

Mixture-Averse Preferences and Heterogeneous Stock Market Participation

Todd Sarver

Duke University

September 9, 2016

ERID Working Paper Number 229

This paper can be downloaded without charge from the Social
Science Research Network Electronic Paper Collection:

<http://ssrn.com/abstract=2837262>

**Economic Research Initiatives at Duke
WORKING PAPERS SERIES**



Mixture-Averse Preferences and Heterogeneous Stock Market Participation*

Todd Sarver[†]

First Version: September 7, 2011

Current Draft: September 9, 2016

Abstract

To study intertemporal decisions under risk, we develop a new recursive model of non-expected-utility preferences. The main axiom of our analysis is called *mixture aversion*, as it captures a dislike of probabilistic mixtures of lotteries. Our representation for mixture-averse preferences can be interpreted as if an individual optimally selects her risk attitude from some feasible set. The representation includes special cases where the choice of risk attitude takes the form of an optimal selection of a reference point. We analyze the implications of the model for both insurance and investment decisions. The main application of the paper shows that mixture-averse preferences can generate endogenous heterogeneity in equilibrium stock market participation, even when consumers have identical preferences and even among wealthy households.

KEYWORDS: mixture aversion, optimal risk attitude, reference point, stock market participation, equity premium puzzle

*I am grateful to David Ahn and Philipp Sadowski for detailed feedback that led to significant improvements in this paper. I also thank Tim Bollerslev, Simone Cerreia-Vioglio, Chris Chambers, Eddie Dekel, David Dillenberger, Larry Epstein, Drew Fudenberg, Faruk Gul, Cosmin Ilut, R. Vijay Krishna, Bart Lipman, Mark Machina, Pietro Ortoleva, Jonathan Parker, Wolfgang Pesendorfer, Chris Shannon, Marciano Siniscalchi, Costis Skiadas, Mengke Wang, and seminar participants at Boston University, Stanford University, Northwestern University, Duke University, Princeton University, University of Pennsylvania, Caltech, Harvard/MIT, UCSD, and UCLA for helpful comments and discussions. This paper was previously circulated under the titles “Optimal Anticipation” and “Optimal Reference Points and Anticipation.”

[†]Duke University, Department of Economics, 213 Social Sciences/Box 90097, Durham, NC 27708. Email: todd.sarver@duke.edu.

1 Introduction

Intertemporal preferences of individuals are of central importance in many economic interactions. They play a key role in the the determination of aggregate macroeconomic variables and prices in financial markets. The explanatory power of models of dynamic choice has been significantly improved in recent years by relaxing the assumption of separability of preferences across states and time, as in [Kreps and Porteus \(1978\)](#), and by incorporating various non-expected-utility preferences, as suggested by [Epstein and Zin \(1989, 1990\)](#) and further pursued in the literature that followed.

The present paper contributes to this literature by using the general recursive framework developed by [Epstein and Zin \(1989\)](#) to study a new class of dynamic risk preferences. The central axiom of our analysis is called *mixture aversion*, as it implies a dislike of probabilistic mixtures of lotteries. Couched in the dynamic structure of our domain, this axiom imposes restrictions on an individual's willingness to trade current consumption for uncertain improvements in future outcomes: Suppose an individual can give up some consumption now in order to increase the probability of a better outcome tomorrow. Our axiom requires that when the initial probability of the better future outcome is higher, this trade becomes more attractive to the individual. In other words, increasing the probability of a good outcome makes additional increases even more desirable.¹

To illustrate, imagine an individual can exert additional (costly) effort in the current period that will increase the probability of a future promotion. Would this individual be more willing to put forth effort if the initial chances of the promotion are low and could be increased slightly, or when the initial probability is already high and could be made certain? Assuming the marginal impact of current effort on the probability of promotion is the same in each case, mixture aversion implies that the individual would exert greater effort in the latter scenario.

As this example demonstrates, mixture aversion is closely related to the *certainty effect* documented by [Allais \(1953\)](#), [Kahneman and Tversky \(1979\)](#), and others—individuals typically assign a premium to increases in probability that lead to certainty. Our axiom implies this property and also extends it to probabilities away from certainty; for example, an individual may be willing to work harder for a promotion when her initial chances are higher, even if it is impossible to make the promotion certain. Mixture aversion is also connected to the literature on probabilistic insurance. As observed by [Kahneman and Tversky \(1979\)](#) and [Wakker, Thaler, and Tversky \(1997\)](#), an individual's willingness to pay for insurance coverage that pays with only half probability in the event of a loss is typically much less than half of the amount that she would be willing to pay for complete

¹In the context of static risk preferences, mixture aversion is sometimes used to refer to quasiconvexity of preferences in probabilities. As we discuss in Section 3.1, our axiom implies this condition and hence it can be thought of as a stronger form of mixture aversion.

(certain) coverage.² While intuitive, this dislike of probabilistic insurance is difficult to reconcile with expected-utility theory and many available non-expected-utility theories.³ Mixture-averse preferences are consistent with a dislike of probabilistic insurance. In fact, in the case where premiums are paid one period in advance, this behavior is a direct implication of our axiom.

We will use our model to study demand for financial assets such as equity and insurance. In our main application, we show that mixture-averse preferences provide a rationale for low levels of stock market participation and for investment by some participating households only in low-risk (e.g., bond) portfolios, both of which are puzzling from the perspective of most existing models. Mixture-averse preferences can also generate a marginal willingness to pay for additional insurance coverage that increases at some levels of coverage, which may help to explain the high prices some consumers pay to decrease their insurance deductibles.

1.1 Preview of Results

Epstein and Zin (1989) provided a general recursive formula that can be used to embed any risk preference developed in a static context into an infinite-horizon dynamic environment. They proved that a value function for this recursive representation exists whenever risk preferences satisfy suitable continuity properties. In Section 2, we begin by introducing our framework and defining the Epstein-Zin representation formally.

Starting from this general Epstein-Zin formula, our analysis in Section 3.1 studies the additional implications of the mixture aversion axiom.⁴ To illustrate, consider a simple consumption-savings problem with a gross return R_t that is i.i.d. across time.

²There are several different formulations of the probabilistic insurance problem. Kahneman and Tversky (1979) formulated the problem in a way that makes it incompatible with expected utility. Our description of probabilistic insurance comes from Wakker, Thaler, and Tversky (1997), which connects more directly with many real-world instances of risk of non-payment from an insurance policy.

³For example, Kahneman and Tversky (1979) showed that the preferences of the majority of subjects in their probabilistic insurance problem were inconsistent with expected-utility theory. Their argument can easily be extended to show that dislike of their formulation of probabilistic insurance is not compatible with any preference that is quasiconcave in probabilities and risk averse. Interestingly, the only prominent non-expected-utility model that has been shown to be consistent with aversion to probabilistic insurance is risk-averse rank-dependent utility (see Segal (1988)), which turns out to be a special case of mixture-averse preferences (see Section S.2 of the Supplementary Appendix).

⁴The assumption that preferences have the recursive structure described in Epstein and Zin (1989) can be broken down into more fundamental assumptions: Chew and Epstein (1991) provided axiomatic foundations for a version of the Epstein-Zin representation. Appendix A.1 contains an axiomatic characterization of the exact form of the Epstein-Zin representation used in this paper.

The representation for mixture-averse preferences has a value function of the form

$$\mathcal{V}(w_t) = \max_{c_t, w_{t+1}} \left\{ u(c_t) + \beta \sup_{\phi \in \Phi} \mathbb{E}_t [\phi(\mathcal{V}(w_{t+1}))] \right\}. \quad (1)$$

In this recursion, the random variables c_t and w_{t+1} evolve according to the constraint $w_{t+1} = (w_t - c_t)R_{t+1}$. The elements of the representation are a utility function u , a discount factor β , and a set of nondecreasing functions Φ that satisfies

$$\sup_{\phi \in \Phi} \phi(x) = x \quad (2)$$

for every real number x in the domain. Equation (2) implies that this representation reduces to time-separable utility for deterministic problems; hence the model is completely standard absent risk. However, when faced with uncertainty, the individual in our model is able to alter her risk attitude through her selection of a transformation ϕ . Since the transformation is chosen to maximize utility, we refer to Equations (1) and (2) as the *optimal risk attitude* (ORA) representation. In Section 3.2, we describe some parametric special cases of the ORA representation that will be used in our applications, including several which can be interpreted more specifically as models of endogenous reference point selection.

While the optimization over risk attitudes in our representation might suggest that the individual is risk loving (or in some sense less risk averse), in fact the opposite is true. It follows as a corollary of our comparative statics result in Section 3.3 that any ORA representation is more risk averse than time-separable expected utility with the same u and β . Using the simple consumption-savings problem described above to illustrate, Equation (2) implies that for any random wealth w_{t+1} ,

$$\sup_{\phi \in \Phi} \mathbb{E}_t [\phi(\mathcal{V}(w_{t+1}))] \leq \mathbb{E}_t \left[\sup_{\phi \in \Phi} \phi(\mathcal{V}(w_{t+1})) \right] = \mathbb{E}_t [\mathcal{V}(w_{t+1})].$$

In particular, the ORA representation can generate high levels of risk aversion for small gambles yet more moderate attitudes toward increases in risk when exposure is already large. In Section 3.2, we illustrate this feature of the model using an insurance example. We show that an individual may be willing to pay more for her last dollar of coverage than her first, which can yield a high willingness to pay for reduced insurance deductibles.

Section 4 uses mixture-averse preferences to examine dynamic investment behavior and asset returns. In Section 4.1, we discuss one of the most widely-documented and puzzling patterns from household finance: Many households have little or no money invested in equity. Recent theoretical and empirical work has made progress in explaining the nonparticipation decisions of some households, but significant aspects of this puzzle of

limited stock market participation remain—such as limited participation by some wealthy households. To shed new light on this behavior, in Section 4.2 we describe an economy with a continuum of agents with identical ORA representations that are homothetic in wealth.

In Section 4.3, we illustrate informally how heterogeneity in risk exposure can arise endogenously in this economy, with some agents holding very little risk in equilibrium— independent of their wealth. One segment of the population chooses a risk attitude that is less sensitive to gains and losses and therefore holds greater consumption risk; the other chooses a risk attitude that provides higher utility for low-risk allocations, but is more sensitive to losses, and therefore holds very little consumption risk. Importantly, this is not just one possible equilibrium, but rather is the unique equilibrium of this economy for a range of parameter values.

In Section 4.4, we provide a calibration of the dynamic general equilibrium of this economy that verifies these informal observations and quantifies the level of participation and the resulting equity premium. Since our model can generate heterogeneity in participation decisions even when agents have identical preferences, it is tractable enough to permit a complete dynamic general equilibrium analysis. In contrast, all existing preference-based explanations of the participation puzzle require preference heterogeneity to generate limited equilibrium stock market participation, and hence these models tend to be restricted to partial equilibrium analysis.

Proofs are contained in the Appendix, and some supporting results and discussions are further relegated to a Supplementary Appendix.

1.2 Related Recursive Models

Given the generality of the recursive formula developed by Epstein and Zin (1989), it has limited empirical content absent additional restrictions on the permissible class of risk preferences. Thus the benefit of their general representation is not that it provides a specific functional form for use in applications, but rather that it provides a framework for easily incorporating any model developed in the world of static risk into dynamic environments.

The most widely-used special case of the Epstein-Zin representation is the infinite-horizon formulation of Kreps and Porteus (1978) expected utility, which we will refer to as Epstein-Zin-Kreps-Porteus (EZKP) utility (see Appendix A.2 for a formal definition). As emphasized by Epstein and Zin (1989) and Weil (1989, 1990), this model permits a separation between risk aversion and intertemporal substitution that is not possible for standard *time-separable* expected utility. Despite its usefulness, EZKP utility is still unable to resolve a number of anomalies associated with expected utility, such as the

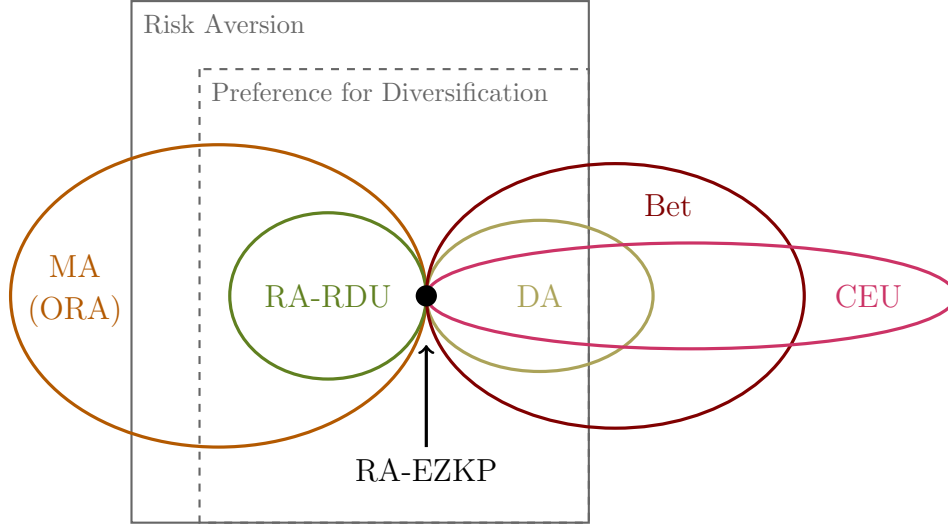


Figure 1. Relationship between mixture-averse (MA) preferences (ORA representation) and other recursive risk preferences: risk-averse Epstein-Zin-Kreps-Porteus expected utility (RA-EZKP), risk-averse rank-dependent utility (RA-RDU), betweenness (Bet), disappointment aversion (DA), cautious expected utility (CEU).

Allais paradox, the Rabin paradox, and the equity premium puzzle. These limitations have motivated a large literature that studies various non-expected-utility theories within the recursive framework of Epstein and Zin (1989).⁵

We summarize the relationship between the ORA representation and several of these prominent theories in Figure 1. The connection to EZKP utility is established in Appendix A.2, where we show that EZKP utility is in fact a special case of the ORA representation, provided the certainty equivalent is risk averse. The other connections depicted in this figure follow from known results and are discussed in detail in Section S.2 of the Supplementary Appendix.

The rectangles in Figure 1 illustrate two properties of the certainty equivalents in these representations that will be relevant for our applications: preference for diversification (quasiconcavity of preferences with respect to random variables)⁶ and risk aversion

⁵For example, a parameterized special case of rank-dependent utility (Quiggin (1982); Yaari (1987); Segal (1989)) was used by Epstein and Zin (1990); the disappointment aversion model of Gul (1991) was used by Bekaert, Hodrick, and Marshall (1997) and Ang, Bekaert, and Liu (2005); other special cases of betweenness preferences (Chew (1983); Dekel (1986)) were used by Epstein and Zin (2001) and Routledge and Zin (2010).

⁶We should be careful to distinguish two related but distinct concepts. Quasiconcavity of preferences in *random variables* is not directly tied to quasiconcavity of preferences in *probabilities*. Mixture aversion implies quasiconvexity of preferences in probabilities (see Section 3.1) and is compatible with (but does not imply) quasiconcavity of preferences in random variables. In this paper, we follow the convention of referring to the latter property as “preference for diversification.” It is worth noting that this terminology is slightly misleading, as it suggests that an investor whose preferences violate this property would not choose to diversify her portfolio in order to reduce risk exposure. In many cases, risk aversion alone is

(monotonicity with respect to second-order stochastic dominance). Preference for diversification is often considered a desirable property since it enhances analytic tractability. However, it turns out that several applications of our model, including some properties of demand for insurance (see Section 3.2) and equilibrium heterogeneity in asset market participation (see Section 4), rely crucially on relaxing preference for diversification. At the same time, it is often desirable (both for realism and technical simplicity) to maintain risk aversion in such applications. Mixture-averse preference are unique among those depicted in Figure 1 in their ability to separate these two properties. The aforementioned sections and Section S.3 of the Supplementary Appendix investigate these issues in greater detail.

2 Preliminaries

2.1 Framework

For any topological space X , let $\Delta(X)$ 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 temporal lotteries. This domain is rich enough to encode not only the atemporal distribution of consumption streams but also how information about future consumption arrives through time. For example, future wealth and hence future consumption may depend on the returns to investments which are realized gradually over a sequence of interim periods. Formally, let C be a compact and connected metrizable space, denoting the consumption space for each period.⁷ A one-period consumption lottery is simply an element of $\Delta(C)$. The space of two-period temporal lotteries is $\Delta(C \times \Delta(C))$, the space of three-period temporal lotteries is $\Delta(C \times \Delta(C \times \Delta(C)))$, and so on.

Extending this idea to the infinite horizon, the domain in this paper is a compact and connected metrizable space D that can be identified (via a homeomorphism) with $C \times \Delta(D)$. Epstein and Zin (1989) showed that such a space is well-defined.⁸ Intuitively, a

sufficient to ensure the optimality of fully diversified portfolios (e.g., if asset returns are i.i.d.).

⁷While the axiomatic analysis will be restricted to compact spaces, the investment application in Sections 4 will allow for the unbounded consumption space \mathbb{R}_+ . It is possible to generalize the axiomatic analysis to infinite-horizon temporal lotteries 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.

⁸See also Theorem 2.1 in Chew and Epstein (1991). Similar constructions were employed by Mertens and Zamir (1985) and Brandenburger and Dekel (1993) in the context of hierarchies of beliefs, and by Gul and Pesendorfer (2004) to develop a space of infinite-horizon decision problems.

lottery over D returns consumption today together with another infinite-horizon temporal lottery beginning tomorrow. Therefore, elements of D will typically be denoted by (c, m) , where $c \in C$ and $m \in \Delta(D)$. The primitive of the axiomatic model is a binary relation \succsim on the space D .

2.2 Epstein-Zin Preferences

In this section, we formally define the Epstein-Zin representation. Their model will serve as the starting point for the analysis of mixture-averse preferences and the optimal risk attitude representation in Section 3.1.

Definition 1 A *certainty equivalent* is a continuous function $W : \Delta([a, b]) \rightarrow \mathbb{R}$ that satisfies $W(\delta_x) = x$ for all $x \in [a, b]$ and is monotone with respect to first-order stochastic dominance.⁹

For any measurable value function $V : D \rightarrow [a, b]$ and any probability $m \in \Delta(D)$, let $m \circ V^{-1}$ denote the distribution (on $[a, b]$) of continuation values induced by m .¹⁰

Definition 2 An *Epstein-Zin (EZ)* representation is a tuple (V, u, W, β) consisting of a continuous function $V : D \rightarrow \mathbb{R}$ that represents \succsim , a continuous and nonconstant function $u : C \rightarrow \mathbb{R}$, a certainty equivalent $W : \Delta([a, b]) \rightarrow \mathbb{R}$ (where $a = \min V$ and $b = \max V$), and a scalar $\beta \in (0, 1)$ such that, for all $(c, m) \in D$,

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

In Appendix A.1, we provide a complete axiomatic characterization of the EZ representation in Definition 2. Chew and Epstein (1991) provided a related characterization of a representation that is not necessarily separable with respect to c and m .¹¹ Our axioms parallel their treatment, but strengthen their separability assumption by applying conditions from Debreu (1960) in an intertemporal context. To focus on the most novel aspects of our model of mixture-averse preferences, the starting point of the theorems in the main text will be a preference \succsim with an EZ representation.

⁹A function $W : \Delta([a, b]) \rightarrow \mathbb{R}$ is *monotone with respect to FOSD* if for any $\mu, \eta \in \Delta([a, b])$ satisfying $\mu([x, b]) \geq \eta([x, b])$ for all $x \in [a, b]$, we have $W(\mu) \geq W(\eta)$.

¹⁰This 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 $V(\hat{c}, \hat{m}) \in E$, which is $m \circ V^{-1}(E)$.

¹¹They considered a nonlinear aggregator of current consumption and the continuation value: $V(c, m) = \psi(c, W(m \circ V^{-1}))$.

3 Dynamic Mixture-Averse Preferences

3.1 Main Axiom and Representation Result

Our main axiom imposes a type of aversion to probabilistic mixtures of lotteries.

Axiom 1 (Mixture Aversion) *For any $c, c' \in C$ and $m, m' \in \Delta(D)$,*

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

Axiom 1 puts structure on the individual's willingness to trade current consumption for changes in the probability of future outcomes. Using current consumption to measure the value of changes to the probability assigned to the lotteries m and m' , this axiom implies that the benefit of increasing the probability that m' is the relevant lottery from zero to one-half is (weakly) less than the benefit of increasing the probability from one-half to one.¹² Equivalently, thinking of Axiom 1 as a restriction on the marginal rate of substitution for (non-infinitesimal) changes in current consumption and future probabilities, it implies that current consumption is an inferior good: The individual becomes more willing to increase the probability weighting of a preferred lottery at the expense of current consumption as the baseline probability of that lottery increases.¹³

For one interpretation of this axiom, consider an individual who can take steps to mentally prepare herself for the uncertainty that she faces. Her planning is the simplest when the future is known (the lottery over future outcomes is degenerate), whereas greater uncertainty about the future makes planning more difficult. In particular, taking the mixture between two lotteries m and m' complicates her planning process. Therefore, it is intuitive that her value for increasing the weighting of m' from zero to one-half is less than half of her value for increasing it from zero to one. In contrast, an individual whose preferences respect the axioms of standard time-separable expected utility would assign the same value to an increase in the probability of m' regardless of its current weighting, and thus would satisfy Axiom 1 as well as its converse.

