \geq f(c') + g(\bar{c}') + h(\bar{m}'). \tag{27}$$

Note that the previous equation only gives a partial representation for $\succsim$. However, by stationarity,

$$\begin{aligned}(\bar{c}, \bar{m}) \succsim (\bar{c}', \bar{m}') &\iff (c, \delta_{(\bar{c},\bar{m})}) \succsim (c, \delta_{(\bar{c}',\bar{m}')}) \\ &\iff g(\bar{c}) + h(\bar{m}) \geq g(\bar{c}') + h(\bar{m}'),\end{aligned} \tag{28}$$

and hence g and h give an additive representation for $\succsim$. In particular, the combination of Equations (27) and (28) implies

$$\begin{aligned}&g(c) + h(\delta_{(\bar{c},\bar{m})}) \geq g(c') + h(\delta_{(\bar{c}',\bar{m}')}) \\ &\iff f(c) + [g(\bar{c}) + h(\bar{m})] \geq f(c') + [g(\bar{c}') + h(\bar{m}')]\end{aligned}$$

Using the uniqueness of additively separable representations (see Debreu (1960) or Theorem 5.4 in Fishburn (1970)), the above implies there exists $\beta > 0$ and $\alpha_1, \alpha_2 \in \mathbb{R}$ such that:

$$\begin{aligned}g(c) &= \beta f(c) + \alpha_1, &&\forall c \in C \\ h(\delta_{(\bar{c},\bar{m})}) &= \beta[g(\bar{c}) + h(\bar{m})] + \alpha_2, &&\forall (\bar{c}, \bar{m}) \in D\end{aligned} \tag{29}$$

Define $u_1 : C \to \mathbb{R}$ and $u_2 : \triangle(D) \to \mathbb{R}$ by $u_1(c) = g(c) + \frac{\alpha_2}{\beta}$ and $u_2(m) = h(m)$. Then, claims 1 and 2 follow directly from Equations (28) and (29).

It remains only to show that $\beta < 1$. Following a similar approach to Gul and Pesendorfer (2004), this can be established using continuity. By nontriviality, there exist $c^*, c_* \in C$ such that $u_1(c^*) > u_1(c_*)$. Fix any $m \in \triangle(D)$ and, with slight abuse of notation, define sequences $\{d_n\}$ and $\{d'_n\}$ in D as follows:[32]

$$d_n = (\underbrace{c^*, \ldots, c^*}_{n}, m) \qquad \text{and} \qquad d'_n = (\underbrace{c_*, \ldots, c_*}_{n}, m)$$

By the compactness of D, there exists $\{n_k\}$ such that the subsequences $\{d_{n_k}\}$ and $\{d'_{n_k}\}$ converge to some d and d' in D, respectively. By continuity, $V(d_{n_k}) \to V(d)$ and $V(d'_{n_k}) \to V(d')$, where V is defined as in condition 1. Therefore, the difference $V(d_{n_k}) - V(d'_{n_k})$ converges to some real number. However, since u_1 and u_2 were shown to satisfy condition 2,

$$\begin{aligned}V(d_{n_k}) - V(d'_{n_k}) &= \left(\sum_{i=0}^{n_k-1} \beta^i u_1(c^*) + \beta^{(n_k-1)} u_2(m)\right) - \left(\sum_{i=0}^{n_k-1} \beta^i u_1(c_*) + \beta^{(n_k-1)} u_2(m)\right) \\ &= \sum_{i=0}^{n_k-1} \beta^i [u_1(c^*) - u_1(c_*)].\end{aligned}$$

[32]More precisely, $d_1 = (c^*, m)$, $d_2 = (c^*, \delta_{(c^*,m)})$, and so on.

Since this difference converges to a real number, it must be that $\beta < 1$. ■

Lemma 6 *Suppose $\succsim$ is represented by $V(c,m) = u_1(c) + u_2(m)$ where $u_1 : C \to \mathbb{R}$ and $u_2 : \triangle(D) \to \mathbb{R}$ are continuous*[33]*, and $\succsim$ satisfies stationarity. Then, $\succsim$ satisfies FOSD (Axiom 6) if and only if for any $m, m' \in \triangle(D)$,*

$$\begin{aligned} m \circ V^{-1}(\{v : v \geq \bar{v}\}) \geq m' \circ V^{-1}(\{v : v \geq \bar{v}\}),\ \forall \bar{v} \in V(D) \\ \implies u_2(m) \geq u_2(m') \end{aligned} \tag{30}$$

Proof: To see that FOSD implies Equation (30), consider any two measures $m, m' \in \triangle(D)$ such that

$$m \circ V^{-1}(\{v : v \geq \bar{v}\}) \geq m' \circ V^{-1}(\{v : v \geq \bar{v}\}), \qquad \forall \bar{v} \in V(D).$$

Fix any $(\bar{c}', \bar{m}') \in D$, and let $\bar{v} = V(\bar{c}', \bar{m}')$. By stationarity, $(c, \delta_{(\bar{c},\bar{m})}) \succsim (c, \delta_{(\bar{c}',\bar{m}')})$ if and only if $V(\bar{c}, \bar{m}) \geq V(\bar{c}', \bar{m}') = \bar{v}$. Therefore,

$$\begin{aligned} m(\{(\bar{c},\bar{m}) : (c, \delta_{(\bar{c},\bar{m})}) \succsim (c, \delta_{(\bar{c}',\bar{m}')})\}) &= m \circ V^{-1}(\{v : v \geq \bar{v}\}) \\ &\geq m' \circ V^{-1}(\{v : v \geq \bar{v}\}) \\ &= m'(\{(\bar{c},\bar{m}) : (c, \delta_{(\bar{c},\bar{m})}) \succsim (c, \delta_{(\bar{c}',\bar{m}')})\}). \end{aligned}$$

Since this condition holds for all $(\bar{c}', \bar{m}') \in D$, FOSD implies $(c, m) \succsim (c, m')$. Thus, $u_2(m) \geq u_2(m')$. The argument that Equation (30) implies the FOSD axiom is similar. ■

Lemma 7 *Suppose $\succsim$ is represented by $V(c,m) = u_1(c) + u_2(m)$ where $u_1 : C \to \mathbb{R}$ and $u_2 : \triangle(D) \to \mathbb{R}$ are nonconstant and continuous. Then, $\succsim$ satisfies convexity (Axiom 7) if and only if u_2 is convex.*

Proof: To see the necessity of the convexity axiom, suppose u_2 is convex and $u_1(c') + u_2(m) \leq u_1(c) + u_2\left(\frac{1}{2}m + \frac{1}{2}m'\right)$. Then,

$$u_1(c') - u_1(c) \leq u_2\left(\tfrac{1}{2}m + \tfrac{1}{2}m'\right) - u_2(m) \leq u_2(m') - u_2\left(\tfrac{1}{2}m + \tfrac{1}{2}m'\right),$$

where the last inequality follows from the convexity of u_2. Hence, $u_1(c') + u_2\left(\frac{1}{2}m + \frac{1}{2}m'\right) \leq u_1(c) + u_2(m')$.

To show sufficiency, suppose that $\succsim$ satisfies convexity. Since u_1 is nonconstant, fix $c^*, c_* \in C$ such that $u_1(c^*) > u_1(c_*)$. First, consider any $m, m' \in \triangle(D)$ such that $|u_2(m) - u_2\left(\frac{1}{2}m + \frac{1}{2}m'\right)| \leq u_1(c^*) - u_1(c_*)$. Then, since C is connected and u_1 is continuous, there exist $c, c' \in C$ such that

$$u_1(c') - u_1(c) = u_2\left(\tfrac{1}{2}m + \tfrac{1}{2}m'\right) - u_2(m).$$

[33] Continuity is not necessary for this result; measurability of u_1 and u_2 are sufficient.

This implies $(c', m) \sim (c, \frac{1}{2}m+\frac{1}{2}m')$, and hence $(c, m') \succsim (c', \frac{1}{2}m+\frac{1}{2}m')$ by the convexity axiom. Therefore,

$$\begin{aligned} & u_2(m') - u_2\left(\tfrac{1}{2}m + \tfrac{1}{2}m'\right) \geq u_1(c') - u_1(c) = u_2\left(\tfrac{1}{2}m + \tfrac{1}{2}m'\right) - u_2(m) \\ \implies & \tfrac{1}{2}u_2(m) + \tfrac{1}{2}u_2(m') \geq u_2\left(\tfrac{1}{2}m + \tfrac{1}{2}m'\right). \end{aligned}$$

Now, take any $m, m' \in \triangle(D)$. Define a function $\psi : [0,1] \to \mathbb{R}$ by $\psi(\alpha) = u_2(\alpha m+(1-\alpha)m')$. This function is continuous by the weak* continuity of v, and since its domain is compact, ψ is therefore uniformly continuous. Thus, there exists $\delta > 0$ such that $|\alpha - \alpha'| \leq \delta$ implies $|\psi(\alpha) - \psi(\alpha')| \leq u_1(c^*) - u_1(c_*)$. By the preceding arguments, this implies ψ is *midpoint convex* on any interval $[\bar{\alpha}, \bar{\alpha}'] \subset [0,1]$ with $|\bar{\alpha} - \bar{\alpha}'| \leq \delta$, that is, for any $\alpha, \alpha' \in [\bar{\alpha}, \bar{\alpha}']$,

$$\tfrac{1}{2}\psi(\alpha) + \tfrac{1}{2}\psi(\alpha') \geq \psi\left(\tfrac{1}{2}\alpha + \tfrac{1}{2}\alpha'\right).$$

It is a stand result that any continuous and midpoint convex function is convex. Thus, ψ is convex on any interval $[\bar{\alpha}, \bar{\alpha}'] \subset [0,1]$ with $|\bar{\alpha} - \bar{\alpha}'| \leq \delta$. This, in turn, is sufficient to ensure that ψ is convex on $[0,1]$. Therefore, for any $\alpha \in [0,1]$,

$$\alpha u_2(m) + (1-\alpha)u_2(m') = \alpha\psi(1) + (1-\alpha)\psi(0) \geq \psi(\alpha) = u_2(\alpha m + (1-\alpha)m').$$

Since m and m' where arbitrary, u_2 is convex. ■

Lemma 8 *Suppose V, u_1, u_2, and β are as in Lemma 5, and suppose V and u_2 satisfy Equation (30). Then, there exists a function $W : \triangle(V(D)) \to \mathbb{R}$ such that $u_2(m) = \beta W(m \circ V^{-1})$ for all $m \in \triangle(D)$. Moreover,*

1. *$W(\delta_v) = v$ for all $v \in V(D)$.*
2. *W is weak* continuous and monotone with respect to FOSD.*
3. *If u_2 is convex, then W is convex.*

Proof: *Proof of existence of W:* First, note that since V is continuous and $V(D)$ is compact, any Borel probability measure on $V(D)$ can be written as $m \circ V^{-1}$ for some $m \in \triangle(D)$ (Part 5 of Theorem 15.14 in Aliprantis and Border (2006)), that is,

$$\{m \circ V^{-1} : m \in \triangle(D)\} = \triangle(V(D)).$$

Fix any $\mu \in \triangle(V(D))$. Let $W(\mu) = \frac{1}{\beta}u_2(m)$ for any $m \in \triangle(D)$ such that $\mu = m \circ V^{-1}$. There exists at least one such m by the preceding arguments. In addition, if $\mu = m \circ V^{-1} = m' \circ V^{-1}$ for $m, m' \in \triangle(D)$, then $u_2(m) = u_2(m')$ by Equation (30). Thus, W is well defined, and by construction, $u_2(m) = \beta W(m \circ V^{-1})$ for all $m \in \triangle(D)$.