Our utility representation is defined as follows.

¹²To be precise, Axiom 1 considers the mixture $\frac{1}{2}m + \frac{1}{2}m'$ of the two lotteries m and m' . Thus uncertainty about which will be the relevant lottery (m or m') and the uncertainty in the lottery itself are both resolving at the same time.

¹³In traditional demand problems, versions of supermodularity have been used to study normal and inferior goods. Formally, Axiom 1 is equivalent to the definition of $(\nabla_i^\lambda, \Delta_i^\lambda)$ -quasisubmodularity from Quah (2007) for $\lambda = 1/2$ and i equal to the consumption dimension of $C \times \Delta(D)$. When the other axioms of the Epstein-Zin representation are also assumed, it can be shown that this notion of quasisubmodularity is satisfied for all $\lambda \in [0, 1]$, a stronger condition that Quah (2007) referred to as C_i -quasisubmodularity. This property is stronger than the traditional definition of quasisubmodularity studied by Milgrom and Shannon (1994), which holds trivially for any additively-separable function.

Definition 3 An *optimal risk attitude (ORA)* representation is a tuple (V, u, Φ, β) consisting of a continuous function $V : D \rightarrow \mathbb{R}$ that represents \succsim , a continuous and non-constant function $u : C \rightarrow \mathbb{R}$, a collection Φ of continuous and nondecreasing functions $\phi : [a, b] \rightarrow \mathbb{R}$ (where $a = \min V$ and $b = \max V$), and a scalar $\beta \in (0, 1)$ such that

$$V(c, m) = u(c) + \beta \sup_{\phi \in \Phi} \int_D \phi(V(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}), \quad (3)$$

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

$$\sup_{\phi \in \Phi} \phi(x) = x, \quad \forall x \in [a, b]. \quad (4)$$

We will sometimes write Equation (3) more compactly by treating V as a random variable defined on the space D and using the expectation operator:

$$V(c, m) = u(c) + \beta \sup_{\phi \in \Phi} \mathbb{E}_m[\phi(V)].$$

Note that the value function V is included explicitly in the definition of the ORA representation. Using similar techniques to Epstein and Zin (1989), we show in Section S.5 of the Supplementary Appendix that a value function exists for any (u, Φ, β) as in Definition 3.

The interpretation of Axiom 1 in terms of mental preparation also extends to the ORA representation. An individual may mentally prepare herself for different future outcomes or different levels of risk, which corresponds to a choice of ϕ in this representation.¹⁴ A larger set of transformations Φ implies greater flexibility in this planning and hence greater ability to tailor her risk attitude to the uncertainty that she faces (cf. our comparative risk aversion result in Section 3.3). Several examples of the ORA representation that lend themselves to more specific interpretations of this mental preparation in terms of anticipation and loss aversion are explored in Section 3.2.

We now state our main representation result.

Theorem 1 Suppose \succsim has an Epstein-Zin representation (V, u, W, β) .¹⁵ The following are equivalent:

¹⁴The selection of a risk attitude in our representation is similar in spirit to models of consumption commitments and adjustment costs. For example, the choice of physical commitments, such as mortgage agreements or purchases of durable consumption goods, impacts risk preferences for future wealth (see Grossman and Laroque (1990); Gabaix and Laibson (2001); Chetty and Szeidl (2007, 2016)). Even more closely related, both conceptually and technically, are Kreps and Porteus (1979), Machina (1984), and Ergin and Sarver (2015), who studied the revealed-preference implications of commitments that are unobservable or psychological in nature. Maccheroni (2002) developed a model of maxmin under risk that had a very different interpretation but also relied on similar techniques.

¹⁵Equivalently, suppose \succsim satisfies Axioms 2–7 in Appendix A.1.

1. The relation \succsim satisfies Axiom 1.
2. The certainty equivalent W in the EZ representation of \succsim is convex in probabilities.
3. The relation \succsim has an optimal risk attitude representation (V, u, Φ, β) .

To avoid any confusion about terminology, we should note that the term mixture aversion has also been used to refer to the related property of quasiconvexity in probabilities. As is evident from the preceding theorem, Axiom 1 (together with the other axioms of the Epstein-Zin representation) implies quasiconvexity in probabilities.¹⁶ Our axiom imposes additional structure beyond quasiconvexity that delivers the interpretable and parsimonious representation in Theorem 1, while still maintaining sufficient generality to permit the behavior in applications that we set out to explain.

We conclude this section with a sketch of the proof of Theorem 1. The intuition for why mixture aversion implies convexity of the certainty equivalent in probabilities should be clear from our discussion of the axiom. The basic intuition for why condition 2 implies 3 comes from standard duality results. Since W is convex, it can be expressed as the supremum of some collection of affine functions. Since any affine function on $\Delta([a, b])$ can be given an expected-utility representation, this implies there exists a collection Φ of continuous functions $\phi : [a, b] \rightarrow \mathbb{R}$ such that, for any $\mu \in \Delta([a, b])$,

$$W(\mu) = \sup_{\phi \in \Phi} \int_a^b \phi(x) d\mu(x). \quad (5)$$

Moreover, since the certainty equivalent satisfies $W(\delta_x) = x$, we have $\sup_{\phi \in \Phi} \phi(x) = x$ for all $x \in [a, b]$. Using the change of variables formula, it follows that for every $(c, m) \in D$,

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

The only missing step in this sketch is showing that the collection Φ contains only *nondecreasing* functions. The proof of this property follows from a result presented

¹⁶Formally, risk preferences satisfy quasiconvexity if $(c, m) \succsim (c, m')$ implies $(c, m) \succsim (c, \alpha m + (1 - \alpha)m')$. While any utility representation that is convex in probabilities satisfies quasiconvexity, it is also well known that the converse is not true: There are many preferences that are quasiconvex in probabilities that cannot be given a convex utility representation (e.g., betweenness preferences other than those that satisfy independence), and therefore such preferences will violate Axiom 1. See [Cerrei-Vioglio \(2009\)](#) for a representation result for the class of all continuous and quasiconcave static risk preferences; a dual version of his representation could be used to represent quasiconvex risk preferences.

in Section S.1 of the Supplementary Appendix: We show that if a convex utility representation W respects a stochastic dominance order, then there exists a collection of expected-utility functions Φ that generates W and respects the same dominance order. In particular, since a certainty equivalent W is by definition monotone with respect to FOSD, this result implies there exists a collection Φ satisfying Equation (5) such that each $\phi \in \Phi$ is nondecreasing.

3.2 Parametric Examples and Some Properties of the Model

In this section, we describe several special cases of the optimal risk attitude certainty equivalent. These parametric examples will be used to illustrate some of the results in subsequent sections, and they will also be used in our application.

Suppose the collection Φ consists of transformations $\phi(x|\gamma, \theta)$ that are indexed by a pair of parameters $\gamma \in \Gamma$ and $\theta \in \Theta$. The first parameter could be interpreted as a target or anticipated utility level, and the second parameter determines risk aversion. Formally, we assume that Γ is an interval of real numbers that contains the range of V , and we impose the following restrictions on the parameterized transformation function:

$$\begin{aligned} \phi(x|\gamma, \theta) &\leq x, \quad \text{with equality if } x = \gamma, \\ \phi(x|\gamma, \theta') &\leq \phi(x|\gamma, \theta), \quad \text{if } \theta' > \theta. \end{aligned} \tag{6}$$

The following examples satisfy these conditions.

Example 1 (Smooth Transformation) For $\gamma \in \mathbb{R}$ and $\theta > 0$, consider the parameterized function

$$\phi(x|\gamma, \theta) = \gamma + \frac{1}{\theta} - \frac{1}{\theta} \exp(-\theta(x - \gamma)). \tag{7}$$

Example 2 (Kinked Transformation) For $\gamma \in \mathbb{R}$ and $\theta \in [0, 1]$, consider the parameterized function

$$\phi(x|\gamma, \theta) = \begin{cases} \gamma + (1 - \theta)(x - \gamma) & \text{if } x \geq \gamma \\ \gamma + (1 + \theta)(x - \gamma) & \text{if } x < \gamma. \end{cases} \tag{8}$$

In the simplest class of examples, θ is a fixed parameter and the individual optimizes only over γ . The set of transformations is therefore $\Phi = \{\phi(\cdot|\gamma, \theta) : \gamma \in \Gamma\}$, which implies the certainty equivalent can be written as

$$W(\mu) = \sup_{\gamma \in \Gamma} \int \phi(x|\gamma, \theta) d\mu(x). \tag{9}$$

Note that this collection Φ satisfies Equation (4) in the definition of the ORA representation by the first restriction in Equation (6). Moreover, as we will establish formally in

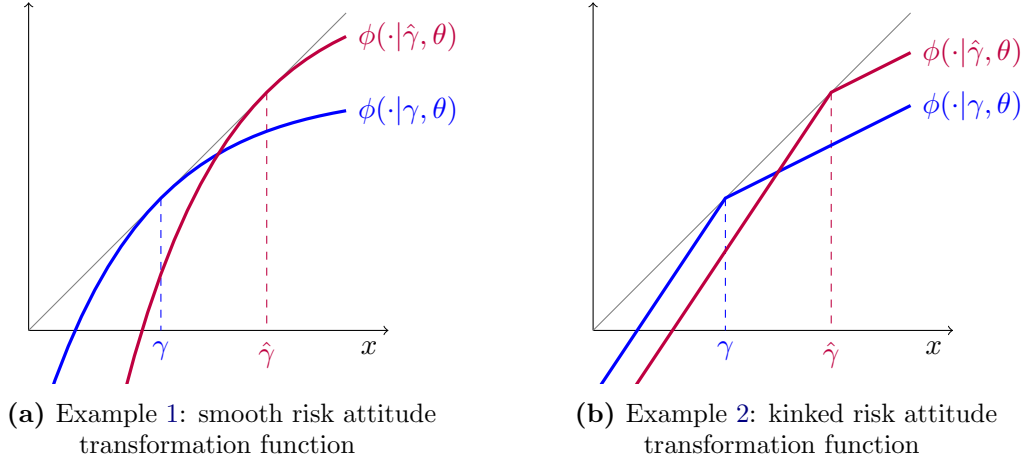


Figure 2. Special cases of the parametric certainty equivalent in Equation (9).

Section 3.3, the second condition in Equation (6) implies that increasing θ leads to an increase in risk aversion. Figure 2 illustrates the transformation functions in Examples 1 and 2 for fixed θ .

Before proceeding to the more general parametric form of the certainty equivalent that will be used in our applications, we discuss some connections to other models and an interesting interpretation of the decision-making process that might give rise to these functional forms. The examples described above are special cases of what Ben-Tal and Teboulle (1986, 2007) referred to as the *optimized certainty equivalent*, which is the special case of Equation (9) where

$$W(\mu) = \sup_{\gamma \in \mathbb{R}} \left\{ \gamma + \int \varphi(x - \gamma) d\mu(x) \right\} \quad (10)$$

for some increasing and concave function φ that satisfies $\varphi(0) = 0$ and $\varphi(x) \leq x$.¹⁷ Ben-Tal and Teboulle (2007, Examples 2.1 and 2.3) also observed several useful properties of the preceding examples that will be related to our analysis. They showed that the certainty equivalent defined by Equations (7) and (9) turns out to be the certainty equivalent of an exponential expected-utility function. This connection is a special case of a more general relationship that we examine in Appendix A.2: Epstein-Zin-Kreps-Porteus (EZKP) expected utility is a special case of the ORA representation, provided the certainty equivalent is risk averse. Ben-Tal and Teboulle (2007) also showed that for any probability distribution μ , taking $\gamma = \text{median}(\mu)$ maximizes Equation (9) when $\phi(x|\gamma, \theta)$ takes the form in Equation (8).¹⁸

¹⁷See also Gollier and Muermann (2010) for a related utility representation.

¹⁸Moreover, it can be shown that the certainty equivalent defined by Equations (8) and (9) is a special

One interpretation of certainty equivalents of the form in Equation (10) involves anticipatory utility and the optimal choice of future reference points. The choice of γ in this certainty equivalent can be interpreted as anticipating some future utility level. The first term γ in the maximization problem captures the anticipatory utility associated with looking forward to this future utility.¹⁹ The term $\varphi(x - \gamma)$ inside the integral can be interpreted as the value of gains or losses relative to the reference level γ . Thus ex-ante anticipation forms a reference level for subsequent outcomes, and gains and losses are measured relative to this reference point. For example, looking forward to a future promotion may bring enjoyment in and of itself, but also heighten the feelings of disappointment if the promotion is not received.²⁰

Some of the most novel implications of our model are obtained from examples that extend beyond the form given in Equation (9). In that specification, the value of θ is fixed; however, θ itself may be a choice variable in the formula for the certainty equivalent. For example, an individual may be able to reduce her sensitivity to risk, but at some psychological cost. Formally, let Θ be any subset of real numbers, let $\tau : \Theta \rightarrow \mathbb{R}$ be a cost function that satisfies $\inf_{\theta \in \Theta} \tau(\theta) = 0$, and define a collection of transformations by

$$\Phi = \{ \phi(\cdot | \gamma, \theta) - \tau(\theta) : \gamma \in \Gamma, \theta \in \Theta \}.$$

The resulting certainty equivalent can be written as

$$W(\mu) = \sup_{\theta \in \Theta} \sup_{\gamma \in \Gamma} \left\{ \int \phi(x | \gamma, \theta) d\mu(x) - \tau(\theta) \right\}. \quad (11)$$

When this certainty equivalent is applied to the transformation functions from Examples 1 and 2, one interpretation along the lines discussed above is that the choice of θ corresponds to the intensity of anticipation of the future utility level γ , where sensitivity to gains and losses and anticipatory utility (now given by $\gamma - \tau(\theta)$) are both increasing in θ (assuming $\tau(\theta)$ is decreasing in θ).

For a concrete example of the behavior associated with the certainty equivalent in Equation (11), consider how an individual's marginal willingness to pay for additional insurance coverage changes with her existing level of coverage. When an insurance policy

case of the dual model of Yaari (1987): $W(\mu) = \int x d(g \circ F_\mu)(x)$ where F_μ is the cumulative distribution for the measure μ and $g(\alpha) = (1 + \theta)\alpha$ for $\alpha \leq 1/2$ and $g(\alpha) = (1 - \theta)\alpha + \theta$ for $\alpha \geq 1/2$.

¹⁹The conscious choice of anticipation in this interpretation should not be confounded with models that treat “anticipation” as uncontrollable anxiety or enjoyment about the future (e.g., Loewenstein (1987); Caplin and Leahy (2001); Kőszegi (2010); Epstein (2008)).

²⁰Existing approaches to modeling reference dependence assume that reference points are directly determined by the status quo (e.g., Markowitz (1952); Kahneman and Tversky (1979); Barberis, Huang, and Santos (2001)) or future expectations (e.g., Gul (1991); Kőszegi and Rabin (2006, 2007); Masatlioglu and Raymond (2016)). Our interpretation of Equation (10) suggests a new alternative whereby individuals exert direct influence on their future reference points through their choice of anticipation.

covers a significant portion of any losses, the individual may have a high marginal willingness to pay for additional coverage that would make the policy complete (or nearly complete) and allow her to avoid loss altogether (see [Sydnor \(2010\)](#)). It is also conceivable that her marginal willingness to pay for small increases in coverage is less at some lower levels of coverage. Intuitively, an individual who is only mentally preparing herself for a small loss has a strong incentive to pay to keep her loss small, whereas an individual who has already resigned herself to the possibility of a large loss might place lower value on marginally reducing her exposure. The following stylized example shows how such behavior is possible within our model.

Example 3 (Reservation Price for Insurance) Suppose the individual has wealth w and faces a loss of amount L with probability $\pi < 1/2$. Suppose the individual evaluates uncertain future wealth using the certainty equivalent defined by Equations (8) and (11) for $\Theta = \{\theta_1, \theta_2\}$, where $0 = \theta_1 < \theta_2 < 1$ and $0 = \tau(\theta_2) < \tau(\theta_1)$.²¹ Let $P(y)$ denote this individual's maximum willingness to pay (reservation price) for $y \in [0, L]$ dollars of insurance coverage paid in the event of a loss. Using array notation for the resulting lotteries over future wealth, it is easy to show from the functional form in Equation (8) that

$$\begin{aligned} P(y) &= W \begin{pmatrix} w - L + y & \pi \\ w & 1 - \pi \end{pmatrix} - W \begin{pmatrix} w - L & \pi \\ w & 1 - \pi \end{pmatrix} \\ &= \max_{\theta \in \Theta} \left\{ \phi(w - \pi L + \pi y | w, \theta) - \tau(\theta) \right\} - \max_{\theta \in \Theta} \left\{ \phi(w - \pi L | w, \theta) - \tau(\theta) \right\}, \end{aligned}$$

where the second equality follows because w is the median outcome of these lotteries and hence $\gamma = w$ is optimal as noted previously. Figure 3 illustrates the function $P(y)$.²²

Example 3 is useful for illustrating both similarities and differences between the ORA representation and existing models. Like many other preferences that satisfy first-order risk aversion (see [Segal and Spivak \(1990\)](#)), an individual with the preferences described in this example is willing to purchase full insurance coverage even at an actuarially unfair rate. However, this example also generates a marginal willingness to pay for insurance that is increasing at some levels of coverage. While such behavior seems quite

²¹This example focuses on the certainty equivalent applied in a static decision problem for ease of exposition. It could equivalently be formulated using risk about future consumption within a (dynamic) ORA representation with $u(c) = c$.

²²For ease of exposition, we assumed that the only nonlinearity in the transformation $\phi(x|\gamma, \theta)$ is the kink at $x = \gamma$. The implication of this assumption is that marginal willingness to pay for insurance is monotonically nondecreasing in the level of coverage. A more realistic example might impose strict concavity of the transformation function at all wealth levels. The implication would be that marginal willingness to pay for insurance is strictly decreasing in the level of coverage except for at those levels at which the optimal θ is changing, resulting in a discrete increase in marginal willingness to pay.

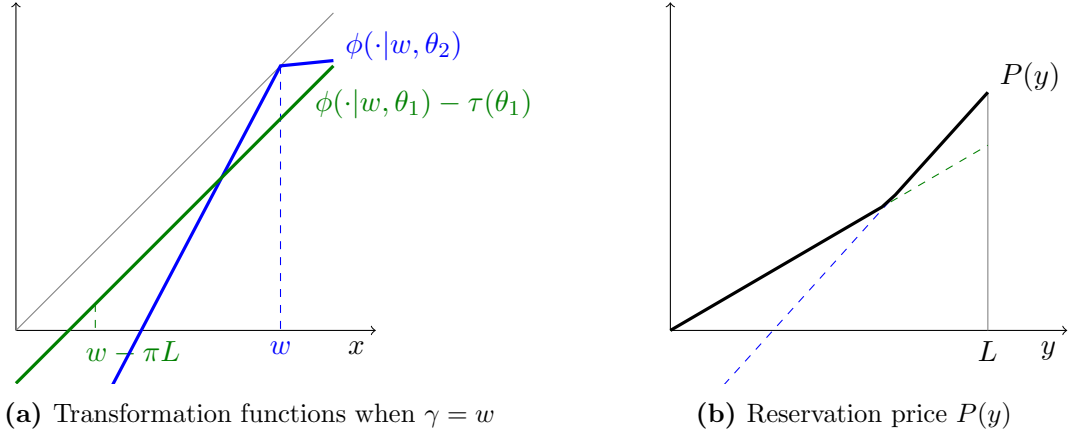


Figure 3. Reservation price for insurance from Example 3.

plausible, we are not aware of any existing models that can induce this demand pattern without simultaneously violating risk aversion: As we show formally in Section S.3 of the Supplementary Appendix, marginal willingness to pay for additional insurance coverage must be decreasing in the level of coverage for any risk preference that satisfies preference for diversification (quasiconcavity in random variables). Moreover, none of the non-expected-utility preferences commonly used in the literature can relax preference for diversification without also violating monotonicity with respect to second-order stochastic dominance.²³ It seems overly restrictive that these two facets of choice should be so tightly linked, especially in the context of insurance decisions where risk aversion plays such a prominent role.

Relaxing preference for diversification (without violating risk aversion) also plays a central role in our application to stock market participation in Section 4. In that section, we use an ORA representation with a certainty equivalent defined by Equations (7) and (11). As we show in Appendix A.2, this specification can equivalently be expressed as a generalization of Epstein-Zin-Kreps-Porteus expected utility:

$$V(c, m) = u(c) + \beta \sup_{\theta \in \Theta} \left\{ -\frac{1}{\theta} \log (\mathbb{E}_m [\exp(-\theta V)]) - \tau(\theta) \right\}. \quad (12)$$

This connection will make it easy to compare our model to the benchmark of EZKP utility (the case where $\Theta = \{\theta\}$) and to show why the generalization to nonsingleton Θ is necessary for our results. This extension of expected utility is also motivated by experimental evidence (from decision problems with a finite outcome space) that finds

²³In contrast, since all of the transformation functions in Example 3 are concave, the preference respects SOSD. See Theorem S.1 in Section S.1 of the Supplementary Appendix for a generalization of this observation to any stochastic order.

the majority of the violations of the expected-utility axioms occur around the boundary of the probability simplex (see Harless and Camerer (1994) for a detailed discussion and an aggregation of numerous earlier studies). Taking $\Theta = \{\theta_L, \theta_H\}$ for $\theta_L < \theta_H$ and $0 = \tau(\theta_H) < \tau(\theta_L)$ in Equation (12) is consistent with this evidence: θ_L will be optimal for lotteries involving significant risk, whereas the individual will switch to θ_H for lotteries near the corners of the simplex.²⁴

3.3 Uniqueness and Comparative Risk Aversion

In this section, we describe the uniqueness properties of the ORA representation and provide a comparative measure of risk aversion.

Most of the elements of the representation will be identified either uniquely or up to an affine transformation. However, there is one technical issue associated with identifying the set of risk attitudes Φ in the representation. Since this set is subjective (i.e., unobserved), it is possible that there are some extremely pessimistic or risk averse transformations that are feasible for the individual but that she would never find optimal for any lottery. For example, if $\phi \leq \hat{\phi}$ (pointwise) for some $\hat{\phi} \in \Phi$, then it can never be determined from the individual's preferences whether or not ϕ is in fact feasible for the individual; since this transformation is dominated, her choices can be rationalized both by including ϕ in Φ and excluding it.

Due to the impossibility of identifying the exact set of feasible risk attitude transformations, it is natural to focus on one of two canonical sets of transformations: a minimal set in the sense that no transformations can be dropped from Φ without altering the implied ranking of some pair of lotteries, or a maximal set in the sense that no transformations can be added without altering the ranking of some pair of lotteries. The results in this section are based on the second approach and identify and compare maximal sets of transformations. This will permit a simple and intuitive characterization of comparative risk aversion whereby a less risk averse individual has a larger set of feasible transformations.²⁵

²⁴This special case of our representation is similar in spirit to, and shares some properties with, the u - v preferences studied by Neilson (1992), Schmidt (1998), and Diecidue, Schmidt, and Wakker (2004), which ascribe one utility function v to certain outcomes and another Bernoulli utility index u to risky outcomes. However, unlike u - v preferences, the representation in Equation (12) is continuous and respects first-order stochastic dominance.

²⁵Another reason for focusing on maximal sets is that there are some technical issues involved in trying to identify a minimal set of transformations, primarily due to the fact that the set of lotteries on an interval has an empty interior within the space of all signed measures. However, in Section S.1 of the Supplementary Appendix, we apply a result from Cerreia-Vioglio, Maccheroni, and Marinacci (2015) to characterize a set of transformations Φ that is minimal in the sense of admitting the smallest possible set of expected-utility preferences.

Definition 4 Let (V, u, Φ, β) be an optimal risk attitude representation. The *maximal extension* of Φ is the set Φ^* of all continuous and nondecreasing functions $\phi : [a, b] \rightarrow \mathbb{R}$ (where $a = \min V$ and $b = \max V$) such that, for any $m \in \Delta(D)$,

$$\int_D \phi(V(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}) \leq \sup_{\hat{\phi} \in \Phi} \int_D \hat{\phi}(V(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}).$$

The ORA representation is *maximal* if $\Phi = \Phi^*$.²⁶

The next result formalizes the uniqueness properties of the ORA representation.

Theorem 2 Two ORA 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

1. $u_2 = \alpha u_1 + \lambda(1 - \beta_1)$,
2. There exists a bijection $f : \Phi_1^* \rightarrow \Phi_2^*$ such that for any $\phi_1 \in \Phi_1^*$ and $\phi_2 = f(\phi_1)$,

$$\phi_2(\alpha x + \lambda) = \alpha \phi_1(x) + \lambda, \quad \forall x \in V_1(D).$$

These conditions imply that $V_2 = \alpha V_1 + \lambda$.

Theorem 2 shows that, modulo an affine transformation, two maximal ORA representations of the same preference must be identical. A direct proof of this result is not provided, since the theorem follows immediately by applying Theorem 3 below to two representations of the same preference.

We now turn to the comparative measure of risk aversion. A more risk averse individual is intuitively more prone to reject a temporal lottery in favor of a deterministic consumption stream, as the following definition from [Chew and Epstein \(1991\)](#) formalizes.²⁷

Definition 5 The relation \succsim_1 is *more risk averse* than \succsim_2 if, for any $(c, m) \in D$ and $\mathbf{c} = (c_0, c_1, c_2, \dots) \in C^{\mathbb{N}}$,

$$(c, m) \succsim_1 \mathbf{c} \implies (c, m) \succsim_2 \mathbf{c}.$$

²⁶An alternative definition of the maximal extension is also possible. It is a standard result that a set of linear functions generating a convex function can be made maximal by taking its closed, convex, comprehensive hull (i.e., the smallest superset that is closed, convex, and contains all pointwise dominated functions). For example, see Theorem 3 in [Machina \(1984\)](#).

²⁷It is immediate from the construction in [Epstein and Zin \(1989\)](#) that $C^{\mathbb{N}}$ can be embedded as a subset of D .

The following result specializes the characterization of comparative risk aversion from [Chew and Epstein \(1991\)](#) for Epstein-Zin representations to the ORA representation.

Theorem 3 *Suppose the relations \succsim_1 and \succsim_2 have ORA representations $(V_1, u_1, \Phi_1, \beta_1)$ and $(V_2, u_2, \Phi_2, \beta_2)$, respectively. Then \succsim_1 is more risk averse than \succsim_2 if and only if $\beta_1 = \beta_2$ and there exist scalars $\alpha > 0$ and $\lambda \in \mathbb{R}$ such that*

1. $u_2 = \alpha u_1 + \lambda(1 - \beta_1)$,
2. There exists an injection $f : \Phi_1^* \rightarrow \Phi_2^*$ such that, for any $\phi_1 \in \Phi_1^*$ and $\phi_2 = f(\phi_1)$,

$$\phi_2(\alpha x + \lambda) = \alpha \phi_1(x) + \lambda, \quad \forall x \in V_1(D).$$

These conditions imply $V_2 \geq \alpha V_1 + \lambda$, with equality for deterministic consumption streams.

This result shows that, modulo an affine transformation, the maximal extension of the set of transformations Φ_1^* of a more risk-averse individual must be a subset of the set Φ_2^* of the less risk-averse individual. Intuitively, having more ways of tailoring one's risk attitude to the lottery being faced decreases risk aversion.

Theorem 3 makes it easy to compare many parametric special cases. For example, holding fixed u and β , if each feasible transformation in Φ_1 is bounded above (pointwise) by some transformation in Φ_2 , then the maximal extension for the first set is a subset of that of the second, $\Phi_1^* \subset \Phi_2^*$. The following corollary lists the implications of this observation for some of the examples considered in [Section 3.2](#).

Corollary 1 *Suppose the relations \succsim_1 and \succsim_2 have ORA representations $(V_1, u_1, \Phi_1, \beta_1)$ and $(V_2, u_2, \Phi_2, \beta_2)$, respectively, and suppose that $u_1 = u_2$ and $\beta_1 = \beta_2$.*

1. *If Φ_2 contains the identity function, $\phi(x) = x$, then \succsim_1 is more risk averse than \succsim_2 . That is, any ORA representation is more risk averse than time-separable expected utility.²⁸*
2. *Suppose as in the certainty equivalent in [Equation \(9\)](#) that $\Phi_i = \{\phi(\cdot|\gamma, \theta_i) : \gamma \in \Gamma\}$, where $\phi(x|\gamma, \theta)$ satisfies [Equation \(6\)](#). Then \succsim_1 is more risk averse than \succsim_2 if and only if $\theta_1 \geq \theta_2$.*
3. *Suppose as in the certainty equivalent in [Equation \(11\)](#) that $\Phi_i = \{\phi(\cdot|\gamma, \theta) - \tau_i(\theta) : \gamma \in \Gamma, \theta \in \Theta\}$, where $\phi(x|\gamma, \theta)$ satisfies [Equation \(6\)](#) and $\inf_{\theta \in \Theta} \tau_i(\theta) = 0$. If $\tau_1(\theta) \geq \tau_2(\theta)$ for all $\theta \in \Theta$, then \succsim_1 is more risk averse than \succsim_2 .*

²⁸If Φ_2 contains the identity function, then $V_2(c, m) = u_2(c) + \beta_2 \mathbb{E}_m[V_2]$.

4 Heterogeneous Stock Market Participation

4.1 Empirical Facts and Existing Explanations

In this section, we show that the optimal risk attitude representation may help to shed some new light on several puzzling patterns related to household stock market participation and asset allocation decisions, as well as equilibrium asset prices. We begin by describing two key empirical facts, explaining how they will be accommodated in our model, and giving an overview of the successes (and shortcomings) of existing models in addressing these facts.

FACT 1: Many households have limited or no participation in equity markets.

FACT 2: Although participation is positively correlated with wealth, a nontrivial fraction of wealthy households hold little or no public (or private) equity.

There is a large literature in household finance documenting these and other patterns in stock market participation (e.g., [Mankiw and Zeldes \(1991\)](#); [Haliassos and Bertaut \(1995\)](#); [Heaton and Lucas \(2000\)](#)). Importantly, [Heaton and Lucas \(2000\)](#) showed that while private business assets substitute for public equity for some households, around 10% of wealthy households hold neither. As noted by [Campbell \(2006, page 1564\)](#), “Limited participation among the wealthy poses a significant challenge to financial theory and is one of the main stylized facts of household finance.” It is true that there has been an increase in the level of stock market participation in recent years, primarily in the form of retirement savings and investment in mutual funds. Even so, as of 2007 only 51.5% of US households had any direct or indirect holdings of stock, and the average share of financial assets held in equity was 52.7% ([Guiso and Sodini \(2013, Tables 1 and 2\)](#)).

In the following sections, we will examine these facts through the lens of the optimal risk attitude representation. We will consider an economy with a continuum of agents with identical preferences that are homogeneous in wealth. Using the utility specification described in Equations (7) and (11), we find that heterogeneity in risk exposure arises endogenously in the dynamic general equilibrium of this economy. The intuition behind our results is relatively straightforward: When agents can optimize over their risk attitudes (subject to some mental cost), they endogenously sort into different types in equilibrium. One segment of the population will choose a risk attitude that is less sensitive to gains and losses and therefore will hold greater consumption risk; the other will choose a risk attitude that is more sensitive to losses and will hold very little consumption risk. Since mentally preparing for greater risk exposure comes at a psychological cost in our representation, inducing a fraction of the population to hold a large amount of risk requires

significant compensation in the form of a larger expected return. In this way, our model will generate both heterogeneous participation and a large equity premium.

The “participation puzzle” described above has received significant attention in recent years, and a number of potential explanations have been proposed. We summarize them below and discuss the similarities and differences from our approach.

EXPLANATION 1: Expected utility with heterogeneity in risk aversion.

Heterogeneity in risk aversion in the population will obviously generate differences in the level of investment across agents. However, closely related to the equity premium puzzle (Mehra and Prescott (1985)), the question is whether the levels of risk aversion required to generate low levels of investment by a large fraction of the population are reasonable. The consumption of stockholders is more volatile and more correlated with stock market returns than that of nonstockholders, and hence the equity premium becomes less of a puzzle when attention is restricted to consumers who invest in the stock market (see Mankiw and Zeldes (1991); Attanasio, Banks, and Tanner (2002); Brav, Constantinides, and Geczy (2002); Vissing-Jørgensen (2002); Vissing-Jørgensen and Attanasio (2003)). At the same time, these papers observe that there remains a deeper puzzle of explaining why so many households hold little or no stock, as the level of risk aversion required to rationalize their choices would be even higher than estimates based on aggregate consumption.

EXPLANATION 2: Participation costs.

Participation costs, in the form of monetary expenses associated with investing or informational costs associated with choosing an optimal portfolio of stocks, can provide a partial resolution of the participation puzzle. A number of studies have found that plausible values of entry costs and ongoing participation costs can rationalize the non-participation decision of many households (e.g., Vissing-Jørgensen (2003); Gomes and Michaelides (2005)). Another important benefit of participation cost models is that they predict that participation increases with wealth. However, these models still fail to explain the lack of participation or minimal risk exposure of some wealthy households, as their benefit from equity investing would dwarf any reasonable value for the participation cost (Vissing-Jørgensen (2003); Briggs, Cesarini, Lindqvist, Östling (2015)).

We view our model as complementary to the participation-cost approach. Since the ORA representation used in this section is homogeneous in wealth, our predictions regarding heterogeneous participation in no way rely on wealth effects. In Section 4.5.1, we discuss how an extension of our model that combines mixture-averse preferences with modest participation costs can explain why risk exposure covaries positively with wealth,

yet even slight heterogeneity in costs or preferences in the population would lead some wealthy households to either not participate in the stock market or hold very conservative portfolios.

EXPLANATION 3: First-order risk aversion or ambiguity aversion.

Other preference-based models have been used to address the participation decisions of wealthy households. These models invoke first-order risk aversion as an explanation for why some households would avoid an actuarially favorable investment opportunity such as the stock market. For example, [Ang, Bekaert, and Liu \(2005\)](#) used the disappointment aversion model of [Gul \(1991\)](#) to study a dynamic asset allocation problem and determined the critical values of the disappointment aversion parameter that lead to nonparticipation. However, [Barberis, Huang, and Thaler \(2006\)](#) observed that the presence of background risk (e.g., uninsurable idiosyncratic income risk) makes it difficult for models of first-order risk aversion to explain nonparticipation in the stock market using reasonable parameter values.²⁹ Their suggested remedy was to assume loss aversion together with narrow framing of portfolio risk, meaning that gains or losses in the stock market are evaluated separately from overall consumption risk. In [Section 4.5.2](#), we discuss why our model is unlikely to be subject to their critique and thus narrow framing is not needed to generate low participation levels. To the contrary, we will show that existing results about the impact of background risk on the risk attitudes of expected-utility maximizers imply that moderate background risk would only serve to *decrease* the optimal level of investment of households in our model.

Ambiguity aversion has also been proposed as an explanation for nonparticipation in the stock market. It is well known from the work of [Dow and Werlang \(1992\)](#) and [Epstein and Wang \(1994\)](#) that ambiguity aversion can lead to portfolio inertia at the risk-free portfolio, and therefore heterogeneity in perceived ambiguity can generate nonparticipation by some agents. More recent work by [Epstein and Schneider \(2007\)](#) has explored how learning under ambiguity can influence stock market participation, and they showed that this mechanism may explain part of the increase in participation rates in recent years.³⁰ Ambiguity aversion and mixture aversion are in many ways complementary—as

²⁹[Barberis, Huang, and Thaler \(2006\)](#) showed that introducing background risk decreases the aversion to exposure to stock market risk for the disappointment aversion model, to the point where the disappointment aversion parameters required for nonparticipation in the stock market would also imply rejection of a 50-50 gamble with a loss of \$10,000 and a gain of \$20,000,000 at a wealth level of \$100,000. They argued that their negative result extends broadly to models that rely on first-order risk aversion to obtain nonparticipation. [Safra and Segal \(2008\)](#) made the related observation that Rabin’s paradox extends to a broad class of non-expected-utility preferences when sufficient independent background risk is introduced.

³⁰Another likely explanation for rising participation rates is a decrease in informational and monetary costs to investing (e.g., see the discussion in [Guiso and Sodini \(2013, page 1454\)](#)).

was the case with participation costs—in that each is ideally suited to address different aspects of household behavior. For example, ambiguity aversion provides a sensible rationale for underdiversification and the home bias (e.g., [Epstein and Miao \(2003\)](#); [Boyle, Garlappi, Uppal, and Wang \(2012\)](#)), neither of which has an obvious connection to our model. On the other hand, the optimal risk attitude representation can generate low levels of stock market participation and low portfolio weights on stock for many participating households even in familiar environments or after long histories, where ambiguity would arguably play a less significant role in investment decisions.

Due to the complexity of dynamic models with heterogeneous preferences, the extant literature on preference-based explanations of the participation puzzle has been restricted almost exclusively to partial equilibrium analysis, taking the data-generating process for asset returns as given and solving for the participation and asset allocation decisions of a single agent. The few papers that have conducted general equilibrium analysis (typically within a one or two period framework) have reported somewhat mixed results. For example, [Cao, Wang, and Zhang \(2005\)](#) showed that increasing the ambiguity dispersion among investors leads simultaneously to a decrease in the participation rate and a decrease in the equity premium.³¹ [Chapman and Polkovnichenko \(2009\)](#) observed similar implications for increases in the dispersion of risk aversion parameters for a variety of non-expected-utility preferences (including disappointment aversion and rank-dependent utility). These results suggest that more research is needed to determine the suitability of these models for jointly explaining the participation rate and asset returns.

This is perhaps the most important point of departure of the ORA model from the previous literature: It can generate equilibrium heterogeneity in participation decisions even when agents have identical and homogeneous preferences. Although our model will not permit representative agent analysis (see [Section 4.3](#)), these conditions imply that equilibrium analysis does not require tracking the distribution of wealth across agents. Consequently, the model is tractable enough for us to conduct a complete dynamic general equilibrium analysis of both market prices and the distribution of participation decisions, taking only the dividend and consumption processes as primitives. This makes it possible to easily evaluate the performance of the ORA representation as a model of both nonparticipation and asset pricing.

4.2 Setting and Risk Preferences

Consider an economy in which aggregate consumption growth follows a first-order Markov process. The current state of the economy $z_t \in Z$ is observed by each individual in the

³¹The intuition behind their result is that greater dispersion in ambiguity implies there is a subset of the population that becomes much more tolerant of ambiguous payoffs and is therefore willing to invest more heavily in stock even at a lower premium.

economy at the start of period t , where Z is finite. We assume that the state is i.i.d. across time, and is distributed according to the probability measure P . Thus the model precludes intertemporal correlation between current and future states. This assumption is certainly restrictive, and is made in order to simplify the analysis and focus attention on the endogenous heterogeneity in risk bearing that arises in equilibrium. A natural next step for future study is to incorporate persistence of shocks and other types of correlation across time into this analysis.

Denote the aggregate consumption endowment by e_t , and let λ_{t+1} denote aggregate consumption growth between period t and $t + 1$: $e_{t+1} = \lambda_{t+1}e_t$. Consumption growth is determined by the state, i.e, $\lambda_{t+1} = \lambda(z_{t+1})$ for some function λ . We assume a continuum of consumers, with the consumption of consumer $i \in [0, 1]$ at time t denoted by $c_{i,t}$. The feasibility constraint in the economy requires that

$$\int_0^1 c_{i,t} di = e_t, \quad \forall t.$$

The distribution of ownership of the endowment allocation is not important for any of the results that follow, as long as it is absolutely continuous with respect to the Lebesgue measure on the unit interval of consumers.

Suppose that markets are complete and there exists a pricing kernel $M_{t,t+1}$ such that the period t price of any random asset paying x_{t+1} in period $t + 1$ is $p_t = \mathbb{E}_t[M_{t,t+1}x_{t+1}]$. Given the independence of consumption growth across time, one might conjecture that the pricing kernel is stationary and depends only on the period $t + 1$ state: $M_{t,t+1} = M(z_{t+1})$ for some function M . We will verify that this indeed the case in equilibrium. Under these assumptions, the time t budget constraint for a consumer with wealth w_t is then

$$c_t + \mathbb{E}[Mw_{t+1}] = w_t, \tag{13}$$

where $c_t \in \mathbb{R}_+$ and $w_{t+1} \in \mathbb{R}_+^Z$.

The consumers in the economy have mixture-averse preferences, where the instantaneous utility function in the optimal risk attitude representation is

$$u(c) = (1 - \beta) \log(c)$$

and the certainty equivalent is given by Equations (7) and (11). Since we are now considering random variables (rather than probability distributions), we change notation slightly to avoid confusion and write \mathcal{R} (rather than W) to denote the operator mapping from random variables to their certainty equivalents. Each consumer therefore has the

following value function for wealth w_t , conditional on the pricing kernel M :

$$\mathcal{V}(w_t; M) = \max_{\substack{c_t \in \mathbb{R}_+ \\ w_{t+1} \in \mathbb{R}_+^Z}} \left\{ (1 - \beta) \log(c_t) + \beta \mathcal{R}(\mathcal{V}(w_{t+1}; M)) \right\}, \quad (14)$$

where the maximization is subject to the budget constraint in Equation (13), and

$$\mathcal{R}(\mathcal{V}(w_{t+1}; M)) = \max_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}} \mathbb{E} \left[\phi(\mathcal{V}(w_{t+1}; M) | \gamma, \theta) - \tau(\theta) \right]$$

for

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

Recall also that $\min_{\theta \in \Theta} \tau(\theta) = 0$. This is a necessary restriction for \mathcal{R} to be a certainty equivalent.³²

To facilitate the interpretation of our results relative to the existing literature, it is useful to note an equivalent formulation of the model. If we let $h_\theta(x) = -\exp(-\theta x)$ for $\theta \in \Theta$, then Corollary 2 in Appendix A.2 implies that we can express \mathcal{R} as

$$\begin{aligned} \mathcal{R}(\mathcal{V}(w_{t+1}; M)) &= \max_{\theta \in \Theta} \left\{ h_\theta^{-1}(\mathbb{E}[h_\theta(\mathcal{V}(w_{t+1}; M))]) - \tau(\theta) \right\} \\ &= \max_{\theta \in \Theta} \left\{ -\frac{1}{\theta} \log(\mathbb{E}[\exp(-\theta \mathcal{V}(w_{t+1}; M))]) - \tau(\theta) \right\}. \end{aligned} \quad (15)$$

Note that Equation (15) reduces to the standard Epstein-Zin-Kreps-Porteus (EZKP) certainty equivalent if $\Theta = \{\theta\}$ and $\tau(\theta) = 0$. The equilibrium analysis in Section 4.4 will include the special case of EZKP utility for comparison. Our results will show that the maximization with respect to θ leads to several significant differences from EZKP preferences: First, having multiple θ permits individuals to be very averse to small gambles while at the same time exhibiting reasonable risk attitudes for larger gambles. Second, equilibrium in this economy often involves heterogeneity in the choice of θ and hence differences in consumption risk across consumers, even when they are ex-ante identical.