Proof of 1: By condition 2 in Lemma 5, $u_2(\delta_{(\bar{c},\bar{m})}) = \beta V(\bar{c},\bar{m})$ for every $(\bar{c},\bar{m}) \in D$, and hence

$$W(\delta_{V(\bar{c},\bar{m})}) = W(\delta_{(\bar{c},\bar{m})} \circ V^{-1}) = \tfrac{1}{\beta} u_2(\delta_{(\bar{c},\bar{m})}) = V(\bar{c},\bar{m}).$$

Proof of 2: To see that W is weak* continuous, take any sequence $\{\mu_n\}$ in $\triangle(V(D))$ that converges to some $\mu \in \triangle(V(D))$. It suffices to show that there exists a subsequence $\{\mu_{n_k}\}$ such that $W(\mu_{n_k}) \to W(\mu)$.[34] For each n, take any $m_n \in \triangle(D)$ such that $\mu_n = m_n \circ V^{-1}$. Since $\triangle(D)$ is compact and metrizable, there is a subsequence $\{m_{n_k}\}$ converging to some $m \in \triangle(D)$. By the continuity of V,

$$m_{n_k} \xrightarrow{w^*} m \implies \mu_{n_k} = m_{n_k} \circ V^{-1} \xrightarrow{w^*} m \circ V^{-1}.$$

This implication follows directly from the definition of weak* convergence, or see Part 1 of Theorem 15.14 in Aliprantis and Border (2006). Thus, $\mu = m \circ V^{-1}$. Since u_2 is weak* continuous,

$$W(\mu_{n_k}) = \tfrac{1}{\beta} u_2(m_{n_k}) \to \tfrac{1}{\beta} u_2(m) = W(\mu).$$

Therefore, W is weak* continuous. To see that W is monotone with respect to FOSD, suppose $\mu, \eta \in \triangle(V(D))$ satisfy $\mu(\{v : v \geq \bar{v}\}) \geq \eta(\{v : v \geq \bar{v}\})$ for all $\bar{v} \in V(D)$. Take any $m, m' \in \triangle(D)$ such that $\mu = m \circ V^{-1}$ and $\eta = m' \circ V^{-1}$. Then, Equation (30) implies $u_2(m) \geq u_2(m')$, and hence $W(\mu) \geq W(\eta)$.

Proof of 3: Suppose u_2 is convex. Fix any $\mu, \eta \in \triangle(V(D))$ and $\alpha \in (0,1)$. Take any $m, m' \in \triangle(D)$ such that $\mu = m \circ V^{-1}$ and $\eta = m' \circ V^{-1}$. Then,

$$\alpha\mu + (1-\alpha)\eta = (\alpha m + (1-\alpha)m') \circ V^{-1},$$

and hence

$$\begin{aligned} W(\alpha\mu + (1-\alpha)\eta) &= \tfrac{1}{\beta} u_2(\alpha m + (1-\alpha)m') \\ &\leq \alpha \tfrac{1}{\beta} u_2(m) + (1-\alpha)\tfrac{1}{\beta} u_2(m') = \alpha W(\mu) + (1-\alpha)W(\eta), \end{aligned}$$

establishing the convexity of W. ■

Proof of Theorem 2: The necessity of the axioms is straightforward. To establish sufficiency, suppose $\succsim$ satisfies Axioms 1–7. By Lemmas 5, 6, 7, and 8, there exists a scalar $\beta \in (0,1)$, a continuous and nonconstant function $u : C \to \mathbb{R}$, and a weak* continuous and convex function $W : \triangle(V(D)) \to \mathbb{R}$ such that

$$V(c,m) = u(c) + \beta W(m \circ V^{-1}), \qquad \forall (c,m) \in D.$$

[34] If W is not continuous at a point μ, there exists $\varepsilon > 0$ and a sequence $\{\mu_n\}$ converging to μ such that $|W(\mu_n) - W(\mu)| > \varepsilon$ for every n. This sequence has no subsequence with the convergence properties described above.

Moreover, $W(\delta_v) = v$ for all $v \in V(D)$, and W is monotone with respect to FOSD. Since V is continuous and D is compact and connected, $V(D) = [a, b]$ for some $a, b \in \mathbb{R}$. Apply Proposition 1 to conclude there exists a collection Φ of continuous and nondecreasing functions $\phi : [a, b] \to \mathbb{R}$ such that

$$W(\mu) = \sup_{\phi \in \Phi} \int \phi(v)\, d\mu(v).$$

Using the change of variables formula, for every $(c, m) \in D$,

$$\begin{aligned} V(c, m) &= u(c) + \beta W(m \circ V^{-1}) \\ &= u(c) + \beta \sup_{\phi \in \Phi} \int_a^b \phi(v)\, d(m \circ V^{-1})(v) \\ &= u(c) + \beta \sup_{\phi \in \Phi} \int_D \phi(V(\bar{c}, \bar{m}))\, dm(\bar{c}, \bar{m}). \end{aligned}$$

In addition,

$$\sup_{\phi \in \Phi} \phi(\bar{v}) = \sup_{\phi \in \Phi} \int \phi(v)\, d\delta_{\bar{v}}(v) = W(\delta_{\bar{v}}) = \bar{v}$$

for all $\bar{v} \in [a, b]$. ■

B.3 Proof of Theorem 6

The first step in the proof is to show that in Equation (17) it is possible to restrict attention to $\gamma \in [a, b]$ instead of considering all $\gamma \in I$. This is accomplished by establishing that for $v \in [a, b]$, $\phi(v|\gamma) \leq \phi(v|a)$ for any $\gamma < a$ and $\phi(v|\gamma) \leq \phi(v|b)$ for any $\gamma > b$. The following lemma establishes that $\phi(\cdot|\cdot)$ is supermodular, which will then be used to prove the desired inequalities.

Lemma 9 *Suppose I and $\phi(\cdot|\cdot)$ satisfy the assumptions of Theorem 6 (either 1, 2, or 3). Then, $\phi(\cdot|\cdot)$ is supermodular: for any $\gamma, \hat{\gamma}, v, \hat{v} \in I$ with $\hat{\gamma} > \gamma$ and $\hat{v} > v$,*

$$\phi(\hat{v}|\hat{\gamma}) + \phi(v|\gamma) \geq \phi(v|\hat{\gamma}) + \phi(\hat{v}|\gamma). \tag{31}$$

Proof: Consider first the case where the first set of assumptions is satisfied, that is, $I = \mathbb{R}_+$, $\phi(\cdot|\cdot)$ satisfies Assumption 1 and is homogeneous, and $\phi(v|0) = 0$. Fix any $\gamma, \hat{\gamma}, v, \hat{v} \in I$ with $\hat{\gamma} > \gamma$ and $\hat{v} > v$. If $\gamma = 0$, then Equation (31) simplifies to $\phi(\hat{v}|\hat{\gamma}) \geq \phi(v|\hat{\gamma})$, which holds since $\phi(\cdot|\hat{\gamma})$ is nondecreasing by Assumption 1. For $\hat{\gamma} > \gamma > 0$, homogeneity implies Equation (31) is equivalent to the following:

$$\hat{\gamma}\phi\Big(\frac{\hat{v}}{\hat{\gamma}}\Big|1\Big) + \gamma\phi\Big(\frac{v}{\gamma}\Big|1\Big) \geq \hat{\gamma}\phi\Big(\frac{v}{\hat{\gamma}}\Big|1\Big) + \gamma\phi\Big(\frac{\hat{v}}{\gamma}\Big|1\Big). \tag{32}$$

To see this holds, note first that $\frac{\hat{v}}{\hat{\gamma}}$ and $\frac{v}{\gamma}$ can both be expressed as convex combinations of $\frac{v}{\hat{\gamma}}$ and $\frac{\hat{v}}{\gamma}$:

$$\frac{\hat{v}}{\hat{\gamma}} = \Big(\frac{\hat{\gamma}\hat{v} - \gamma\hat{v}}{\hat{\gamma}\hat{v} - \gamma v}\Big)\Big(\frac{v}{\hat{\gamma}}\Big) + \Big(\frac{\gamma\hat{v} - \gamma v}{\hat{\gamma}\hat{v} - \gamma v}\Big)\Big(\frac{\hat{v}}{\gamma}\Big)$$
$$\frac{v}{\gamma} = \Big(\frac{\hat{\gamma}\hat{v} - \hat{\gamma}v}{\hat{\gamma}\hat{v} - \gamma v}\Big)\Big(\frac{v}{\hat{\gamma}}\Big) + \Big(\frac{\hat{\gamma}v - \gamma v}{\hat{\gamma}\hat{v} - \gamma v}\Big)\Big(\frac{\hat{v}}{\gamma}\Big).$$

Therefore, the concavity of $\phi(\cdot|1)$ implies

$$\hat{\gamma}\phi\Big(\frac{\hat{v}}{\hat{\gamma}}\Big|1\Big) \geq \hat{\gamma}\left[\Big(\frac{\hat{\gamma}\hat{v} - \gamma\hat{v}}{\hat{\gamma}\hat{v} - \gamma v}\Big)\phi\Big(\frac{v}{\hat{\gamma}}\Big|1\Big) + \Big(\frac{\gamma\hat{v} - \gamma v}{\hat{\gamma}\hat{v} - \gamma v}\Big)\phi\Big(\frac{\hat{v}}{\gamma}\Big|1\Big)\right]$$
$$\gamma\phi\Big(\frac{v}{\gamma}\Big|1\Big) \geq \gamma\left[\Big(\frac{\hat{\gamma}\hat{v} - \hat{\gamma}v}{\hat{\gamma}\hat{v} - \gamma v}\Big)\phi\Big(\frac{v}{\hat{\gamma}}\Big|1\Big) + \Big(\frac{\hat{\gamma}v - \gamma v}{\hat{\gamma}\hat{v} - \gamma v}\Big)\phi\Big(\frac{\hat{v}}{\gamma}\Big|1\Big)\right].$$

Summing these two inequalities yields Equation (32).

The proofs in the cases of the second and third sets of assumptions are similar. For homogeneous $\phi(\cdot|\cdot)$ on $I = (-\infty, 0)$, the arguments are almost identical to those above, but use the fact that $\phi(v|\gamma) = -\gamma\phi(\frac{v}{-\gamma}| - 1)$. For shift invariant $\phi(\cdot|\cdot)$ on $I = \mathbb{R}$, the claim can be established by noting the Equation (31) is equivalent to

$$\phi(\hat{v} - \hat{\gamma}|0) + \phi(v - \gamma|0) \geq \phi(v - \hat{\gamma}|0) + \phi(\hat{v} - \gamma|0),$$

then expressing $\hat{v}-\hat{\gamma}$ and $v-\gamma$ as convex combinations of $v-\hat{\gamma}$ and $\hat{v}-\gamma$ and using concavity. ■

By Lemma 9, for any $v \in [a, b]$ and $\gamma \in I$ such that $a > \gamma$, $\phi(v|a)+\phi(a|\gamma) \geq \phi(a|a)+\phi(v|\gamma)$. Since $\phi(a|a) = a \geq \phi(a|\gamma)$ by Assumption 1, this implies that $\phi(v|a) \geq \phi(v|\gamma)$, as desired. Likewise, for any $v \in [a, b]$ and $\gamma \in I$ such that $\gamma > b$, $\phi(b|\gamma) + \phi(v|b) \geq \phi(v|\gamma) + \phi(b|b)$. Since $\phi(b|b) = b \geq \phi(b|\gamma)$, this implies $\phi(v|b) \geq \phi(v|\gamma)$. Thus, it is possible to restrict attention $\gamma \in [a, b]$:

$$\sup_{\gamma\in I}\int \phi(v|\gamma)\, d\mu(v) = \sup_{\gamma\in[a,b]}\int \phi(v|\gamma)\, d\mu(v).$$

Next, to see that a maximizer exists in the interval $[a, b]$, begin by showing some continuity properties of $\phi(\cdot|\cdot)$.