Standard techniques can be applied to prove the existence and uniqueness of the value

³²One could obtain similar heterogeneity in participation by using the kinked transformation functions from Equation (8). This alternative specification could be useful for incorporating first-order risk aversion. However, our choice of certainty equivalent is convenient for several reasons: First, it shows that first-order risk aversion is not needed to generate a fraction of the population that holds minimal risk exposure. Second, it is smooth and therefore permits techniques based on differentiation. Third, it allows the model to make realistic predictions about the impact of incorporating uninsurable idiosyncratic income risk (see Section 4.5.2).

function for this problem. Moreover, the value function takes the form

$$\mathcal{V}(w_t; M) = \Lambda(M) + \log(w_t) \tag{16}$$

for a constant $\Lambda(M)$ that depends on the pricing kernel. Additional details and a proof of these claims can be found in Section S.4.1 of the Supplementary Appendix.

4.3 Illustration of Equilibrium

In this section, we provide an informal description of the three possible types of equilibrium that can arise in this model. For ease of illustration, the discussion will initially focus on equilibrium in a static (one-period) model. We then proceed to describe how equilibrium in our infinite-horizon model has a similar structure. Formal statements of the equilibrium conditions for the three cases are relegated to Section S.4.2 in the Supplementary Appendix, which also contains a summary of the numerical procedure that will be used in Section 4.4.

4.3.1 Benchmark: Static Model

For ease of illustration, we first consider a simple model with a single period, two states, and two types. This section will therefore focus on the following simple specification:

$$\begin{aligned} Z &= \{z^l, z^h\} & \Theta &= \{\theta_L, \theta_H\}, \quad \theta_H > \theta_L \\ e(z^l) &< e(z^h) & \tau(\theta) &= \begin{cases} \tau_H = 0 & \text{if } \theta = \theta_H \\ \tau_L > 0 & \text{if } \theta = \theta_L. \end{cases} \end{aligned}$$

We have dropped time subscripts for the one-period analysis in this section. Assume each of the continuum of consumers has equal (unit) ownership of the endowment (this assumption is made for ease of exposition but is not imposed in the analysis in the following section). Thus a consumption allocation $c = (c(z^l), c(z^h))$ is affordable given the pricing kernel M if

$$\mathbb{E}[Mc] = \mathbb{E}[Me].$$

Assume each consumer evaluates a random consumption allocation c by applying the certainty equivalent in Equation (15) to the logarithm of $c(z)$:³³

$$\mathcal{R}(\log(c)) = \max_{\theta \in \{\theta_L, \theta_H\}} \left\{ \log \left(\mathbb{E}[c^{-\theta}]^{-\frac{1}{\theta}} \right) - \tau(\theta) \right\}.$$

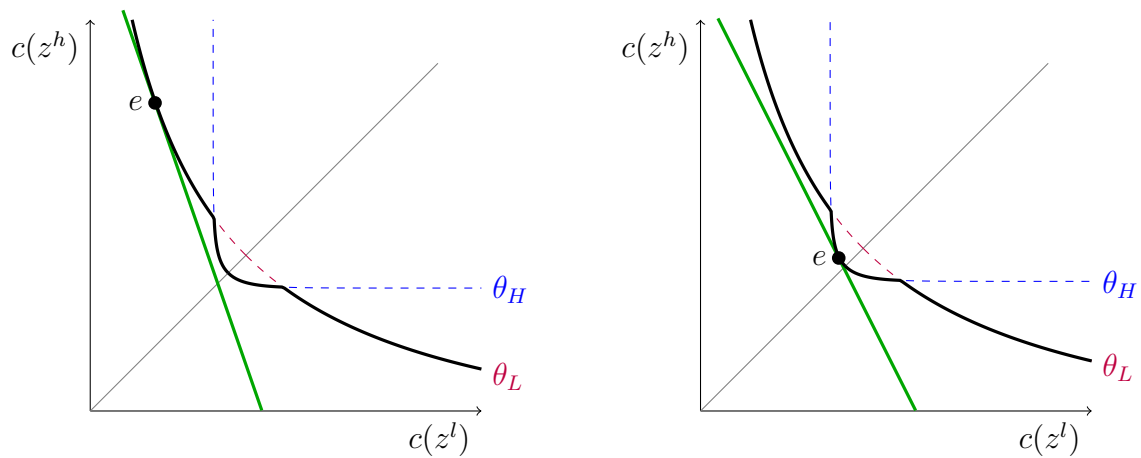
³³To make this analysis as representative as possible of the infinite-horizon model, we treat consumption just as we would treat future wealth in the general model. Given the log form of the value function in Equation (16), we therefore apply the certainty equivalent \mathcal{R} to $\log(c)$ rather than to c .

Thus, for fixed θ , each consumer evaluates uncertain consumption using a CRRA certainty equivalent with a coefficient of relative risk aversion of $\theta + 1$. However, since θ can vary with the risk being faced, this preference will violate the expected-utility axioms. Intuitively, for small gambles it will be optimal to choose θ_H in order to avoid the utility cost $\tau_L > 0$; when faced with larger risks, a consumer may then find it optimal to select θ_L to decrease the sensitivity to this more uncertain outcome.

Figure 4 illustrates the indifference curves of a consumer with the risk preferences described above. Note that θ_H is optimal for allocations near the certainty line, whereas θ_L is optimal for allocations where the level of consumption differs greatly between states. This figure also illustrates the equilibrium allocations and budget lines for three possible cases. In cases 1 and 2, all consumers choose the same type in equilibrium (θ_L and θ_H , respectively) and consume their endowment: $c(z) = e(z)$ for $z \in \{z^l, z^h\}$. In these first two cases, the equilibrium allocation and prices are the same for our economy with a continuum of consumers as they would be for an economy with a single representative agent with the same preferences.

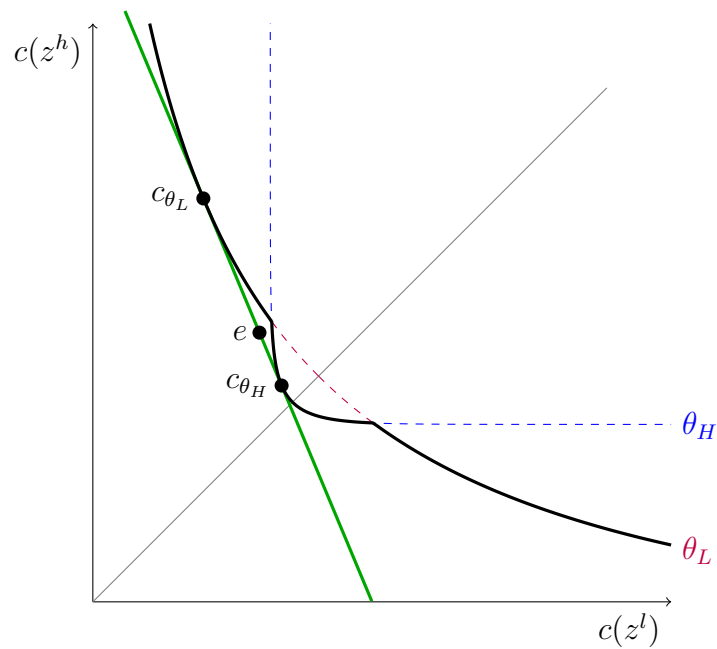
The distinctive part of our analysis arises in case 3, where equilibrium necessarily involves heterogeneous types, with some consumers selecting θ_H and others θ_L . As is evident from Figure 4, preferences are not quasiconcave in the consumption allocation, and therefore equilibrium may not exist in a representative agent economy (non-existence would be an issue precisely in case 3). However, for our economy with a continuum of consumers, the theorem of [Aumann \(1966\)](#) ensures the existence of an equilibrium.³⁴ The crux of his theorem is that the average (i.e., integral) of the (possibly non-convex) upper-contour sets of individual preferences is convex. The economic interpretation of this mathematical condition is central to our application: Rather than every consumer keeping their endowment, there may be welfare improvements associated with heterogeneous allocations that lead to the same average aggregate consumption. Equilibrium in case 3 takes this form, with consumers who select type θ_L choosing allocation c_{θ_L} and consumers who select type θ_H choosing allocation c_{θ_H} . Note that these two allocations give the same utility, which is strictly higher than the utility from consuming the endowment e . The fraction of consumers selecting each of these types and allocations is determined by the market-clearing condition: $e = \alpha c_{\theta_L} + (1 - \alpha)c_{\theta_H}$ where α is the fraction of the population that selects type θ_L .

³⁴A related result for a large but finite economy in which consumers' preferences are permitted to violate quasiconcavity can be found in [Starr \(1969\)](#), who showed that divergence from equilibrium shrinks with the number of consumers.



(a) Case 1: All consumers choose type θ_L

(b) Case 2: All consumers choose type θ_H



(c) Case 3: Heterogeneous types

Figure 4. Three possible cases of equilibrium illustrated for a static model with two states. In cases 1 and 2, equilibrium consumption is identical for each consumer and proportional to the aggregate endowment. In case 3, type θ_L consumers choose allocation c_{θ_L} and type θ_H consumers choose allocation c_{θ_H} . The fraction of consumers selecting each type is determined by the market-clearing condition.

4.3.2 Equilibrium in the Infinite-Horizon Model

Equilibrium in the infinite-horizon model will also fall into one of the three cases illustrated informally for the static model: In each period, consumers will either all select the same type in equilibrium (θ_L or θ_H), or there will be heterogeneity in the choice of type. We defer the details of how to solve for the pricing kernel and check the market-clearing conditions in each of these cases to Section S.4.2 in the Supplementary Appendix, but we present the numerical results of our equilibrium analysis in the following section.

4.4 Calibration

In this section, we numerically solve the infinite-horizon model. We will highlight several key features of individual risk attitudes and market equilibrium for various parameter values: heterogeneity in market participation and risk exposure, asset returns, and attitudes toward small and large idiosyncratic gambles.

4.4.1 Parameter Values

Recall that aggregate consumption growth is determined by a state which is i.i.d. across time. Assume there are two states, $Z = \{z^l, z^h\}$, and each occurs with equal probability: $P(z^l) = P(z^h) = 0.5$ (we relax this assumption in Section 4.4.3). Aggregate consumption satisfies $e_{t+1} = \lambda(z_{t+1})e_t$. For ease of comparison to existing results, we calibrate the model using the same mean and standard deviation for aggregate consumption growth as in Mehra and Prescott (1985):

$$\lambda(z^l) = \mu - \sigma, \quad \lambda(z^h) = \mu + \sigma$$

for $\mu = 1.018$ and $\sigma = 0.036$. We will also analyze the returns of an asset that pays a stream of dividends $\{d_t\}$. The growth rate of dividends satisfies $d_{t+1} = \lambda^d(z_{t+1})d_t$, where

$$\lambda^d(z^l) = \mu - \sigma_d, \quad \lambda^d(z^h) = \mu + \sigma_d$$

for $\sigma_d = 0.10$.

There is a continuum of consumers with identical preferences. Each consumer has a value function $\mathcal{V}(w_t; M)$ for wealth (conditional on the pricing kernel) that satisfies

Equations (13), (14), and (15). We consider the following simple specification:

$$\Theta = \{\theta_L, \theta_H\}, \quad \theta_H > \theta_L$$

$$\tau(\theta) = \begin{cases} \tau_H = 0 & \text{if } \theta = \theta_H \\ \tau_L > 0 & \text{if } \theta = \theta_L. \end{cases}$$

Several values of θ_L , θ_H , and τ_L will be considered in Table 1.

As part of the calibration of the model, we also describe the attitudes toward 50-50 gambles of various scales for the different parameter values under consideration. Examining the gains needed to compensate for losses ranging from small to large provides another gauge of the overall risk attitudes associated with different specifications, and hence of whether parameter values generate reasonable behavior outside of this specific investment application.³⁵ Similar explorations have been used to evaluate the choice of parameters in other models (e.g., Epstein and Zin (1990); Kandel and Stambaugh (1991)).

Imagine a consumer is offered a one-time gamble over future wealth at some initial period t . Inserting the explicit formula for the value function from Equation (16) into Equation (15), the consumer evaluates the random future wealth w_{t+1} resulting from this gamble according to

$$\mathcal{R}(\mathcal{V}(w_{t+1}; M)) = \Lambda(M) + \max_{\theta \in \Theta} \left\{ \log \left(\mathbb{E}[w_{t+1}^{-\theta}]^{-\frac{1}{\theta}} \right) - \tau(\theta) \right\}. \quad (17)$$

We will use Equation (17) to evaluate atemporal wealth gambles in the calibration results that follow.³⁶

4.4.2 Numerical Results

Table 1 summarizes our results. For comparison, the first two columns of the table describe EZKP utility with coefficients of relative risk aversion ($\theta_H + 1$) equal to 4 and 18, respectively. The last three columns describe different parameter values for the optimal risk attitude representation.

Panel A in the table describes the gains needed for an individual to accept an atemporal 50-50 gamble for various possible loss values when initial wealth is \$300,000. Panel B describes the consumption growth rates of each type $\theta \in \{\theta_L, \theta_H\}$ that is selected by

³⁵There are many estimates of what constitute reasonable values for the coefficient of relative risk aversion in a CRRA expected-utility model; however, one cannot rely on these estimates since the risk attitude of individuals with ORA preferences may change with their exposure to risk.

³⁶Epstein and Zin (1989, Section 5) were the first to demonstrate that, assuming the stochastic process driving the economy is i.i.d. across time, the certainty equivalent in a recursive non-expected-utility model also represents the preferences over timeless wealth gambles.

some segment of the population in equilibrium. We write $\lambda_\theta(z)$ to denote the consumption growth rate of type θ consumers in the current period: $\lambda_\theta(z) = c_{i,t+1}(z)/c_{i,t}$ for any consumer i who selects type θ in period t . In addition to describing the mean and standard deviation of consumption growth for each type, the last row of this panel indicates the fraction of current aggregate wealth held by consumers who choose type θ_L or, equivalently, the fraction of the current period aggregate endowment consumed by type θ_L consumers.³⁷

Panel C indicates the equilibrium values of the pricing kernel and asset returns. In this panel, R^f denotes the gross risk-free rate and R denotes the gross return on an asset paying the stochastic dividend stream $\{d_t\}$. While not the main focus of the calibration exercise, the results in this panel illustrate that the model generates a moderate equity premium for reasonable parameter values.

Turning to the first specification in the table, note that EZKP1 generates reasonable aversion to the largest gambles in Panel A, but it is almost risk neutral for small gambles. It also only generates an equity premium of 1.4%. In contrast, specification EZKP2 assumes a coefficient of relative risk aversion of 18 and generates a more realistic equity premium of 5.9%.³⁸ However, the gambling behavior for this specification highlights the concerns about assuming such a large coefficient of risk aversion. The individual will reject any gamble in which she will lose over \$20,000 (7% of wealth) with even odds, regardless of the size of the possible gain that could be won.³⁹ These specifications of EZKP utility provide a parametric illustration of a paradoxical implication of expected utility that was shown by [Rabin \(2000\)](#) to hold more generally: Any expected-utility preference must have either implausibly low aversion to small gambles or excessive aversion to large gambles.

Specification ORA1 improves the risk attitudes for binary gambles over both EZKP1 and EZKP2: It increases the risk aversion for small gambles relative to these specifications, while drastically reducing the extreme risk aversion of EZKP2 for large gambles. At the same time, it maintains a moderately large equity premium of 4.4%.⁴⁰

³⁷Just as in the single-period illustration from Section 4.3.1, in a heterogeneous-type equilibrium in the infinite-horizon economy, the fraction α of aggregate wealth held by consumers who choose type θ_L is pinned down by the market clearing condition, in this case applied to consumption growth rates rather than levels: $\lambda(z) = \alpha\lambda_{\theta_L}(z) + (1 - \alpha)\lambda_{\theta_H}(z)$ for $z \in \{z^l, z^h\}$. See Section S.4.2 in the Supplementary Appendix for details.

³⁸It is well known that EZKP utility with a high degree of risk aversion can generate a large equity premium. The calibration in specification EZKP2 is consistent with the estimate from [Kocherlakota \(1996, page 51\)](#) that a coefficient of relative risk aversion of 17.95 will satisfy the Euler equation for the equity premium.

³⁹Similar observations related to large-scale idiosyncratic risks (e.g., occupational earnings risk) led [Mehra and Prescott \(1985\)](#), [Lucas \(2003\)](#), and others to argue that the coefficient of relative risk aversion should be bounded above by 10.

⁴⁰One caveat in interpreting these results is that the simple two-state stochastic process in our analysis implies perfect correlation between consumption and asset returns. Imposing more realistic correlation between aggregate consumption and returns tends to lower the equity premium in most models.

Table 1. Calibration Results: EZKP and ORA Models

The table has results for two specifications of the EZKP model and three specifications of the ORA model. Consumption and dividend growth rates are summarized in Section 4.4.1, R^f denotes the risk-free rate, and R denotes the gross return on asset paying the stochastic dividend stream $\{d_t\}$.

Risk-Preference Model					
	EZKP1	EZKP2	ORA1	ORA2	ORA3
θ_H	3.000	17.000	25.000	25.000	100.000
θ_L	–	–	3.000	4.000	3.000
τ_L	–	–	0.020	0.025	0.020
β^{-1}	1.010	1.010	1.010	1.010	1.010

Panel A: Binary 50-50 Gambles

Loss	Gain that leads to indifference for initial wealth \$300,000				
\$100	100.13	100.60	100.87	100.87	103.48
\$400	402.14	409.84	414.37	414.37	462.37
\$1,000	1,013.51	1,063.85	1,094.95	1,094.95	1,518.31
\$2,000	2,054.80	2,273.08	2,420.72	2,420.72	9,254.96
\$5,000	5,357.20	7,170.61	8,995.81	8,995.81	18,991.43
\$10,000	11,539.60	27,901.22	26,396.79	32,281.88	26,396.79
\$20,000	27,302.60	∞	45,692.48	58,228.29	45,692.48
\$30,000	50,274.57	∞	75,052.03	110,405.61	75,052.03

Panel B: Equilibrium Consumption Growth by Type

$\lambda_{\theta_H}(z^l)$	0.9820	0.9820	0.9927	0.9880	1.0011
$\lambda_{\theta_H}(z^h)$	1.0540	1.0540	1.0277	1.0342	1.0095
$\mathbb{E}(\lambda_{\theta_H})$	1.0180	1.0180	1.0102	1.0111	1.0053
$\sigma(\lambda_{\theta_H})$	0.0360	0.0360	0.0175	0.0231	0.0042
$\lambda_{\theta_L}(z^l)$	–	–	0.9346	0.9399	0.9374
$\lambda_{\theta_L}(z^h)$	–	–	1.1709	1.1919	1.1576
$\mathbb{E}(\lambda_{\theta_L})$	–	–	1.0527	1.0659	1.0475
$\sigma(\lambda_{\theta_L})$	–	–	0.1182	0.1260	0.1101
% type θ_L	–	–	18.36	12.55	30.04

Panel C: Pricing Kernel and Asset Returns

$M(z^l)$	1.1149	1.5508	1.4047	1.5192	1.3796
$M(z^h)$	0.8400	0.4339	0.5700	0.4633	0.5934
R^f	1.0231	1.0077	1.0128	1.0088	1.0137
$\mathbb{E}(R)$	1.0374	1.0667	1.0567	1.0645	1.0550
$\sigma(R)$	0.1019	0.1048	0.1038	0.1046	0.1036
$\mathbb{E}(R) - R^f$	0.0143	0.0590	0.0439	0.0557	0.0413

Panel B illustrates the heterogeneity in consumption risk for the two segments of the population under ORA1. Just over 18% of the aggregate wealth in the population is held by consumers who choose type θ_L . The consumption growth rate λ_{θ_L} for this fraction of the population has a standard deviation of 0.1182, over three times that of aggregate consumption growth. The majority of the population chooses type θ_H , and the consumption growth rate λ_{θ_H} for this segment of the population has a standard deviation of 0.0175, roughly half that of aggregate consumption. Thus a significant portion of aggregate risk is consolidated in a small segment of the population. Moreover, the additional risk borne by type θ_L consumers is compensated by a substantially higher expected consumption growth rate—1.05 (5%) rather than the expected consumption growth rate of 1.01 (1%) for type θ_H consumers.⁴¹

These results are consistent with the observed patterns of investment summarized in Section 4.1. There has traditionally been a small segment of the population that invests in the stock market, with the majority of the population investing primarily in low-risk and low-return savings instruments. We should be careful to point out that our model predicts that even type θ_H consumers continue to bear some consumption risk. However, as we discuss in Section 4.5.1, since the utility difference between the equilibrium allocation with minimal risk exposure and one with no exposure to stock market risk is relatively small, introducing modest participation costs could drive some of these consumers completely out of the market—even those with relatively high wealth.

The last two columns illustrate the impact of changing parameters in the ORA representation. Specification ORA2 increases the values of θ_L and τ_L . The result is an increase in the equity premium, up to 5.6%, and a decrease in the fraction of the population choosing type θ_L and holding greater consumption risk, down to under 13%. For this specification, aversion to large gambles on the scale of \$30,000 (10% of wealth) becomes overly large, although still more reasonable than for EZKP2. Specification ORA3 illustrates the impact of instead increasing θ_H . This generates a high level of risk aversion for small-scale gambles, but the attitude toward very large gambles is the same as under ORA1. It is interesting to note that due to changes in equilibrium consumption heterogeneity, the equity premium actually decreases slightly for this specification.

4.4.3 Comparative Statics of Participation and Risk Sharing

We now examine how risk sharing in the population responds to changes in fundamentals. There are many possible changes to aggregate uncertainty that have been considered

⁴¹Cross-sectional differences in expected returns due to differences in portfolio allocations have been suggested as one potential driver of wealth inequality. However, our model lends a very different welfare interpretation to these differences, since the portfolio choices of type θ_H and θ_L consumers yield the same ex-ante utility.

in the literature, including changes to the conditional expectation and the conditional volatility of consumption growth. In this section, we continue to focus on a consumption growth process that is i.i.d. across time and examine a particular change to its (unconditional) distribution. To facilitate comparison to existing results, we maintain the same mean and standard deviation of consumption growth and only vary its skewness.

Let π denote the probability of state z^l . In the analysis in Table 1, this value was fixed at $\pi = 0.5$. We now consider the impact of varying this probability while holding fixed the first and second moments of consumption and dividend growth. Specifically, growth rates take the following values:

$$\begin{aligned} \lambda(z^l) &= \mu - \sigma \sqrt{\frac{1-\pi}{\pi}} & \lambda(z^h) &= \mu + \sigma \sqrt{\frac{\pi}{1-\pi}} \\ \lambda^d(z^l) &= \mu - \sigma_d \sqrt{\frac{1-\pi}{\pi}} & \lambda^d(z^h) &= \mu + \sigma_d \sqrt{\frac{\pi}{1-\pi}}. \end{aligned} \tag{18}$$

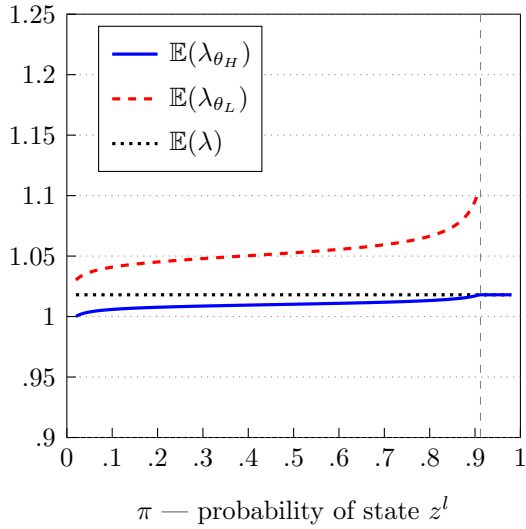
Using the parameter values from specification ORA1, Figure 5 illustrates the impact of changing skewness on the expected value and standard deviation of the consumption growth rates of each type, on the distribution of types in the population, and on asset returns. As π increases, aggregate consumption growth transitions from being skewed to the left to being skewed to the right. In response to this change, the fraction of aggregate wealth held by type θ_L consumers decreases while simultaneously the consumption risk held by these consumers increases. The standard deviation of λ_{θ_L} spikes to roughly 9 times that of aggregate risk as π approaches 0.91, and once π exceeds this threshold all consumers select type θ_H and hold identical portfolio allocations.

While these comparative statics concern changes in the unconditional distribution of an i.i.d. process, they are suggestive that the ORA model may provide a mechanism to understand intertemporal variation in investment levels and movements in and out of the stock market. If consumption growth follows a stationary Markov process with state-dependent mean, standard deviation, or skewness, one might expect the equilibrium participation rate to change with the current state. Extending our analysis to more general stochastic processes and relating its predictions to recent studies on the dynamics of household portfolio decisions is an obvious area for future exploration.

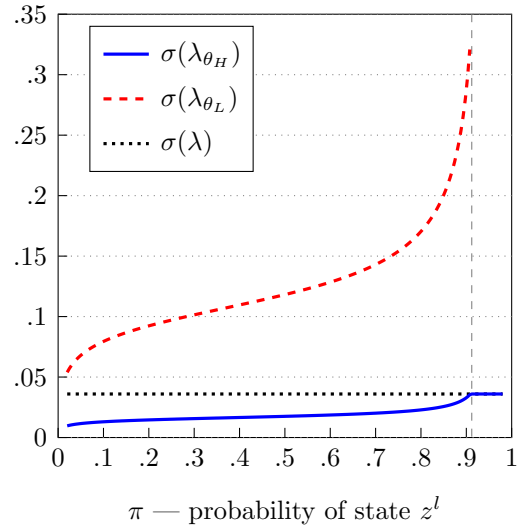
4.5 Discussion and Extensions

4.5.1 Participation Costs

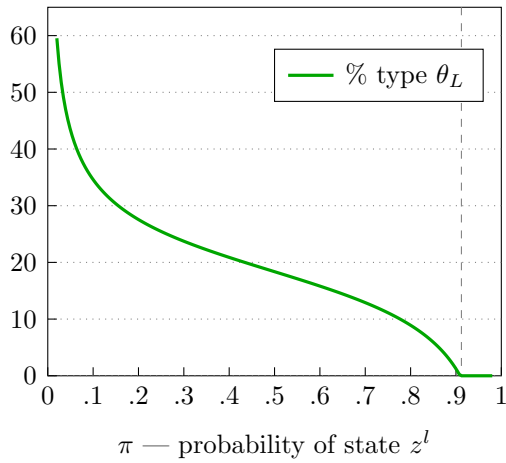
As we noted earlier, the combination of mixture-averse preferences with participation costs can be used to further refine the predictions of our model. The analysis in the



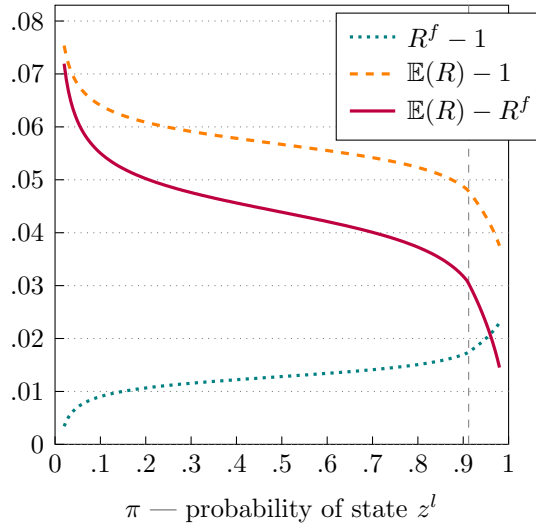
(a) Expected value of consumption growth rate by type



(b) Standard deviation of consumption growth rate by type



(c) Percent of aggregate wealth held by consumers choosing type θ_L in a given period



(d) Expected returns and risk premium

Figure 5. Expected value and standard deviation of consumption growth by type, division of population into types, and returns. Each is plotted as a function of the probability π of state z^l . Consumption and dividend growth rates satisfy Equation (18), and parameters values are set according to the ORA1 specification in Table 1.

previous section showed that consumers who select type θ_H in equilibrium invest significantly less in stocks, and their utility gain from these investments are relatively small given their high level of local risk aversion (recall that utility for low-risk allocations is equivalent to that of an EZKP utility maximizer with a coefficient of relative risk aversion of $\theta_H + 1$). This implies that the level of participation costs required to induce nonparticipation within the ORA model will be much lower than previous estimates based on expected-utility preferences.

An important consequence of this observation is that, even for wealthy households, much more moderate financial or informational costs can lead to nonparticipation, which allows our model to address one of the most puzzling facts from household finance. Moreover, since the required costs are small, slight cross-sectional variation in risk aversion or participation costs can generate the combination of complete nonparticipation by some households, small but positive investment in stocks by some households, and high levels of investment by others, where the portion of wealth invested in stocks is positively (but not perfectly) correlated with wealth. We leave the formal analysis of this extension as an important direction for future research.

4.5.2 Idiosyncratic Income Risk

As we discussed in Section 4.1, one difficulty for models that rely on first-order risk aversion to explain nonparticipation is that the introduction of background risk tends to significantly decrease the aversion to stock market risk for such preferences, making it harder to rationalize nonparticipation using reasonable parameter values. This observation motivated [Barberis, Huang, and Thaler \(2006\)](#) to incorporate narrow framing of stock market risk into their model of participation.

Interestingly enough, this property of models with first-order risk aversion differs dramatically from the implications of background risk for expected utility: For CRRA expected-utility preferences, the introduction of independent background risk only serves to increase effective risk aversion and hence to decrease the optimal investment in stocks (see [Gollier and Pratt \(1996\)](#)). In fact, the asset-pricing literature has used incomplete consumption insurance (in the form of persistent idiosyncratic income shocks) to generate *greater* aversion to stock market risk in order to help explain the level of the equity premium (see [Constantinides and Duffie \(1996\)](#); [Brav, Constantinides, and Geczy \(2002\)](#)).

Narrow framing is unlikely to be needed to generate realistic predictions with our model. Based on existing theoretical results for expected utility, it is easy to see that introducing small or moderate amounts of background risk into the ORA model will similarly increase aversion to stock market risk and decrease the optimal level of investment in stocks. Assuming the amount of background risk is not sufficiently large to change the optimal risk attitude θ , consumers in our model behave exactly like CRRA expected-utility

maximizers, and hence the comparative statics results in [Gollier and Pratt \(1996\)](#) apply. However, there is an important caveat to this argument: If the amount of background risk is sufficiently large, then consumers who would otherwise choose the risk attitude θ_H may switch to choosing θ_L , thereby decreasing their local risk aversion. Determining the relevant scale of uninsurable idiosyncratic risk in relation to our model is an empirical question that is beyond the scope of the present paper.

4.5.3 Optimal Expectations and Speculative Behavior

In the ORA representation, the individual optimizes her risk attitude for any given distribution of future outcomes. There is a literature that considers the dual problem of optimizing beliefs for a fixed utility function (e.g., [Brunnermeier and Parker \(2005\)](#); [Gollier and Muermann \(2010\)](#); [Bénabou and Tirole \(2011\)](#); [Macera \(2014\)](#)). The models in this literature usually predict distortions of future behavior in response to changes in current beliefs, or require distortions of current behavior as a form of self-signaling. These features are in sharp contrast to our model of dynamically-consistent choice.

The differences between these two approaches allows each to address slightly different questions. For example, [Brunnermeier and Parker \(2005\)](#) showed that optimal expectations can generate endogenous heterogeneity in investment behavior in equilibrium, where ex-ante identical consumers select opposing beliefs and take stock market positions that bet against each another. Optimal expectations can thus help to explain speculative investment behavior. Our application addresses a different type of heterogeneity, namely, greater risk bearing by one segment of the population. Exploring the interactions between optimal beliefs and optimal risk attitudes is a potentially interesting avenue for further research.

A Epstein-Zin and Kreps-Porteus Preferences

A.1 Epstein-Zin Representation Result

In this section, we provide an axiomatic characterization of the Epstein-Zin representation in Definition 2. As noted in the main text, the axioms in this section will parallel the treatment in Chew and Epstein (1991), but strengthen their separability assumption.

The first three axioms are entirely standard.

Axiom 2 (Weak Order) *The relation \succsim is complete and transitive.*

Axiom 3 (Nontriviality) *There exist $c, c' \in C$ and $m \in \Delta(D)$ such that $(c, m) \succ (c', m)$.*

Axiom 4 (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. It states that the preference between any pair of alternatives remains the same if those alternatives are pushed back one period into the future.

Axiom 5 (Stationarity) *For any $c, \hat{c}, \check{c} \in C$ and $\hat{m}, \hat{m}' \in \Delta(D)$,*

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

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 6 (Separability) *For any $c, c', \hat{c}, \check{c} \in C$ and $\hat{m}, \hat{m}' \in \Delta(D)$,*

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

Condition 1 in Axiom 6 says that the comparison of c today and \hat{c} tomorrow versus c' today and \check{c}' tomorrow is the same regardless of the lottery (\hat{m} or \hat{m}') following tomorrow's consumption. Likewise, condition 2 says that comparison of c today and lottery \hat{m} following tomorrow versus c' today and \hat{m}' following tomorrow is the same for any consumption tomorrow (\hat{c} or \check{c}'). Note that Axiom 6 only applies to temporal lotteries in which the one-step-ahead continuation is deterministic. Intuitively, in the case of deterministic consumption streams, Definition 2 reduces to a standard time-separable intertemporal utility function.

The next axiom ensures that preferences respect the first-order stochastic dominance order on $\Delta(D)$. Recall that in the case of monetary gambles, FOSD roughly corresponds to increasing the probability of better monetary outcomes. The same is true in this setting, with (\hat{c}', \hat{m}') being a better continuation path than (\hat{c}, \hat{m}) following current consumption c if and only if $(c, \delta_{(\hat{c}', \hat{m}')}) \succsim (c, \delta_{(\hat{c}, \hat{m})})$.⁴²

Axiom 7 (FOSD) For any $c \in C$ and $m, m' \in \Delta(D)$, if for all $(\hat{c}, \hat{m}) \in D$,

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

then $(c, m) \succsim (c, m')$.⁴³

The following result characterizes the Epstein-Zin representation from Definition 2.

Proposition 1 The relation \succsim satisfies Axioms 2–7 if and only if it has an Epstein-Zin representation (V, u, W, β) .

A.2 The Independence Axiom and EZKP Utility

It is immediate from the ORA representation that mixture-averse preferences are a generalization of time-separable expected utility—simply let Φ contain the identity mapping. Perhaps less obvious is the relationship with recursive [Kreps and Porteus \(1978\)](#) utility, formally defined as follows.

Definition 6 An *Epstein-Zin-Kreps-Porteus (EZKP)* representation is a tuple (V, u, h, β) consisting of a continuous function $V : D \rightarrow \mathbb{R}$ that represents \succsim , a continuous and nonconstant function $u : C \rightarrow \mathbb{R}$, a continuous and strictly increasing function $h : [a, b] \rightarrow \mathbb{R}$ (where $a = \min V$ and $b = \max V$), and a scalar $\beta \in (0, 1)$ such that, for all $(c, m) \in D$,

$$V(c, m) = u(c) + \beta h^{-1}(\mathbb{E}_m[h(V)]).$$

The commonly used (and empirically more relevant) case of EZKP utility is where h is a concave transformation, and therefore risk aversion is increased relative to time-separable expected utility. This case of EZKP utility will be shown to satisfy mixture aversion. In this section, we establish the relationship using the axioms as well as directly from the representations.

EZKP utility is the special case of the Epstein-Zin representation where the certainty equivalent takes the expected-utility form. It therefore satisfies a version of the independence axiom.

⁴²When preferences are continuous, this first-order stochastic dominance assumption is equivalent to the “recursivity” axiom that has appeared in various forms in the literature on dynamic preferences, including [Chew and Epstein \(1991\)](#): $(c, \delta_{(\hat{c}, \hat{m})}) \succsim (c, \delta_{(\hat{c}', \hat{m}')}) \iff (c, \alpha \delta_{(\hat{c}, \hat{m})} + (1-\alpha)m) \succsim (c, \alpha \delta_{(\hat{c}', \hat{m}')} + (1-\alpha)m)$.

⁴³Implicit in this axiom is the assumption that the set $\{(\hat{c}', \hat{m}') : (c, \delta_{(\hat{c}', \hat{m}')}) \succsim (c, \delta_{(\hat{c}, \hat{m})})\}$ is Borel measurable for each $(\hat{c}, \hat{m}) \in D$. However, if the continuity axiom is imposed, then each of these sets is closed and hence measurable.

Axiom 8 (Independence) For any $c \in C$, $m, m', m'' \in \Delta(D)$, and $\alpha \in (0, 1)$,

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

The following proposition characterizes the EZKP representation and shows the class of representations that are compatible with both independence and mixture aversion. The techniques needed for the first part of this result are essentially the same as those used by [Kreps and Porteus \(1978\)](#) in a finite horizon setting and [Chew and Epstein \(1991\)](#) for nonseparable preferences in the infinite horizon domain.

Proposition 2 Suppose \succsim has an EZ representation.⁴⁴ Then \succsim satisfies Axiom 8 if and only if it has an EZKP representation (V, u, h, β) . Moreover, \succsim also satisfies Axiom 1 if and only if h is concave.

Proposition 2 shows that risk-averse EZKP preferences (i.e., those represented using concave h) are a subset of mixture-averse preferences. The following result further illustrates the connection by describing the parametric functional form of the ORA representation corresponding to EZKP utility.

Proposition 3 Suppose $h : [a, b] \rightarrow \mathbb{R}$ is differentiable,⁴⁵ concave, and $h' > 0$. Then for any measurable function $V : D \rightarrow [a, b]$ and $m \in \Delta(D)$,

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

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

The following corollary highlights the particular case of this equivalence that is used in our investment application in Section 4.⁴⁶

Corollary 2 For $\gamma \in \mathbb{R}$ and $\theta \in \Theta \subset \mathbb{R}_{++}$, define $\phi(x|\gamma, \theta)$ as in Equation (7), and suppose $\tau : \Theta \rightarrow \mathbb{R}$ satisfies $\inf_{\theta \in \Theta} \tau(\theta) = 0$. Then for any measurable function $V : D \rightarrow [a, b]$ and $m \in \Delta(D)$,

$$\sup_{\theta \in \Theta} \sup_{\gamma \in \mathbb{R}} \mathbb{E}_m[\phi(V|\gamma, \theta) - \tau(\theta)] = \sup_{\theta \in \Theta} \left\{ -\frac{1}{\theta} \log(\mathbb{E}_m[\exp(-\theta V)]) - \tau(\theta) \right\}.$$

⁴⁴Equivalently, suppose \succsim satisfies Axioms 2–7 in Appendix A.1.

⁴⁵Differentiability is only assumed for expositional simplicity. If h is not differentiable at a point γ , then $h'(\gamma)$ in Proposition 3 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.

⁴⁶This result can also be proved using the observations in Example 2.1 in [Ben-Tal and Teboulle \(2007\)](#).

B Proofs

B.1 Proofs of Proposition 1 and Theorem 1

Lemma 1 *The relation \succsim satisfies weak order, nontriviality, continuity, stationarity, and separability (Axiom 2–6) if and only if there exist continuous and nonconstant functions $u_1 : C \rightarrow \mathbb{R}$ and $u_2 : \Delta(D) \rightarrow \mathbb{R}$ and a scalar $\beta \in (0, 1)$ such that the following hold:*

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

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

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

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 6, stationarity (Axiom 5) implies that $(c, \delta_{(\hat{c}, \hat{m})}) \succsim (c, \delta_{(\hat{c}', \hat{m}')}})$ if and only if $(c', \delta_{(\hat{c}, \hat{m})}) \succsim (c', \delta_{(\hat{c}', \hat{m}')}})$. Therefore, the assumed axioms are sufficient to apply Theorem 3 of [Debreu \(1960\)](#) to obtain continuous functions $f : C \rightarrow \mathbb{R}$, $g : C \rightarrow \mathbb{R}$, and $h : \Delta(D) \rightarrow \mathbb{R}$ such that

$$(c, \delta_{(\hat{c}, \hat{m})}) \succsim (c', \delta_{(\hat{c}', \hat{m}')}}) \iff f(c) + g(\hat{c}) + h(\hat{m}) \geq f(c') + g(\hat{c}') + h(\hat{m}'). \quad (19)$$

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

$$\begin{aligned} (\hat{c}, \hat{m}) \succsim (\hat{c}', \hat{m}') &\iff (c, \delta_{(\hat{c}, \hat{m})}) \succsim (c, \delta_{(\hat{c}', \hat{m}')}}) \\ &\iff g(\hat{c}) + h(\hat{m}) \geq g(\hat{c}') + h(\hat{m}'), \end{aligned} \quad (20)$$

and hence g and h give an additive representation for \succsim . In particular, the combination of Equations (19) and (20) implies

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

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

$$\begin{aligned} g(c) &= \beta f(c) + \alpha_1, & \forall c \in C \\ h(\delta_{(\hat{c}, \hat{m})}) &= \beta [g(\hat{c}) + h(\hat{m})] + \alpha_2, & \forall (\hat{c}, \hat{m}) \in D \end{aligned} \quad (21)$$

Define $u_1 : C \rightarrow \mathbb{R}$ and $u_2 : \Delta(D) \rightarrow \mathbb{R}$ 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 (20) and (21).