Lemma 10 *Suppose I, $\phi(\cdot|\cdot)$, and $a, b \in I$ satisfy the assumptions of Theorem 6 (either 1, 2, or 3). Fix any sequences $\{v_n\}$ and $\{\gamma_n\}$ in $[a, b]$. Then,*

$$\big[v_n \to v, \gamma_n \to \gamma\big] \implies \phi(v_n|\gamma_n) \to \phi(v|\gamma). \tag{33}$$

Consequently, $\gamma_n \to \gamma$ implies $\phi(\cdot|\gamma_n) \to \phi(\cdot|\gamma)$ uniformly on $[a, b]$.

Proof: Consider the case where the first set of assumptions is satisfied. By homogeneity and the real-valued assumption, for any $\gamma \in [a, b]$ and v in a neighborhood of $[a, b]$, $\gamma\phi(\frac{v}{\gamma}|1) =$

$\phi(v|\gamma) > -\infty$. Since $\phi(\cdot|1)$ is concave by Assumption 1, this implies $\phi(\cdot|1)$ is continuous at each $\frac{v}{\gamma}$ for $v, \gamma \in [a,b]$. Therefore, if $\gamma_n \to \gamma \in [a,b]$ and $v_n \to v \in [a,b]$, then

$$\phi(v_n|\gamma_n) = \gamma_n \phi\Big(\frac{v_n}{\gamma_n}\Big|1\Big) \to \gamma \phi\Big(\frac{v}{\gamma}\Big|1\Big) = \phi(v|\gamma),$$

proving Equation (33). The proofs in the cases of the second and third sets of assumptions are similar.

It is a standard result that Equation (33) implies the asserted uniform convergence property. For completeness, this property is established as follows: Suppose for a contradiction that $\phi(\cdot|\gamma_n)$ does not converge to $\phi(\cdot|\gamma)$ uniformly. Then, there exists $\varepsilon > 0$, a subsequence $\{\gamma_{k_n}\}$, and a sequence $\{v_n\} \subset [a,b]$ such that $|\phi(v_n|\gamma_{k_n}) - \phi(v_n|\gamma)| \geq \varepsilon$ for all n. Since $[a,b]$ is compact, it is without loss to assume that $v_n \to v$ for some $v \in [a,b]$. But then

$$\varepsilon \leq \big|\phi(v_n|\gamma_{k_n}) - \phi(v_n|\gamma)\big| \leq \big|\phi(v_n|\gamma_{k_n}) - \phi(v|\gamma)\big| + \big|\phi(v|\gamma) - \phi(v_n|\gamma)\big|.$$

Since both of the terms on the right converge to 0 by Equation (33), this is a contradiction. ■

Now, define $f : [a,b] \times \triangle([a,b]) \to \mathbb{R}$ by

$$f(\gamma,\mu) = \int \phi(v|\gamma)\, d\mu(v).$$

To see that f is continuous, suppose $\gamma_n \to \gamma$ and $\mu_n \xrightarrow{w^*} \mu$. Fix any $\varepsilon > 0$. Since $\phi(\cdot|\gamma)$ is continuous by Equation (33), the definition of weak convergence implies there exists N such that $n \geq N$ implies $|f(\gamma,\mu_n) - f(\gamma,\mu)| < \varepsilon$. In addition, since $\phi(\cdot|\gamma_n) \to \phi(\cdot|\gamma)$ uniformly by Lemma 10, there exists N' such that $n \geq N'$ implies $|\phi(v|\gamma_n) - \phi(v|\gamma)| < \varepsilon$ for all $v \in [a,b]$. Consequently, for any $\eta \in \triangle([a,b])$, $n \geq N'$ implies

$$|f(\gamma_n,\eta) - f(\gamma,\eta)| \leq \int |\phi(v|\gamma_n) - \phi(v|\gamma)|\, d\eta(v) \leq \varepsilon.$$

Thus, $n \geq \max\{N, N'\}$ implies

$$|f(\gamma_n,\mu_n) - f(\gamma,\mu)| \leq |f(\gamma_n,\mu_n) - f(\gamma,\mu_n)| + |f(\gamma,\mu_n) - f(\gamma,\mu)| < 2\varepsilon.$$

Hence, f is continuous.

Since $[a,b]$ is compact, continuity of f implies that for every $\mu \in \triangle([a,b])$, there exists $\hat{\gamma} \in [a,b]$ such that $f(\hat{\gamma},\mu) = \sup_{\gamma\in[a,b]} f(\gamma,\mu)$, that is,

$$\int \phi(v|\hat{\gamma})\, d\mu(v) = \sup_{\gamma\in[a,b]} \int \phi(v|\gamma)\, d\mu(v) = \sup_{\gamma\in I} \int \phi(v|\gamma)\, d\mu(v).$$

Moreover, by Berge's Maximum theorem (see, e.g., Theorem 17.31 in Aliprantis and Border (2006)), the function $W : \triangle([a, b]) \to [a, b]$ defined by

$$W(\mu) = \max_{\gamma\in[a,b]} f(\gamma, \mu) = \max_{\gamma\in I} \int \phi(v|\gamma)\, d\mu(v)$$

is continuous.

B.4 Proof of Theorem 7

Let $B(Z)_+$ denote the set of all bounded and nonnegative functions from Z into $\mathbb{R}$. The following lemma establishes the existence of a value function for unit wealth.