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 $m \in \Delta(D)$ and, with slight abuse of notation, define sequences $\{d_n\}$ and $\{d'_n\}$ in D as follows:⁴⁷

$$d_n = (\underbrace{c^*, \dots, c^*}_n, m) \quad \text{and} \quad d'_n = (\underbrace{c_*, \dots, c_*}_n, m)$$

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, $V(d_{n_k}) \rightarrow V(d)$ and $V(d'_{n_k}) \rightarrow V(d')$, 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,

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

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

Lemma 2 *Suppose \succsim is represented by $V(c, m) = u_1(c) + u_2(m)$ where $u_1 : C \rightarrow \mathbb{R}$ and $u_2 : \Delta(D) \rightarrow \mathbb{R}$ are continuous,⁴⁸ and \succsim satisfies stationarity. Then, \succsim satisfies FOSD (Axiom 7) if and only if, for any $m, m' \in \Delta(D)$,*

$$m \circ V^{-1}([x, \infty)) \geq m' \circ V^{-1}([x, \infty)), \quad \forall x \in V(D) \implies u_2(m) \geq u_2(m'). \quad (22)$$

Proof: To see that FOSD implies Equation (22), consider any two measures $m, m' \in \Delta(D)$ such that

$$m \circ V^{-1}([x, \infty)) \geq m' \circ V^{-1}([x, \infty)), \quad \forall x \in V(D).$$

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

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

⁴⁷More precisely, $d_1 = (c^*, m)$, $d_2 = (c^*, \delta_{d_1})$, and so on.

⁴⁸Continuity is not necessary for this result; measurability of u_1 and u_2 are sufficient.

Since this condition holds for all $(\hat{c}, \hat{m}) \in D$, the FOSD axiom implies $(c, m) \succsim (c, m')$. Thus $u_2(m) \geq u_2(m')$. The argument that Equation (22) implies the FOSD axiom is similar. ■

Lemma 3 *Suppose \succsim is represented by $V(c, m) = u_1(c) + u_2(m)$ where $u_1 : C \rightarrow \mathbb{R}$ and $u_2 : \Delta(D) \rightarrow \mathbb{R}$ are nonconstant and continuous. Then, \succsim satisfies mixture aversion (Axiom 1) if and only if u_2 is convex.*

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

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

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

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

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

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 mixture aversion axiom. Therefore,

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

Now, take any $m, m' \in \Delta(D)$. Define a function $\psi : [0, 1] \rightarrow \mathbb{R}$ by $\psi(\alpha) = u_2(\alpha m + (1-\alpha)m')$. This function is continuous by the weak* continuity of u_2 , 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 $\alpha, \alpha' \in [\bar{\alpha}, \bar{\alpha}']$,

$$\frac{1}{2}\psi(\alpha) + \frac{1}{2}\psi(\alpha') \geq \psi(\frac{1}{2}\alpha + \frac{1}{2}\alpha').$$

It is a standard 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' were arbitrary, u_2 is convex. ■

Lemma 4 Suppose V , u_1 , u_2 , and β are as in Lemma 1, and suppose V and u_2 satisfy Equation (22). Then, there exists a function $W : \Delta(V(D)) \rightarrow \mathbb{R}$ such that $u_2(m) = \beta W(m \circ V^{-1})$ for all $m \in \Delta(D)$. Moreover,

1. $W(\delta_x) = x$ for all $x \in V(D)$.
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 $m \in \Delta(D)$ (Part 5 of Theorem 15.14 in Aliprantis and Border (2006)), that is,

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

Fix any $\mu \in \Delta(V(D))$. Let $W(\mu) = \frac{1}{\beta} u_2(m)$ for any $m \in \Delta(D)$ 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 $m, m' \in \Delta(D)$, then $u_2(m) = u_2(m')$ by Equation (22). Thus W is well defined, and by construction, $u_2(m) = \beta W(m \circ V^{-1})$ for all $m \in \Delta(D)$.

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

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

Proof of 2: To see that W is weak* continuous, take any sequence $\{\mu_n\}$ in $\Delta(V(D))$ that converges to some $\mu \in \Delta(V(D))$. It suffices to show that there exists a subsequence $\{\mu_{n_k}\}$ such that $W(\mu_{n_k}) \rightarrow W(\mu)$.⁴⁹ For each n , take any $m_n \in \Delta(D)$ such that $\mu_n = m_n \circ V^{-1}$. Since $\Delta(D)$ is compact and metrizable, there is a subsequence $\{m_{n_k}\}$ converging to some $m \in \Delta(D)$. 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,

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

Therefore, W is weak* continuous. To see that W is monotone with respect to FOSD, suppose $\mu, \eta \in \Delta(V(D))$ satisfy $\mu([x, b]) \geq \eta([x, b])$ for all $x \in [a, b] \equiv V(D)$. Take any $m, m' \in \Delta(D)$ such that $\mu = m \circ V^{-1}$ and $\eta = m' \circ V^{-1}$. Then, Equation (22) implies $u_2(m) \geq u_2(m')$, and hence $W(\mu) \geq W(\eta)$.

⁴⁹If 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.

Proof of 3: Suppose u_2 is convex. Fix any $\mu, \eta \in \Delta(V(D))$ and $\alpha \in (0, 1)$. Take any $m, m' \in \Delta(D)$ 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

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

establishing the convexity of W . ■

Proof of Proposition 1: The necessity of the axioms is straightforward. To establish sufficiency, suppose \succsim satisfies Axioms 2–7. By Lemmas 1, 2, and 4, there exists a continuous function $V : D \rightarrow \mathbb{R}$, a scalar $\beta \in (0, 1)$, a continuous and nonconstant function $u : C \rightarrow \mathbb{R}$, and a weak* continuous function $W : \Delta(V(D)) \rightarrow \mathbb{R}$ such that

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

Moreover, $W(\delta_x) = x$ for all $x \in V(D)$, and W is monotone with respect to FOSD. Finally, since V is continuous and D is compact and connected, $V(D) = [a, b]$ for some $a, b \in \mathbb{R}$. ■

Proof of Theorem 1: *Proof of 3 \Rightarrow 1:* The necessity of Axiom 1 is straightforward.

Proof of 1 \Rightarrow 2: By Lemmas 3 and 4, the certainty equivalent W is convex.

Proof of 2 \Rightarrow 3: Apply Part 1 of Corollary S.1 in the Supplementary Appendix to conclude there exists a collection Φ of continuous and nondecreasing functions $\phi : [a, b] \rightarrow \mathbb{R}$ such that

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

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

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

In addition,

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

for all $\bar{x} \in [a, b]$. ■

B.2 Proof of Proposition 2

Suppose \succsim has an Epstein-Zin representation (V, u, W, β) . Since \succsim also satisfies the expected-utility axioms when restricted to lotteries $m \in \Delta(D)$, there exists a continuous function $f : D \rightarrow \mathbb{R}$ such that $\mathbb{E}_m[f] \geq \mathbb{E}_{m'}[f]$ if and only if $(c, m) \succsim (c, m')$. (The particular c is irrelevant by the separability of the EZ representation.) Therefore, $W(m \circ V^{-1}) \geq W(m' \circ V^{-1})$ if and only if $\mathbb{E}_m[f] \geq \mathbb{E}_{m'}[f]$, and hence there exists a monotone transformation h such that $h(W(m \circ V^{-1})) = \mathbb{E}_m[f]$ for all $m \in \Delta(D)$. Continuity of h follows from continuity of f and W and connectedness of the domain. Since W is a certainty equivalent, $W(\delta_{(c,m)} \circ V^{-1}) = V(c, m)$ for any $(c, m) \in D$, and hence

$$h(V(c, m)) = h(W(\delta_{(c,m)} \circ V^{-1})) = \mathbb{E}_{\delta_{(c,m)}}[f] = f(c, m).$$

Thus, $f = h \circ V$, which implies $W(m \circ V^{-1}) = h^{-1}(\mathbb{E}_m[h(V)])$ for any $m \in \Delta(D)$.

To prove the second claim, note that $m \mapsto \mathbb{E}_m[h(V)]$ is a linear function of m . Therefore, $h^{-1}(\mathbb{E}_m[h(V)])$ is convex in m if and only if h^{-1} is convex, i.e., h is concave. Since Axiom 1 corresponds to the convexity of the certainty equivalent by Theorem 1, this completes the proof.

B.3 Proof of Proposition 3

The concavity of h implies that $h(x) - h(\gamma) \leq h'(\gamma)(x - \gamma)$ for any $\gamma, x \in [a, b]$. Rearranging terms yields

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

with equality if $\gamma = x$. For any $y \in h([a, b])$, letting $x = h^{-1}(y)$, this implies

$$\gamma + \frac{y - h(\gamma)}{h'(\gamma)} \leq h^{-1}(y),$$

with equality if $\gamma = h^{-1}(y)$. The results follows by taking $y = \mathbb{E}_m[h(V)]$.

B.4 Proof of Theorem 3

It is immediate that if $V_2 \geq \alpha V_1 + \lambda$, with equality for deterministic consumption streams, then \succsim_1 is more risk averse than \succsim_2 . For the other direction, first note that for any ORA representation (V, u, Φ, β) and any $\mathbf{c} = (c_0, c_1, c_2, \dots) \in C^{\mathbb{N}}$,

$$V(\mathbf{c}) = \sum_{t=0}^{\infty} \beta^t u(c_t).$$

The following lemma shows that the range of V on D is the same as its range when restricted to deterministic consumption streams in $C^{\mathbb{N}}$. As a consequence, it is possible to construct

a measurable mapping from any continuation value in $V(D)$ to a deterministic consumption stream that gives the same continuation value.

Lemma 5 *Fix any ORA representation (V, u, Φ, β) . There exist $c_*, c^* \in C$ such that, for any $(c, m) \in D$,*

$$\frac{1}{1-\beta}u(c_*) \leq V(c, m) \leq \frac{1}{1-\beta}u(c^*).$$

Therefore,

$$V(D) = V(C^{\mathbb{N}}) = \left[\frac{u(c_*)}{1-\beta}, \frac{u(c^*)}{1-\beta} \right]$$

and there exists a measurable function $g : V(D) \rightarrow C^{\mathbb{N}}$ such that $V(g(x)) = x$ for all $x \in V(D)$.

Proof: By the continuity of V and the compactness of D , there exist $(c_*, m_*), (c^*, m^*) \in D$ such that, for all $(c, m) \in D$,

$$V(c_*, m_*) \leq V(c, m) \leq V(c^*, m^*).$$

Since each $\phi \in \Phi$ is nondecreasing,

$$V(c^*, m^*) = u(c^*) + \beta \sup_{\phi \in \Phi} \mathbb{E}_{m^*}[\phi(V)] \leq u(c^*) + \beta \max_{(c, m) \in D} V(c, m) = u(c^*) + \beta V(c^*, m^*).$$

Rearranging terms gives

$$V(c^*, m^*) \leq \frac{u(c^*)}{1-\beta}.$$

A similar argument gives

$$\frac{u(c_*)}{1-\beta} \leq V(c_*, m_*),$$

which proves the first claim. Since C is connected and V is continuous, it follows that $V(D) = V(C^{\mathbb{N}}) = \left[\frac{u(c_*)}{1-\beta}, \frac{u(c^*)}{1-\beta} \right]$.

The existence of a function $g : V(D) \rightarrow C^{\mathbb{N}}$ such that $V(g(x)) = x$ for all $x \in V(D)$ follows immediately from the range condition above. That g can be chosen to be measurable is less trivial, but follows from Theorem 18.17 in [Aliprantis and Border \(2006\)](#). \blacksquare

Continuing the proof of Theorem 3, the definition of more risk averse applied to deterministic consumption streams implies that, for any $\mathbf{c}, \mathbf{c}' \in C^{\mathbb{N}}$,

$$V_1(\mathbf{c}') \geq V_1(\mathbf{c}) \implies V_2(\mathbf{c}') \geq V_2(\mathbf{c}). \quad (23)$$

Note that we have not established that the preferences over deterministic consumption streams is the same for these two individuals, since $V_1(\mathbf{c}') > V_1(\mathbf{c})$ does not necessarily imply $V_2(\mathbf{c}') > V_2(\mathbf{c})$. However, the following lemma shows that given the separable structure of the representations, this is in fact implied.

Lemma 6 Fix any two ORA representations $(V_1, u_1, \Phi_1, \beta_1)$ and $(V_2, u_2, \Phi_2, \beta_2)$. Suppose that, for any $\mathbf{c}, \mathbf{c}' \in C^{\mathbb{N}}$, $V_1(\mathbf{c}') \geq V_1(\mathbf{c})$ implies $V_2(\mathbf{c}') \geq V_2(\mathbf{c})$. Then $\beta_1 = \beta_2$ and there exist $\alpha > 0$ and $\lambda \in \mathbb{R}$ such that $u_2 = \alpha u_1 + \lambda(1 - \beta_1)$. Hence, for any $\mathbf{c} \in C^{\mathbb{N}}$, $V_2(\mathbf{c}) = \alpha V_1(\mathbf{c}) + \lambda$.

Proof: First, show that u_1 and u_2 are ordinally equivalent. Let $c_*, c^* \in C$ be such that $u_2(c^*) > u_2(c_*)$. Such consumption values exist by the nontriviality axiom. Now fix any $c, c' \in C$. If $u_1(c') \geq u_1(c)$ then $u_2(c') \geq u_2(c)$. This follows by applying Equation (23) to the consumption streams $\mathbf{c} = (c, c, c, \dots)$ and $\mathbf{c}' = (c', c', c', \dots)$. Also, if $u_1(c') > u_1(c)$ then $u_2(c') > u_2(c)$. To see that this must hold, choose $t \in \mathbb{N}$ sufficiently large that

$$u_1(c') + \beta_1^t u_1(c_*) > u_1(c) + \beta_1^t u_1(c^*).$$

By Equation (23) this implies

$$u_2(c') + \beta_2^t u_2(c_*) \geq u_2(c) + \beta_2^t u_2(c^*).$$

Since $u_2(c^*) > u_2(c_*)$ this implies $u_2(c') > u_2(c)$, as claimed. Thus we have shown that $u_1(c') \geq u_1(c) \iff u_2(c') \geq u_2(c)$.

Next, fix any consumption streams $\mathbf{c} = (c_0, c_1, c_2, \dots)$ and $\mathbf{c}' = (c'_0, c'_1, c'_2, \dots)$. We need to show that $V_1(\mathbf{c}') > V_1(\mathbf{c})$ implies $V_2(\mathbf{c}') > V_2(\mathbf{c})$. Suppose to the contrary that $V_2(\mathbf{c}') = V_2(\mathbf{c})$. By the continuity of u_1 and the connectedness of C , there exists a consumption stream $\mathbf{c}'' = (c''_0, c''_1, c''_2, \dots)$ such that $u_1(c''_t) < u_1(c'_t)$ for some $t \in \mathbb{N}$, $c''_{t'} = c'_{t'}$ for all $t' \neq t$, and $V_1(\mathbf{c}') > V_1(\mathbf{c}'') > V_1(\mathbf{c})$. Since u_1 and u_2 are ordinally equivalent, this requires that $u_2(c''_t) < u_2(c'_t)$ and hence $V_2(\mathbf{c}'') < V_2(\mathbf{c}') = V_2(\mathbf{c})$, contradicting Equation (23). Thus we have shown that $V_1(\mathbf{c}') \geq V_1(\mathbf{c}) \iff V_2(\mathbf{c}') \geq V_2(\mathbf{c})$.

Since these representations are ordinally equivalent for all deterministic consumption streams, it follows from the uniqueness of additively separable representations (see [Debreu \(1960\)](#) or Theorem 5.4 in [Fishburn \(1970\)](#)) that $\beta_1 = \beta_2$ and there exist $\alpha > 0$ and $\gamma \in \mathbb{R}$ such that $u_2 = \alpha u_1 + \gamma$. Let $\lambda = \gamma/(1 - \beta_1)$, and hence $u_2 = \alpha u_1 + \lambda(1 - \beta_1)$. ■

Fix any $(c, m) \in D$, and let $x = V_1(c, m)$. Define $g_1 : V_1(D) \rightarrow C^{\mathbb{N}}$ and $g_2 : V_2(D) \rightarrow C^{\mathbb{N}}$ as in Lemma 5. Note that $(c, m) \sim_1 g_1(x)$. Since \succsim_1 is more risk averse than \succsim_2 , this implies $(c, m) \succsim_2 g_1(x)$. Thus, by Lemma 6,

$$V_2(c, m) \geq V_2(g_1(x)) = \alpha V_1(g_1(x)) + \lambda = \alpha V_1(c, m) + \lambda.$$

More explicitly, letting $\beta \equiv \beta_1 = \beta_2$,

$$u_2(c) + \beta \sup_{\phi \in \Phi_2} \int_D \phi(V_2(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}) \geq \alpha u_1(c) + \alpha \beta \sup_{\phi \in \Phi_1} \int_D \phi(V_1(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}) + \lambda.$$

Since $u_2 = \alpha u_1 + \lambda(1 - \beta)$, this implies that for any $m \in \Delta(D)$,

$$\sup_{\phi \in \Phi_2} \int_D \phi(V_2(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}) \geq \alpha \sup_{\phi \in \Phi_1} \int_D \phi(V_1(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}) + \lambda. \quad (24)$$

Suppose $\phi_1 \in \Phi_1^*$, and define $\phi_2 : V_2(D) \rightarrow \mathbb{R}$ by

$$\phi_2(x) = \alpha \phi_1\left(\frac{x - \lambda}{\alpha}\right) + \lambda.$$

To establish condition 2, it must be shown that $\phi_2 \in \Phi_2^*$. By definition,

$$\int_D \phi_1(V_1(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}) \leq \sup_{\hat{\phi} \in \Phi_1} \int_D \hat{\phi}(V_1(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}), \quad \forall m \in \Delta(D). \quad (25)$$

Note also that for any $(\hat{c}, \hat{m}) \in D$,

$$V_2(\hat{c}, \hat{m}) = V_2(g_2(V_2(\hat{c}, \hat{m}))) = \alpha V_1(g_2(V_2(\hat{c}, \hat{m}))) + \lambda. \quad (26)$$

Fix any $m \in \Delta(D)$, and let $\tilde{m} = m \circ V_2^{-1} \circ g_2^{-1}$. Then

$$\begin{aligned} & \alpha \int_D \phi_1\left(\frac{V_2(\hat{c}, \hat{m}) - \lambda}{\alpha}\right) dm(\hat{c}, \hat{m}) + \lambda \\ &= \alpha \int_D \phi_1(V_1(g_2(V_2(\hat{c}, \hat{m})))) dm(\hat{c}, \hat{m}) + \lambda && \text{(by Equation (26))} \\ &= \alpha \int_D \phi_1(V_1(\hat{c}, \hat{m})) d\tilde{m}(\hat{c}, \hat{m}) + \lambda && \text{(change of variables)} \\ &\leq \alpha \sup_{\hat{\phi} \in \Phi_1} \int_D \hat{\phi}(V_1(\hat{c}, \hat{m})) d\tilde{m}(\hat{c}, \hat{m}) + \lambda && \text{(by Equation (25))} \\ &\leq \sup_{\hat{\phi} \in \Phi_2} \int_D \hat{\phi}(V_2(\hat{c}, \hat{m})) d\tilde{m}(\hat{c}, \hat{m}) && \text{(by Equation (24))} \\ &= \sup_{\hat{\phi} \in \Phi_2} \int_D \hat{\phi}(V_2(g_2(V_2(\hat{c}, \hat{m})))) dm(\hat{c}, \hat{m}) && \text{(change of variables)} \\ &= \sup_{\hat{\phi} \in \Phi_2} \int_D \hat{\phi}(V_2(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}). && \text{(by Equation (26))} \end{aligned}$$

Since this is true for any $m \in \Delta(D)$, we have shown that $\phi_2 \in \Phi_2^*$. Therefore, there is an injection $f : \Phi_1^* \rightarrow \Phi_2^*$ defined by

$$f(\phi_1)(x) = \alpha \phi_1\left(\frac{x - \lambda}{\alpha}\right) + \lambda$$

for $x \in V_2(D)$. This establishes condition 2 and completes the proof.

References

- Aliprantis, C., and K. Border (2006): *Infinite Dimensional Analysis*, 3rd edition. Berlin, Germany: Springer-Verlag. [44,47,15,16,18]
- Allais, M. (1953): “Le Comportement de l’Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l’Ecole Americaine,” *Econometrica*, 21, 503–546. [2]
- Ang, A., G. Bekaert, and J. Liu (2005): “Why Stocks May Disappoint,” *Journal of Financial Economics*, 76, 471–508. [6,22]
- Attanasio, O. P., J. Banks, and S. Tanner (2002): “Asset Holding and Consumption Volatility,” *Journal of Political Economy*, 110, 771–792, [21]
- Aumann, R. J. (1966): “Existence of Competitive Equilibria in Markets with a Continuum of Traders,” *Econometrica*, 34, 1–17. [27]
- Barberis, N., M. Huang, and T. Santos (2001): “Prospect Theory and Asset Prices,” *Quarterly Journal of Economics*, 116, 1–53. [14]
- Barberis, N., M. Huang, and R. H. Thaler (2006): “Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing,” *American Economic Review*, 96, 1069–1090. [22,36]
- 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. [6]
- Ben-Tal, A., and M. Teboulle (1986): “Expected Utility, Penalty Functions, and Duality in Stochastic Nonlinear Programming,” *Management Science*, 32, 1445–1466. [13]
- Ben-Tal, A., and M. Teboulle (2007): “An Old-New Concept of Convex Risk Measures: The Optimized Certainty Equivalent,” *Mathematical Finance*, 17, 449–476. [13,40]
- Bénabou, R., and J. Tirole (2011): “Identity, Morals and Taboos: Beliefs as Assets,” *Quarterly Journal of Economics*, 126, 805–855. [37]
- Blackwell, D. (1965): “Discounted Dynamic Programming,” *Annals of Mathematical Statistics*, 36, 226–235. [9,12,16]
- Boyle, P., L. Garlappi, R. Uppal, and T. Wang (2012): “Keynes Meets Markowitz: The Trade-Off Between Familiarity and Diversification,” *Management Science*, 58, 253–272. [23]
- Brandenburger, A., and E. Dekel (1993): “Hierarchies of Beliefs and Common Knowledge,” *Journal of Economic Theory*, 59, 189–198. [7]
- Brav, A., G. M. Constantinides, and C. C. Geczy (2002): “Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence,” *Journal of Political Economy*, 110, 793–824. [21,36]
- Briggs, J. S., D. Cesarini, E. Lingqvist, and R. Östling (2015): “Windfall Gains and Stock Market Participation,” NBER working paper. [21]
- Brunnermeier, M., and J. Parker (2005): “Optimal Expectations,” *American Economic Review*, 95, 1092–1118. [37]
- Campbell, J. Y. (2006): “Household Finance,” *Journal of Finance*, 61, 1553–1604. [20]

- Cao, H. H., T. Wang, and H. H. Zhang (2005): “Model Uncertainty, Limited Market Participation, and Asset Prices,” *Review of Financial Studies*, 18, 1219–1251. [23]
- Caplin, A., and J. Leahy (2001): “Psychological Expected Utility Theory and Anticipatory Feelings,” *Quarterly Journal of Economics*, 116, 55–79. [14]
- Cerreia-Vioglio, S. (2009): “Maxmin Expected Utility on a Subjective State Space: Convex Preferences under Risk,” working paper. [11,3]
- Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2015): “Cautious Expected Utility and the Certainty Effect,” *Econometrica*, 83, 693–728. [3,5]
- Cerreia-Vioglio, S., F. Maccheroni, and M. Marinacci (2015): “Stochastic Dominance Analysis without the Independence Axiom,” *Management Science*, forthcoming. [17,1,2,3,4]
- Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011): “Uncertainty Averse Preferences,” *Journal of Economic Theory*, 146, 1275–1330. [3]
- Chapman, D. A., and V. Polkovnichenko (2009): “First-Order Risk Aversion, Heterogeneity, and Asset Market Outcomes,” *Journal of Finance*, 64, 1863–1887. [23]
- Chatterjee, K., and R. V. Krishna (2011): “A Nonsmooth Approach to Nonexpected Utility Theory under Risk,” *Mathematical Social Sciences*, 62, 166–175. [2,4]
- 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. [6,5]
- Chew, S. H., and L. G. Epstein (1991): “Recursive Utility Under Uncertainty,” in A. Khan and N. Yannelis (Eds.), *Equilibrium Theory in Infinity Dimensional Spaces*, Springer Verlag. [3,7,8,18,19,38,39,40]
- Chew, S. H., and L. G. Epstein, and U. Segal (1991): “Mixture Symmetry and Quadratic Utility,” *Econometrica*, 59, 139–163. [7]
- Chew, S. H., E. Karni, and Z. Safra (1987): “Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities,” *Journal of Economic Theory*, 42, 370–381. [4,7]
- Chetty, R., and A. Szeidl (2007): “Consumption Commitments and Risk Preferences,” *Quarterly Journal of Economics*, 122, 831–877. [10]
- Chetty, R., and A. Szeidl (2016): “Consumption Commitments and Habit Formation,” *Econometrica*, 84, 855–890. [10]
- Constantinides, G. M., and D. Duffie (1996): “Asset Pricing with Heterogeneous Consumers,” *Journal of Political Economy*, 104, 219–240. [36]
- 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. [8,38,41,48]
- Dekel, E. (1986): “An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom,” *Journal of Economic Theory*, 40, 304–318. [6,5]
- Dekel, E. (1989): “Asset Demands without the Independence Axiom,” *Econometrica*, 57, 163–169. [7]

- Diecidue, E., U. Schmidt, and P. P. Wakker (2004): “The Utility of Gambling Reconsidered,” *Journal of Risk and Uncertainty*, 29, 241–259. [17]
- Dow, J., and S. R. Werlang (1992): “Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio,” *Econometrica*, 60, 197–204. [22]
- Ekeland, I., and T. Turnbull (1983): *Infinite-Dimensional Optimization and Convexity*. Chicago: The University of Chicago Press. [14]
- Epstein, L. G. (2008): “Living with Risk,” *Review of Economic Studies*, 75, 1121–1141. [14]
- Epstein, L. G., and J. Miao (2003): “A Two-Person Dynamic Equilibrium Under Ambiguity,” *Journal of Economic Dynamics and Control*, 27, 1253–1288. [23]
- Epstein, L. G., and M. Schneider (2007): “Learning Under Ambiguity,” *Review of Economic Studies*, 74, 1275–1303. [22]
- Epstein, L. G., and T. Wang (1994): “Intertemporal Asset Pricing Under Knightian Uncertainty,” *Econometrica*, 62, 283–322. [22]
- 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. [2,3,5,6,7,10,18,30,12,13]
- Epstein, L. G., and S. Zin (1990): “‘First-Order’ Risk Aversion and the Equity Premium Puzzle,” *Journal of Monetary Economics*, 26, 387–407. [2,6,30]
- Epstein, L. G., and S. Zin (2001): “The Independence Axiom and Asset Returns,” *Journal of Empirical Finance*, 8, 537–572. [6]
- Ergin, H., and T. Sarver (2015): “Hidden Actions and Preferences for Timing of Resolution of Uncertainty,” *Theoretical Economics*, 10, 489–541. [10]
- Fishburn, P. C. (1970): *Utility Theory for Decision Making*. New York: John Wiley and Sons. [41,48]
- Gabaix, X., and D. Laibson (2001): “The 6D Bias and the Equity-Premium Puzzle,” *NBER Macroeconomics Annual*, 16, 257–312. [10]
- Ghirardato, P., F. Maccheroni, and M. Marinacci (2004): “Differentiating Ambiguity and Ambiguity Attitude,” *Journal of Economic Theory*, 118, 133–173. [3]
- Ghirardato, P., and M. Siniscalchi (2012): “Ambiguity in the Small and in the Large,” *Econometrica*, 80, 2827–2847. [3]
- Gollier, C., and A. Muermann (2010): “Optimal Choice and Beliefs with Ex Ante Savoring and Ex Post Disappointment,” *Management Science*, 56, 1272–1284. [13,37]
- Gollier, C., and J. W. Pratt (1996): “Risk Vulnerability and the Tempering Effect of Background Risk,” *Econometrica*, 64, 1109–1123. [36,37]
- Gomes, F., and A. Michaelides (2005): “Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence,” *Journal of Finance*, 60, 869–904. [21]
- Grant, S., A. Kajii, and B. Polak (2000): “Temporal Resolution of Uncertainty and Recursive Non-Expected Utility Models,” *Econometrica*, 68, 425–434. [5]

- Grossman, S., and G. Laroque (1990): “Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods,” *Econometrica*, 58, 25–51. [10]
- Guiso, L., and P. Sodini (2013): “Household Finance: An Emerging Field,” in G. M. Constantinides, M. Harris and R. M. Stulz (Eds.), *Handbook of the Economics of Finance*, Volume 2, Amsterdam, The Netherlands: Elsevier. [20,22]
- Gul, F. (1991): “A Theory of Disappointment Aversion,” *Econometrica*, 59, 667–686. [6,14,22,5]
- Gul, F., and W. Pesendorfer (2004): “Self-Control and the Theory of Consumption,” *Econometrica*, 72, 119–158. [7,42]
- Haliassos, M., and C. C. Bertaut (1995): “Why Do So Few Hold Stocks?,” *The Economic Journal*, 105, 1110–1129. [20]
- Harless, D., and C. Camerer (1994): “The Predictive Utility of Generalized Expected Utility Theories,” *Econometrica*, 62, 1251–1289. [17]
- Heaton, J., and D. Lucas (2000): “Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk,” *Journal of Finance*, 55, 1163–1198. [20]
- Kandel, S., and R. F. Stambaugh (1991): “Asset Returns and Intertemporal Preferences,” *Journal of Monetary Economics*, 27, 39–71. [30]
- Kahneman, D., and A. Tversky (1979): “Prospect Theory: An Analysis of Decision Under Risk,” *Econometrica*, 47, 263–292. [2,3,14,4]
- Kocherlakota, N. (1990): “Disentangling the Coefficient of Relative Risk Aversion from the Elasticity of Intertemporal Substitution: An Irrelevance Result,” *Journal of Finance*, 45, 175–190. [10]
- Kocherlakota, N. (1996): “The Equity Premium: It’s Still a Puzzle,” *Journal of Economic Literature*, 34, 42–71. [31]
- Kőszegi, B. (2010): “Utility from Anticipation and Personal Equilibrium,” *Economic Theory*, 44, 415–444. [14]
- Kőszegi, B., and M. Rabin (2006): “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics*, 121, 1133–1165. [14,5]
- Kőszegi, B., and M. Rabin (2007): “Reference-Dependent Risk Attitudes,” *American Economic Review*, 97, 1047–1073. [14,5]
- Kreps, D., and E. Porteus (1978): “Temporal Resolution of Uncertainty and Dynamic Choice Theory,” *Econometrica*, 46, 185–200. [2,5,39,40]
- Kreps, D., and E. Porteus (1979): “Temporal von Neumann-Morgenstern and Induced Preferences,” *Journal of Economic Theory*, 20, 81–109. [10]
- Lucas, R. E. (2003): “Macroeconomic Priorities,” *American Economic Review*, 93, 1–14. [31]
- Loewenstein, G. (1987): “Anticipation and the Valuation of Delayed Consumption,” *The Economics Journal*, 97, 666–684. [14]
- Maccheroni, F. (2002): “Maxmin Under Risk,” *Economic Theory*, 19, 823–831. [10]
- Macera, R. (2014): “Dynamic Beliefs,” working paper. [37]

- Machina, M. (1982): “‘Expected Utility’ Analysis without the Independence Axiom,” *Econometrica*, 50, 277–324. [1,2]
- Machina, M. (1984): “Temporal Risk and the Nature of Induced Preferences,” *Journal of Economic Theory*, 33, 199–231. [10,18,2]
- Mankiw, N. G., and S. P. Zeldes (1991): “The Consumption of Stockholders and Nonstockholders,” *Journal of Financial Economics*, 29, 97–112. [20,21]
- Marinacci, M., and L. Montrucchio (2010): “Unique Solutions for Stochastic Recursive Utilities,” *Journal of Economic Theory*, 145, 1776–1804. [9,13]
- Markowitz, H. (1952): “The Utility of Wealth,” *Journal of Political Economy*, 60, 151–158. [14]
- Masatlioglu, Y., and C. Raymond (2016): “A Behavioral Analysis of Stochastic Reference Dependence,” *American Economic Review*, forthcoming. [14,5]
- Mehra, R., and E. Prescott (1985): “The Equity Premium: A Puzzle,” *Journal of Monetary Economics*, 15, 145–161. [21,29,31]
- Mertens, J.-F., and S. Zamir (1985): “Formulation of Bayesian Analysis for Games with Incomplete Information,” *International Journal of Game Theory*, 14, 1–29. [7]
- Milgrom, P., and C. Shannon (1994): “Monotone Comparative Statics,” *Econometrica*, 62, 157–180. [9]
- Neilson, W. S. (1992): “Some Mixed Results on Boundary Effects,” *Economic Letters*, 39, 275–278. [17]
- Phelps, R. R. (1993): *Convex Functions, Monotone Operators, and Differentiability*. Berlin, Germany: Springer-Verlag. [14]
- Quah, J. (2007): “The Comparative Statics of Constrained Optimization Problems,” *Econometrica*, 75, 401–431. [9]
- Quiggin, J. (1982): “A Theory of Anticipated Utility,” *Journal of Economic Behavior and Organization*, 3, 323–343. [6,4]
- Rabin, M. (2000): “Risk Aversion and Expected-Utility Theory: A Calibration Theorem,” *Econometrica*, 68, 1281–1292. [31]
- Routledge, B., and S. Zin (2010): “Generalized Disappointment Aversion and Asset Prices,” *Journal of Finance*, 65, 1303–1332. [6]
- Safra, Z., and U. Segal (2008): “Calibration Results for Non-Expected Utility Theories,” *Econometrica*, 76, 1143–1166. [22]
- Sarver, T. (2012): “Optimal Reference Points and Anticipation,” CMS-EMS Discussion Paper 1566, Northwestern University. [9]
- Schmidt, U. (1998): “A Measurement of the Certainty Effect,” *Journal of Mathematical Psychology*, 42, 32–47. [17]
- Segal, U. (1988): “Probabilistic Insurance and Anticipated Utility,” *Journal of Risk and Insurance*, 55, 287–297. [3]
- Segal, U. (1989): “Anticipated Utility: A Measure Representation Approach,” *Annals of Operations Research*, 19, 359–373. [6,4]

- Segal, U., and A. Spivak (1990): “First Order versus Second Order Risk Aversion,” *Journal of Economic Theory*, 51, 111–125. [15]
- Starr, R. M. (1969): “Quasi-Equilibria in Markets with Non-Convex Preferences,” *Econometrica*, 37, 25–38. [27]
- Sydnor, J. (2010): “(Over)insuring Modest Risks,” *American Economic Journal: Applied Economics*, 2, 177–199. [15,6]
- Tversky, A., and D. Kahneman (1992): “Advances in Prospect Theory: Cumulative Representation of Uncertainty,” *Journal of Risk and Uncertainty*, 5, 297–323. [4]
- Vissing-Jørgensen, A. (2002): “Limited Asset Market Participation and the Elasticity of Intertemporal Substitution,” *Journal of Political Economy*, 110, 825–853. [21]
- Vissing-Jørgensen, A. (2003): “Perspectives on Behavioral Finance: Does ‘Irrationality’ Disappear with Wealth? Evidence from Expectations and Actions,” *NBER Macroeconomics Annual*, 139–208. [21]
- Vissing-Jørgensen, A., and O. P. Attanasio (2003): “Stock-Market Participation, Intertemporal Substitution, and Risk-Aversion,” *American Economic Review*, 93, 383–391. [21]
- Wakker, P. (1994): “Separating Marginal Utility and Probabilistic Risk Aversion,” *Theory and Decision*, 36, 1–44. [4]
- Wakker, P., R. Thaler, and A. Tversky (1997): “Probabilistic Insurance,” *Journal of Risk and Uncertainty*, 15, 7–28. [2,3]
- Weil, P. (1989): “The Equity Premium Puzzle and the Risk-Free Rate Puzzle,” *Journal of Monetary Economics*, 24, 401–421. [5]
- Weil, P. (1990): “Nonexpected Utility in Macroeconomics,” *Quarterly Journal of Economics*, 105, 29–42. [5]
- Yaari, M. E. (1987): “The Dual Theory of Choice under Risk,” *Econometrica*, 55, 95–115. [6,14,4]

Supplementary Appendix for MIXTURE-AVERSE PREFERENCES AND HETEROGENEOUS STOCK MARKET PARTICIPATION

S.1 Local Expected-Utility Analysis

When applying optimal risk attitude model, one important consideration is how properties of the set of transformations Φ in the representation relate to properties of the corresponding risk preference. In this section, we show that the certainty equivalent for an ORA representation respects a stochastic order if and only if each of the transformations $\phi \in \Phi$ also respects this order. This result is similar in spirit to the local expected-utility analysis introduced in the influential paper by [Machina \(1982\)](#). After presenting the main theorem of this section, we will make precise connections to Machina's results and the many generalizations and extensions that appeared in the literature that followed. We will also describe how the expected-utility core recently developed by [Cerreia-Vioglio, Maccheroni, and Marinacci \(2015\)](#) can be related to the ORA representation.

The main result of this section applies to any convex utility representation on a set of lotteries $\Delta(X)$, where X is any compact metric space. Of particular interest is the special case where X is an interval, e.g., a set of monetary outcomes or the set of continuation values for an ORA representation. We first state a general definition of stochastic orders generated by sets of functions.

Definition S.1 Let \mathcal{C} be a set in the space of real-valued continuous functions $C(X)$ for some compact metric space X . The order $\geq_{\mathcal{C}}$ on $\Delta(X)$ generated by \mathcal{C} is defined by

$$\mu \geq_{\mathcal{C}} \eta \iff \int \phi(x) d\mu(x) \geq \int \phi(x) d\eta(x), \forall \phi \in \mathcal{C}.$$

A function $W : \Delta(X) \rightarrow \mathbb{R}$ is *monotone with respect to the order $\geq_{\mathcal{C}}$* if $\mu \geq_{\mathcal{C}} \eta$ implies $W(\mu) \geq W(\eta)$.

This definition includes as special cases all of the stochastic orders typically used in economics. For example, if X is a subset of the real numbers, the first-order stochastic dominance order is generated by taking \mathcal{C} to be the set of all nondecreasing continuous functions; the second-order stochastic dominance order is generated by taking \mathcal{C} to be the set of all nondecreasing and concave continuous functions; and so on.

For any set \mathcal{C} of continuous functions in $C(X)$, let $\langle \mathcal{C} \rangle$ denote smallest closed convex cone containing \mathcal{C} and all the constant functions, that is, $\langle \mathcal{C} \rangle$ is the closed convex hull of the set of all affine transformations of functions in \mathcal{C} . It is easy to see that the stochastic

order generated by $\langle \mathcal{C} \rangle$ is the same as the stochastic order generated by \mathcal{C} . The following result shows that a convex function respects the order generated by a set \mathcal{C} if and only if it can be expressed as the supremum of some subset of $\langle \mathcal{C} \rangle$.

Theorem S.1 *Suppose $W : \Delta(X) \rightarrow \mathbb{R}$ for some compact metric space X , and suppose $\mathcal{C} \subset C(X)$. The following are equivalent:*

1. *W is lower semicontinuous in the topology of weak convergence, convex, and monotone with respect to the order $\geq_{\mathcal{C}}$.*
2. *There exists a set of functions $\Phi \subset \langle \mathcal{C} \rangle$ such that*

$$W(\mu) = \sup_{\phi \in \Phi} \int \phi(x) d\mu(x). \quad (\text{S.1})$$

The following corollary of Theorem S.1 takes $X = [a, b]$ and relates monotonicity with respect to first-order stochastic dominance (FOSD) and second-order stochastic dominance (SOSD) to properties of the local utility functions. Part 1 of this corollary is applied to the set of continuation values $[a, b] = V(D)$ in the proof of Theorem 1 to show that a recursive representation with a convex and FOSD-monotone certainty equivalent can be given an optimal risk attitude representation.

Corollary S.1 *Suppose $W : \Delta([a, b]) \rightarrow \mathbb{R}$ is lower semicontinuous in the topology of weak convergence and convex. Then:*

1. *W is monotone with respect to FOSD if and only if it satisfies Equation (S.1) for some collection Φ of nondecreasing continuous functions $\phi : [a, b] \rightarrow \mathbb{R}$.*
2. *W is monotone with respect to SOSD if and only if it satisfies Equation (S.1) for some collection Φ of nondecreasing and concave continuous functions $\phi : [a, b] \rightarrow \mathbb{R}$.*

Corollary S.1 is a variation of the main local expected-utility results from 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. A number of papers have since explored relaxations of this differentiability assumption. Most recently, Cerreia-Vioglio, Maccheroni, and Marinacci (2015) showed that Machina's results can be extended to any Gateaux differentiable utility function and any integral stochastic order (as in Definition S.1). Since the ORA representation is in general not differentiable, these results will not suffice for the analysis in this paper. Theorem S.1 and Corollary S.1 complement the existing literature by showing that one can relax differentiability to the much weaker requirement of lower semicontinuity when dealing with convex functions.⁵⁰

⁵⁰Local 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 (2011) relaxed the assumption of differentiability and obtained several local expected-utility results for concave and Lipschitz continuous functions.

Theorem S.1 also provides a way of relating the ORA representation to the expected-utility core analyzed by Cerreia-Vioglio (2009), Cerreia-Vioglio, Dillenberger, and Ortolova (2015), and Cerreia-Vioglio, Maccheroni, and Marinacci (2015).⁵¹

Definition S.2 The *expected-utility core* of a function $W : \Delta([a, b]) \rightarrow \mathbb{R}$ is the binary relation \succeq on $\Delta([a, b])$ defined by

$$\mu \succeq \eta \iff W(\alpha\mu + (1 - \alpha)\nu) \geq W(\alpha\eta + (1 - \alpha)\nu), \quad \forall \alpha \in (0, 1], \forall \nu \in \Delta([a, b]).$$

It follows immediately from this definition that \succeq is consistent with W in the sense that

$$\mu \succeq \eta \implies W(\mu) \geq W(\eta).$$

If W satisfies independence,⁵² then the converse is also true and hence \succeq is a complete and transitive binary relation on $\Delta([a, b])$ that is represented by W . However, if W does not satisfy independence, then \succeq is necessarily incomplete. As discussed in Cerreia-Vioglio, Maccheroni, and Marinacci (2015), the expected-utility core \succeq is the largest relation that is consistent with W and satisfies independence.