Lemma 11 *Suppose ρ, β, r, and $\phi(\cdot|\cdot)$ are as in the statement of Theorem 7. Then, there exists a unique function $L \in B(Z)_+$ such that*

$$L(z) = \sup_{\substack{c\in[0,1]\\ w\in\triangle(K)}} \left[\frac{c^\rho}{\rho} + (1-c)^\rho \beta \sup_{\gamma\geq 0} \int_Z \phi\Big((w\cdot r(z'))^\rho L(z') \Big| \gamma \Big)\, dP(z'|z) \right]. \tag{34}$$

Proof: Define $T : B(Z)_+ \to \mathbb{R}^Z$ by

$$Tf(z) = \sup_{\substack{c\in[0,1]\\ w\in\triangle(K)}} \left[\frac{c^\rho}{\rho} + (1-c)^\rho \beta \sup_{\gamma\geq 0} \int_Z \phi\Big((w\cdot r(z'))^\rho f(z') \Big| \gamma \Big)\, dP(z'|z) \right].$$

The proof consists of showing that T has a unique fixed point L.

By the assumptions on r and the finiteness of Z, there exists $\bar{r} > 0$ such that $\beta\bar{r}^\rho < 1$ and $\int r_k(z')\, dP(z'|z) \leq \bar{r}$ for all $z \in Z$ and $1 \leq k \leq K$. Fix any $b > 0$. Recall that $\phi(v|\gamma) \leq v$ for all $v \geq 0$, and hence for any $f \in B(Z)_+$ with $f \leq b$,

$$\phi\Big((w\cdot r(z'))^\rho f(z') \Big| \gamma \Big) \leq (w\cdot r(z'))^\rho f(z') \leq b(w\cdot r(z'))^\rho.$$

Together with Jensen's inequality, this implies that for any $f \in B(Z)_+$ with $f \leq b$,

$$\begin{aligned} \sup_{\substack{w\in\triangle(K)\\ \gamma\geq 0}} \int_Z \phi\Big((w\cdot r(z'))^\rho f(z') \Big| \gamma \Big)\, dP(z'|z) &\leq b \sup_{w\in\triangle(K)} \int_Z (w\cdot r(z'))^\rho\, dP(z'|z) \\ &\leq b \sup_{w\in\triangle(K)} \left(\int_Z w\cdot r(z')\, dP(z'|z) \right)^\rho \leq b\bar{r}^\rho. \end{aligned} \tag{35}$$

Now, fix any $b \geq \frac{1}{\rho(1-\beta\bar{r}^\rho)}$ and let $A = \{f \in B(Z) : 0 \leq f \leq b\}$. Then, $\frac{1}{\rho} + \beta b\bar{r}^\rho \leq b$, so Equation (35) implies that for $f \in A$,

$$Tf(z) \leq \sup_{c\in[0,1]} \left(\frac{c^\rho}{\rho} + (1-c)^\rho \beta b\bar{r}^\rho\right) \leq \frac{1}{\rho} + \beta b\bar{r}^\rho \leq b.$$

Thus, $Tf \in A$ for all $f \in A$. To show that T has a unique fixed point in A, it suffices to verify that T satisfies the conditions of Theorem 9 on this set A. Since each $\phi(\cdot|\gamma)$ is nondecreasing, monotonicity of T is immediate. To verify the second condition of Theorem 9, first let $\bar{c} \in (0,1)$ denote the unique solution to $c^{\rho-1} = \rho(1-c)^{\rho-1}\beta b\bar{r}^\rho$, that is, $\bar{c} = [1 + (\rho\beta b\bar{r}^\rho)^{1/(1-\rho)}]^{-1}$. Note that for any $f \in A$, we can restrict attention to $c \in [\bar{c}, 1]$:

$$Tf(z) = \sup_{\substack{c\in[\bar{c},1]\\ w\in\triangle(K)}} \left[\frac{c^\rho}{\rho} + (1-c)^\rho \beta \sup_{\gamma\geq 0} \int_Z \phi\Big((w\cdot r(z'))^\rho f(z')\Big|\gamma\Big)\, dP(z'|z)\right].$$

This equality holds because the supremum over $c \in [0,1]$ is attained by a unique maximizing c, given by

$$c^{\rho-1} = \rho(1-c)^{\rho-1}\beta \sup_{\substack{w\in\triangle(K)\\ \gamma\geq 0}} \int_Z \phi\Big((w\cdot r(z'))^\rho f(z')\Big|\gamma\Big)\, dP(z'|z).$$

It therefore follows from Equation (35) that this maximizing c is weaker larger than $\bar{c}$. Let $\delta = \frac{\bar{c}^\rho}{\rho} > 0$. Then, for any $f \in A$ and $\alpha \in [0,1]$, the homogeneity property from Lemma 2 implies[35]

$$\begin{aligned} T(\alpha f)(z) &= \sup_{\substack{c\in[\bar{c},1]\\ w\in\triangle(K)}} \left[\frac{c^\rho}{\rho} + \alpha(1-c)^\rho \beta \sup_{\gamma\geq 0} \int_Z \phi\Big((w\cdot r(z'))^\rho f(z')\Big|\gamma\Big)\, dP(z'|z)\right] \\ &\geq \sup_{\substack{c\in[\bar{c},1]\\ w\in\triangle(K)}} \left[\alpha\frac{c^\rho}{\rho} + (1-\alpha)\frac{\bar{c}^\rho}{\rho} + \alpha(1-c)^\rho \beta \sup_{\gamma\geq 0} \int_Z \phi\Big((w\cdot r(z'))^\rho f(z')\Big|\gamma\Big)\, dP(z'|z)\right] \\ &= \alpha Tf(z) + (1-\alpha)\delta. \end{aligned}$$

Since the conditions of Theorem 9 are satisfied, conclude that T has a unique fixed point L in A. To see that L is the unique fixed point in $B(Z)_+$, fix any function $f \in B(Z)_+$. Since $b \geq \frac{1}{\rho(1-\beta\bar{r}^\rho)}$ was arbitrary, the preceding arguments hold if we take $b \geq f$ for this function f. It follows that f can only be a fixed point of T if $f = L$. ■

[35] Strictly speaking, this equality follows from applying Lemma 2 to the distribution of $(w\cdot r(\cdot))^\rho f(\cdot)$ under the probability measure $P(\cdot|z)$ and using the change of variables formula. Alternatively, the arguments used to prove that lemma can be immediately adapted to expectations of integrable random variables.

Taking L as in Lemma 11, define $J : \mathbb{R}_+ \times Z \to \mathbb{R}_+$ by $J(a,z) = a^\rho L(z)$. Then, Equation (34) together with the homogeneity property from Lemma 2 imply

$$\begin{aligned}
&\sup_{\substack{c\in[0,a]\\ w\in\triangle(K)}} \left[\frac{c^\rho}{\rho} + \beta \sup_{\gamma\geq 0} \int_Z \phi\Big(J\big(w\cdot r(z')(a-c),\, z'\big)\Big|\gamma\Big)\, dP(z'|z)\right] \\
&\qquad = \sup_{\substack{c\in[0,a]\\ w\in\triangle(K)}} \left[\frac{c^\rho}{\rho} + \beta \sup_{\gamma\geq 0} \int_Z \phi\Big(\big(w\cdot r(z')(a-c)\big)^\rho L(z')\Big|\gamma\Big)\, dP(z'|z)\right] \\
&\qquad = \sup_{\substack{c\in[0,a]\\ w\in\triangle(K)}} \left[\frac{c^\rho}{\rho} + (a-c)^\rho \beta \sup_{\gamma\geq 0} \int_Z \phi\Big((w\cdot r(z'))^\rho L(z')\Big|\gamma\Big)\, dP(z'|z)\right] \\
&\qquad = a^\rho \sup_{\substack{\hat{c}\in[0,1]\\ w\in\triangle(K)}} \left[\frac{\hat{c}^\rho}{\rho} + (1-\hat{c})^\rho \beta \sup_{\gamma\geq 0} \int_Z \phi\Big((w\cdot r(z'))^\rho L(z')\Big|\gamma\Big)\, dP(z'|z)\right] \\
&\qquad = a^\rho L(z) = J(a,z).
\end{aligned}$$

It remains only to show that there exist maximizing c, w, and γ that attain the supremum in this expression. It is clear from the expressions for $J(a,z)$ given above that a maximizing $c \in [0,a]$ exists. To see the existence of maximizing w and γ, first fix any $z \in Z$ and $c \in [0,a]$. Then, define a function $G : \triangle(K) \to \mathbb{R}$ by

$$G(w) = \sup_{\gamma\geq 0} \int_Z \phi\Big(\big(w\cdot r(z')(a-c)\big)^\rho L(z')\Big|\gamma\Big)\, dP(z'|z).$$

It suffices to show that the supremum over γ in this expression is attained and that G is continuous in w. If $c = a$, then the supremum is attained by $\gamma = 0$, and $G(w) = 0$ for any $w \in \triangle(K)$. Alternatively, consider the case of $c < a$. The finiteness of Z and K together with the assumption $r_k(z) > 0$ imply there exists $\underline{r} > 0$ such that $r_k(z) \geq \underline{r}$ for all $z \in Z$. It is also clear from Equation (34) that $L(z) \geq \frac{1}{\rho}$ for all $z \in Z$. Let $\overline{L} = \max_{z\in Z} L(z)$ and $\overline{r} = \max_{1\leq k\leq K} \max_{z\in Z} r_k(z)$. Then, letting $l \equiv (\underline{r}(a-c))^\rho \frac{1}{\rho}$ and $u \equiv (\overline{r}(a-c))^\rho \overline{L}$, we have

$$0 < l \leq \big(w\cdot r(z')(a-c)\big)^\rho L(z') \leq u, \qquad \forall w \in \triangle(K),\ z' \in Z.$$

Therefore, for any $w \in \triangle(K)$, the distribution of $(w\cdot r(\cdot)(a-c))^\rho L(\cdot)$ has support in the closed and bounded interval $[l,u]$. Therefore, Theorem 6 implies that the supremum over γ is attained by some $\gamma \in [l,u]$. Moreover, $w_n \to w$ implies the random variable $(w_n \cdot r(\cdot)(a-c))^\rho L(\cdot)$ converges pointwise (hence, almost surely) to $(w\cdot r(\cdot)(a-c))^\rho L(\cdot)$ on Z. Since almost-sure convergence of a sequence of random variables implies weak convergence of their distributions, the continuity established in Theorem 6 is sufficient to conclude that G is a continuous function. Therefore, it is maximized by some $w \in \triangle(K)$.

References

Aliprantis, C., and K. Border (2006): *Infinite Dimensional Analysis*, 3rd edition. Berlin, Germany: Springer-Verlag.

Backus, D., B. Routledge, and S. Zin (2004): "Exotic Preferences for Macroeconomics," *NBER Macroeconomics Annual*, 19, 319–390.

Barberis, N., and M. Huang (2001): "Mental Accounting, Loss Aversion, and Individual Stock Returns," *Journal of Finance*, 56, 1247–1292.

Barberis, N., M. Huang, and T. Santos (2001): "Prospect Theory and Asset Prices," *Quarterly Journal of Economics*, 116, 1–53.

Barillas, F., L. P. Hansen, and T. J. Sargent (2009): "Doubts or Variability?" *Journal of Economic Theory*, 144, 2388–2418.

Bekaert, G., R. J. Hodrick, and D. A. Marshall (1997): "The Implications of First-Order Risk Aversion for Asset Market Risk Premiums," *Journal of Monetary Economics*, 40, 3–39.

Bénabou, R., and J. Tirole (2011): "Indentity, Morals and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, 126, 805–855.

Benartzi, S., and R. Thaler (1995): "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics*, 110, 75–92.

Blackwell, D. (1965): "Discounted Dynamic Programming," *Annals of Mathematical Statistics*, 36, 226–235.

Brandenburger, A., and E. Dekel (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189–198.

Brunnermeier, M., and J. Parker (2005): "Optimal Expectations," *American Economic Review*, 95, 1092–1118.

Campbell, J. Y. (1999): "Asset Prices, Consumption, and the Business Cycle," in J. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Vol. 1, North-Holland, Amsterdam.

Caplin, A., and J. Leahy (2001): "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics*, 116, 55–79.

Chatterjee, K., and R. V. Krishna (2010): "A Nonsmooth Approach to Nonexpected Utility Theory under Risk," Mimeo.

Chew, S. H. (1983): "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox," *Econometrica*, 51, 1065–1092.

Cochrane, J. H. (2005): *Asset Pricing*, 2nd edition. Princeton, NJ: Princeton University Press.

Debreu, G. (1960): "Topological Methods in Cardinal Utility Theory," in K. J. Arrow, S. Karlin, and P. Suppes (Eds.), *Mathematical Methods in the Social Sciences*, Stanford, California: Stanford University Press.

Dekel, E. (1986): "An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom," *Journal of Economic Theory*, 40, 304–318.

Ekeland, I., and T. Turnbull (1983): *Infinite-Dimensional Optimization and Convexity.* Chicago: The University of Chicago Press.

Epstein, L. G. (2008): "Living with Risk," *Review of Economic Studies*, 75, 1121–1141.

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

Epstein, L. G., and S. Zin (1990): "'First-Order' Risk Aversion and the Equity Premium Puzzle," *Journal of Monetary Economics*, 26, 387–407.

Epstein, L. G., and S. Zin (1991): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99, 263–286.

Epstein, L. G., and S. Zin (2001): "The Independence Axiom and Asset Returns," *Journal of Empirical Finance*, 8, 537–572.

Ergin, H., and T. Sarver (2011): "Hidden Actions and Preferences for Timing of Resolution of Uncertainty," Working paper, Northwestern University.

Fishburn, P. C. (1970): *Utility Theory for Decision Making.* New York: John Wiley and Sons.

Gabaix, X., and D. Laibson (2001): "The 6D Bias and the Equity-Premium Puzzle," *NBER Macroeconomics Annual*, 16, 257–312.

Grossman, S., and G. Laroque (1990): "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, 58, 25–51.

Gul, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59, 667–686.

Gul, F., and W. Pesendorfer (2004): "Self-Control and the Theory of Consumption," *Econometrica*, 72, 119-158.

Hansen, L. P., J. Heaton, J. Lee, and N. Roussanov (2007): "Intertemporal Substitution and Risk Aversion," in *Handbook of Econometrics*, vol. 6A, edited by J. Heckman and E. E. Leamer. Amsterdam: North-Holland.

Hansen, L. P., and R. Jagannathan (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

Hansen, L. P., and T. J. Sargent (2001): "Robust Control and Model Uncertainty," *American Economic Review*, 91, 60–66.

Hansen, L. P., T. J. Sargent, and T. Tallarini (1999): "Robust Permanent Income and Pricing," *Review of Economic Studies*, 66, 873–907.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–292.

Kocherlakota, N. (1996): "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 34, 42–71.

Kőszegi, B., and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121, 1133–1165.

Kőszegi, B., and M. Rabin (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97, 1047–1073.

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

Kreps, D., and E. Porteus (1979): "Temporal von Neumann-Morgenstern and Induced Preferences," *Journal of Economic Theory*, 20, 81–109.

Lucas, R. E. (2003): "Macroeconomic Priorities," *American Economic Review*, 93, 1–14.

Machina, M. (1982): "'Expected Utility' Analysis without the Independence Axiom," *Econometrica*, 50, 277–324.

Machina, M. (1984): "Temporal Risk and the Nature of Induced Preferences," *Journal of Economic Theory*, 33, 199–231.

Maenhout, P. J. (2004): "Robust Portfolio Rules and Asset Pricing," *Review of Financial Studies*, 17, 951–983.

Marinacci, M., and L. Montrucchio (2010): "Unique Solutions for Stochastic Recursive Utilities," *Journal of Economic Theory*, 145, 1776–1804.

Markowitz, H. (1952): "The Utility of Wealth," *Journal of Political Economy*, 60, 151–158.

Mehra, R., and E. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Mertens, J.-F., and S. Zamir (1985): "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1–29.

Munkres, J. (2000): *Topology*. Upper Saddle River, NJ: Prentice Hall.

Phelps, R. R. (1993): *Convex Functions, Monotone Operators, and Differentiability*. Berlin, Germany: Springer-Verlag.

Rabin, M. (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68, 1281–1292.

Routledge, B., and S. Zin (2010): "Generalized Disappointment Aversion and Asset Prices," *Journal of Finance*, 65, 1303–1332.

Segal, U., and A. Spivak (1990): "First Order versus Second Order Risk Aversion," *Journal of Economic Theory*, 51, 111–125.

Skiadas, C. (2003): "Robust Control and Recursive Utility," *Finance and Stochastics*, 7, 475–489.

Strzalecki, T. (2011): "Axiomatic Foundations of Multiplier Preferences," *Econometrica*, 79, 47–73.

Tallarini, T. (2000): "Risk-Sensitive Real Business Cycles," *Journal of Monetary Economics*, 45, 507–532.

Weil, P. (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

Weil, P. (1990): "Nonexpected Utility in Macroeconomics," *Quarterly Journal of Economics*, 105, 29–42.