Applying the results of Cerreia-Vioglio, Maccheroni, and Marinacci (2015) together with Theorem S.1 gives the following corollary.⁵³

Corollary S.2 Suppose $W : \Delta([a, b]) \rightarrow \mathbb{R}$ is continuous in the topology of weak convergence and convex. Let \succeq denote the expected-utility core of W . Then:

1. There exist a collection of continuous functions \mathcal{C} such that $\succeq = \succeq_{\mathcal{C}}$, i.e.,

$$\mu \succeq \eta \iff \int_a^b \phi(x) d\mu(x) \geq \int_a^b \phi(x) d\eta(x), \quad \forall \phi \in \mathcal{C}.$$

2. If a collection of continuous functions Φ satisfies Equation (S.1) for W , then $\mathcal{C} \subset \langle \Phi \rangle$.
3. There exists a collection of continuous functions Φ^m that satisfies Equation (S.1) for W and satisfies $\langle \mathcal{C} \rangle = \langle \Phi^m \rangle$.

Parts 1 and 2 of Corollary S.2 come from Cerreia-Vioglio, Maccheroni, and Marinacci (2015, Lemma 1). For part 3, note that $\mu \succeq_{\mathcal{C}} \eta$ implies $W(\mu) \geq W(\eta)$. Therefore,

⁵¹The expected-utility core is the risk counterpart of the revealed unambiguous preference relation studied by Ghirardato, Maccheroni, and Marinacci (2004), Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2011), and Ghirardato and Siniscalchi (2012).

⁵²Say that W satisfies independence if $W(\mu) \geq W(\eta)$ implies $W(\alpha\mu + (1 - \alpha)\nu) \geq W(\alpha\eta + (1 - \alpha)\nu)$.

⁵³I thank David Dillenberger for suggesting the connection to the expected-utility core and Simone Cerreia-Vioglio for detailed comments about how to formalize this result.

Theorem S.1 implies there exists a set $\Phi^m \subset \langle \mathcal{C} \rangle$ that satisfies Equation (S.1).⁵⁴ Note that parts 2 and 3 of this result together imply that $\langle \Phi^m \rangle \subset \langle \Phi \rangle$ for any set of continuous functions Φ that satisfies Equation (S.1). Thus the expected-utility core provides a way of identifying a set of transformations that is minimal in terms of the associated set of expected-utility preferences.

S.2 Related Non-Expected-Utility Preferences

In this section, we discuss the relationship between the optimal risk attitude representation and other non-expected-utility theories. The conclusions of this discussion are summarized in Figure 1 of the paper.

S.2.1 Probability Weighting and Rank-Dependent Utility

An important alternative to expected utility is the *rank-dependent utility* model (see, e.g., Quiggin (1982), Yaari (1987), Segal (1989)). For a probability measure $\mu \in \Delta([a, b])$, the certainty equivalent for rank-dependent utility takes the form

$$W(\mu) = h^{-1} \left(\int h(x) d(g \circ F_\mu)(x) \right),$$

where $h : [a, b] \rightarrow \mathbb{R}$ is continuous and strictly increasing, F_μ is the cumulative distribution function for the measure μ , and $g : [0, 1] \rightarrow [0, 1]$ is continuous, strictly increasing, and onto. The function g in the representation permits distortions of the cumulative probabilities. If $g(p) = p$ for all $p \in [0, 1]$, then the expression above reduces to the certainty equivalent for expected utility. However, when $g(F_\mu(x)) > F_\mu(x)$ for some x , the probability of obtaining an outcome below x is distorted upward, capturing the intuition that low probability bad events may be overweighted. Reweighting of probabilities also played an important role in the prospect theory of Kahneman and Tversky (1979) and Tversky and Kahneman (1992).

Chew, Karni, and Safra (1987) showed that a rank-dependent utility preference is risk averse (dislikes mean-preserving spreads) if and only if both h and g are concave. In this case, the certainty equivalent W is a convex function (see Wakker (1994, Observation 2) or Chatterjee and Krishna (2011, Proposition 4.6)). The argument is as follows: For concave g ,

$$g(F_{\alpha\mu+(1-\alpha)\eta}(x)) = g(\alpha F_\mu(x) + (1-\alpha)F_\eta(x)) \geq \alpha g(F_\mu(x)) + (1-\alpha)g(F_\eta(x))$$

⁵⁴Theorem 1 in Cerreia-Vioglio, Maccheroni, and Marinacci (2015) provides a similar result to part 3 of this corollary for Gateaux differentiable functions: If W is Gateaux differentiable, then $\langle \nabla W \rangle = \langle \mathcal{C} \rangle$. The ORA representation is in general not differentiable, and hence the connection between the dual representation of W and the set \mathcal{C} representing the expected-utility core instead relies on Theorem S.1.

for all $x \in [a, b]$. Since h is increasing, this FOSD relationship between the transformed distributions implies

$$\int h(x) d(g \circ F_{\alpha\mu+(1-\alpha)\eta})(x) \leq \alpha \int h(x) d(g \circ F_{\mu})(x) + (1 - \alpha) \int h(x) d(g \circ F_{\eta})(x).$$

Finally, concavity of h implies h^{-1} is convex, which yields the convexity of W . Since any recursive model with a convex certainty equivalent can be expressed as an ORA representation by Theorem 1, this shows that recursive risk-averse rank-dependent utility preferences are a special case of mixture-averse preferences.

In a recent paper, Masatlioglu and Raymond (2016) found a somewhat surprising relationship between rank-dependent utility and a model of endogenous reference points developed by Kőszegi and Rabin (2006, 2007). They showed that whenever the choice-acclimating personal equilibrium (CPE) concept for reference point formation from Kőszegi and Rabin (2006, 2007) leads to risk preference that respects first-order stochastic dominance, that preference conforms to rank-dependent utility. The implication for our model is that any CPE representation that respects both FOSD and SOSD also satisfies mixture aversion.

S.2.2 Disappointment Aversion, Betweenness, and Quasiconcave Risk Preferences

Another important class of non-expected-utility preferences are the *betweenness* preferences developed by Chew (1983) and Dekel (1986). One of the more widely-used special cases of betweenness preferences is the *disappointment aversion* model of Gul (1991). Grant, Kajii, and Polak (2000, Lemma 2) showed that any betweenness preference that has a convex representation must be an expected-utility preference. Thus, in a dynamic setting, the only overlap of recursive betweenness preferences and mixture-averse preferences is EZKP expected utility.

Another intriguing related model is the *cautious expected utility* representation recently proposed by Cerreia-Vioglio, Dillenberger, and Ortoleva (2015). The certainty equivalent for this model is the minimum of a set of expected-utility certainty equivalents. This representation has a nontrivial intersection with betweenness preferences that includes risk-averse disappointment aversion preferences. However, since cautious expected utility preferences are quasiconcave with respect to lotteries, they only overlap with mixture-averse preferences in the case of linear indifference curves, i.e., betweenness preferences. Therefore, by the previous observations, the intersection of recursive cautious expected utility and mixture-averse preferences is again EZKP utility.

S.3 Preference for Diversification

In this section, we consider random variables defined on some fixed probability space.⁵⁵ We will use \tilde{x} to denote a random variable and $\mu_{\tilde{x}}$ to denote the distribution of that random variable. As stated in the paper, a risk preference with a certainty equivalent W satisfies preference for diversification if it is quasiconcave with respect to random variables.

Definition S.3 A certainty equivalent $W : \Delta([a, b]) \rightarrow [a, b]$ exhibits *preference for diversification* (PD) if, for any random variables \tilde{x} and \tilde{y} and any $\alpha \in [0, 1]$,

$$W(\mu_{\tilde{x}}) \geq W(\mu_{\tilde{y}}) \implies W(\mu_{\alpha\tilde{x}+(1-\alpha)\tilde{y}}) \geq W(\mu_{\tilde{y}}).$$

Preference for diversification is a useful property of a model of risk preferences for several reasons: Together with homotheticity of preferences, PD permits representative agent analysis; it also implies the sufficiency of first-order conditions in maximization problems (e.g., portfolio choice). However, while each of these arguments supports PD as enhancing the analytic tractability of economic models, neither speaks to its descriptive realism, and in the paper we observed several important reasons for relaxing this condition. In Section 4, we showed that relaxing PD permits heterogeneity in stock market participation even when agents have identical preferences. In this section, we show that obtaining the properties of demand for insurance discussed in Section 3.2 also requires violating PD. We then discuss the connection between PD and risk aversion.

S.3.1 Preference for Diversification and Insurance Demand

Suppose, as in Example 3 from Section 3.2, that an individual has wealth w and faces a loss of amount L with probability π . Let $P(y)$ denote this individual's maximum willingness to pay (reservation price) for $y \in [0, L]$ dollars of insurance coverage paid in the event of a loss. The following property relates the willingness to pay for additional insurance to the existing level of coverage.

Definition S.4 An individual has a *nonincreasing marginal willingness to pay for insurance coverage* if $P(y + \varepsilon) - P(y)$ is nonincreasing in y for every $y, \varepsilon \geq 0$ such that $y + \varepsilon \leq L$.

In Section 3.2, we argued that the above definition may be overly restrictive and that violations of this property may better match observed insurance choices—people often accept large increases in premiums in order to reduce their insurance deductibles (Sydnor (2010)). We also showed that the ORA representation can permit a marginal willingness

⁵⁵Note that we are presuming that the probability measure is known, and therefore we are still working within a framework of objective risk, as opposed to subjective beliefs.

to pay for additional insurance coverage that increases at some levels of coverage (while still maintaining risk aversion). In contrast, the following result shows that any risk preference that satisfies preference for diversification must exhibit nonincreasing marginal willingness to pay for insurance.

Proposition S.1 *If the certainty equivalent W for an individual's risk preferences exhibits preference for diversification, then this individual has a nonincreasing marginal willingness to pay for insurance coverage.*

The implications of Proposition S.1 are immediate in the case of expected utility. If an individual does not have a nonincreasing marginal willingness to pay for insurance coverage, then she must violate preference for diversification. For an expected-utility maximizer, preference for diversification is satisfied if and only if her Bernoulli utility function is concave. Thus in order to have a marginal willingness to pay for insurance coverage that increases at some levels of coverage (e.g., near full coverage), the individual must violate risk aversion. The next section shows that this conclusion also extends to many non-expected-utility models.

S.3.2 Preference for Diversification and SOSD

The connections between preference for diversification and risk aversion (monotonicity with respect to SOSD) has been well documented for a number of models. Dekel (1989) showed that preference for diversification implies risk aversion for any risk preference. He also showed that the converse is true for preferences that are quasiconcave in probabilities. The following proposition summarizes his result as well as related observations for rank-dependent utility from Chew, Karni, and Safra (1987).

Proposition S.2 1. (Dekel (1989)) *If W is quasiconcave in probabilities and respects SOSD, then it satisfies preference for diversification.*

2. (Chew, Karni, and Safra (1987)) *If W is a rank-dependent utility certainty equivalent and it respects SOSD, then it satisfies preference for diversification.*

Note that Proposition S.2 applies to all of the risk preferences discussed in Section S.2. Thus, other than our model of mixture-averse preferences, most of the non-expected-utility preferences considered in the literature are encompassed by this result. The only prominent theory that we are aware of that is not covered by this result is the *quadratic utility* model of Chew, Epstein, and Segal (1991). To our knowledge, it remains an open question whether quadratic utility can violate PD while still respecting SOSD.

S.4 Derivations Used in the Investment Application

In this section, we provide some supporting results for the investment application in Section 4. First, we establish the existence and functional form of the value function for consumers' preferences. Second, we describe the precise equilibrium conditions for the economy for each of the three cases, and we provide the formulas for asset returns. These conditions will also serve as the basis of the numerical analysis in that section.

S.4.1 Properties of the Value Function

For ease of reference, we repeat Equations (13), (14), and (15) from the paper:

$$c_t + \mathbb{E}[Mw_{t+1}] = w_t, \quad (\text{S.2})$$

$$\mathcal{V}(w_t; M) = \max_{\substack{c_t \in \mathbb{R}_+ \\ w_{t+1} \in \mathbb{R}_+^Z}} \left\{ (1 - \beta) \log(c_t) + \beta \mathcal{R}(\mathcal{V}(w_{t+1}; M)) \right\}, \quad (\text{S.3})$$

$$\mathcal{R}(\mathcal{V}(w_{t+1}; M)) = \max_{\theta \in \Theta} \left\{ -\frac{1}{\theta} \log(\mathbb{E}[\exp(-\theta \mathcal{V}(w_{t+1}; M))]) - \tau(\theta) \right\}. \quad (\text{S.4})$$

Since equilibrium may involve heterogeneity in the choice of θ , it is useful to define the value of the optimization problem conditional on choosing risk attitude $\theta \in \Theta$ in a given period:

$$\mathcal{V}_\theta(w_t; M) = \max_{\substack{c_t \in \mathbb{R}_+ \\ w_{t+1} \in \mathbb{R}_+^Z}} \left\{ (1 - \beta) \log(c_t) - \beta \frac{1}{\theta} \log(\mathbb{E}[\exp(-\theta \mathcal{V}(w_{t+1}; M))]) - \tau(\theta) \right\}, \quad (\text{S.5})$$

subject to the budget constraint in Equation (S.2). Thus

$$\mathcal{V}(w_t; M) = \max_{\theta \in \Theta} \mathcal{V}_\theta(w_t; M).$$

The following proposition summarizes the relevant properties of \mathcal{V} and \mathcal{V}_θ that will be used in the equilibrium analysis in Section S.4.2.

Proposition S.3 *For any pricing kernel $M : Z \rightarrow \mathbb{R}_{++}$, there exists a unique value function \mathcal{V} for the problem described in Equations (S.2), (S.3), and (S.4). This value function takes the form*

$$\mathcal{V}(w_t; M) = \Lambda(M) + \log(w_t),$$

where

$$\Lambda(M) = \log(1 - \beta) + \frac{\beta}{1 - \beta} \log(\beta) + \frac{\beta}{1 - \beta} \max_{\theta \in \Theta} \left\{ \log \left(\mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-\frac{\theta+1}{\theta}} \right) - \tau(\theta) \right\}.$$

Moreover, the maximizing consumption and state-contingent future wealth in the conditional (on $\theta \in \Theta$) optimization problem in Equation (S.5) are

$$\begin{aligned} c_{\theta,t} &= (1 - \beta)w_t, \\ w_{\theta,t+1}(z) &= \beta w_t \mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-1} M(z)^{-\frac{1}{\theta+1}}, \end{aligned}$$

and therefore \mathcal{V}_θ takes the form

$$\begin{aligned} \mathcal{V}_\theta(w_t; M) &= \beta \Lambda(M) + (1 - \beta) \log(1 - \beta) + \beta \log(\beta) \\ &\quad + \beta \log \left(\mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-\frac{\theta+1}{\theta}} \right) - \beta \tau(\theta) + \log(w_t). \end{aligned}$$

The proof of Proposition S.3 is contained in Section S.6.3. For this particular specification, the existence of the value function follows from the theorem of Blackwell (1965).⁵⁶

S.4.2 Equilibrium Conditions and Numerical Procedure

Equilibrium in the model falls into one of three cases. These correspond roughly to the three cases described in the static context in Section 4.3 and depicted in Figure 4, but adapted to deal with consumption growth rather than consumption levels. As we explain at the end of this section, checking each of these possible cases is precisely the numerical procedure used in the calibration in Section 4.4:

1. *All consumers choose $\theta = \theta_L$ in equilibrium:* Combining the two optimality conditions from Proposition S.3, the each consumer $i \in [0, 1]$ must satisfy

$$c_{i,t+1}(z) = \beta c_{i,t} \mathbb{E} \left[M^{\frac{\theta_L}{\theta_L+1}} \right]^{-1} M(z)^{-\frac{1}{\theta_L+1}}.$$

Aggregating across consumers, this implies the following relationship between aggregate consumption growth and the pricing kernel:

$$\lambda(z) = \beta \mathbb{E} \left[M^{\frac{\theta_L}{\theta_L+1}} \right]^{-1} M(z)^{-\frac{1}{\theta_L+1}}.$$

The solution to this equation is⁵⁷

$$M(z) = \beta \mathbb{E} \left[\lambda^{-\theta_L} \right]^{-1} (\lambda(z))^{-(\theta_L+1)}. \quad (\text{S.6})$$

⁵⁶It is worth noting that, as with most recursive non-expected-utility models, the usual techniques from Blackwell (1965) can only be applied in a very restricted set of special cases of the ORA model. Fortunately, for other homogeneous specifications, recent results by Marinacci and Montrucchio (2010) can be used to establish the existence and uniqueness of the value function (see also Theorem 7 in Sarver (2012) which builds on their results).

⁵⁷Equation (S.6) illustrates that when consumption growth is i.i.d. across time and all consumers select the same type θ_L , the pricing kernel is the same as for time-separable expected utility with a coefficient of relative risk aversion $\theta_L + 1$ and discount factor $\hat{\beta} = \beta \mathbb{E} [\lambda^{-\theta_L}]^{-1}$. This formula could also be obtained

To check whether or not this is in fact an equilibrium, we need to determine whether θ_L is actually optimal given these prices. Note that by Proposition S.3, $\mathcal{V}_{\theta_L}(w_t; M) \geq \mathcal{V}_{\theta_H}(w_t; M)$ holds if and only if

$$e^{-\tau_L} \mathbb{E} \left[M^{\frac{\theta_L}{\theta_L+1}} \right]^{-\frac{\theta_L+1}{\theta_L}} \geq \mathbb{E} \left[M^{\frac{\theta_H}{\theta_H+1}} \right]^{-\frac{\theta_H+1}{\theta_H}}. \quad (\text{S.7})$$

If this condition is satisfied, then we have found an equilibrium. If not, then there is not an equilibrium in which all consumers choose $\theta = \theta_L$.

2. *All consumers choose $\theta = \theta_H$ in equilibrium:* The analysis is similar to the previous case. To satisfy the consumer optimality conditions given aggregate consumption growth, the pricing kernel must be

$$M(z) = \beta \mathbb{E} \left[\lambda^{-\theta_H} \right]^{-1} (\lambda(z))^{-(\theta_H+1)}. \quad (\text{S.8})$$

If these prices satisfy $\mathcal{V}_{\theta_H}(w_t; M) \geq \mathcal{V}_{\theta_L}(w_t; M)$, which is equivalent to

$$\mathbb{E} \left[M^{\frac{\theta_H}{\theta_H+1}} \right]^{-\frac{\theta_H+1}{\theta_H}} \geq e^{-\tau_L} \mathbb{E} \left[M^{\frac{\theta_L}{\theta_L+1}} \right]^{-\frac{\theta_L+1}{\theta_L}}, \quad (\text{S.9})$$

then we have found an equilibrium. If not, then there is not an equilibrium in which all consumers choose $\theta = \theta_H$.

3. *Consumers are heterogeneous, with some choosing $\theta = \theta_L$ and some choosing $\theta = \theta_H$ in equilibrium:* If both θ_L and θ_H are optimal, then they must give the same indirect utility. This gives our first equilibrium restriction on prices: $\mathcal{V}_{\theta_L}(w_t; M) = \mathcal{V}_{\theta_H}(w_t; M)$ or, equivalently,

$$e^{-\tau_L} \mathbb{E} \left[M^{\frac{\theta_L}{\theta_L+1}} \right]^{-\frac{\theta_L+1}{\theta_L}} = \mathbb{E} \left[M^{\frac{\theta_H}{\theta_H+1}} \right]^{-\frac{\theta_H+1}{\theta_H}}. \quad (\text{S.10})$$

We use individual optimality and market clearing conditions to give our second restriction. Let $\alpha \in (0, 1)$ denote the fraction of time t wealth held by consumers who choose type θ_L . Then the total time t consumption by type θ_L consumers is αe_t , and the total time t consumption by type θ_H consumers is $(1 - \alpha)e_t$. By the consumers' optimality conditions, total consumption of these two groups at time $t + 1$ are as follows:

$$\begin{aligned} \text{type } \theta_L \text{ consumers: } & \alpha \beta e_t \mathbb{E} \left[M^{\frac{\theta_L}{\theta_L+1}} \right]^{-1} M(z)^{-\frac{1}{\theta_L+1}} \\ \text{type } \theta_H \text{ consumers: } & (1 - \alpha) \beta e_t \mathbb{E} \left[M^{\frac{\theta_H}{\theta_H+1}} \right]^{-1} M(z)^{-\frac{1}{\theta_H+1}}. \end{aligned}$$

by appealing to the results of [Kocherlakota \(1990\)](#), who observed the equivalence of the equilibrium conditions for EZKP utility and time-separable expected utility (with a different discount factor) in i.i.d. environments.

Market clearing requires that these sum to e_{t+1} or, equivalently,

$$\lambda(z) = \alpha\beta\mathbb{E}\left[M^{\frac{\theta_L}{\theta_L+1}}\right]^{-1} M(z)^{-\frac{1}{\theta_L+1}} + (1-\alpha)\beta\mathbb{E}\left[M^{\frac{\theta_H}{\theta_H+1}}\right]^{-1} M(z)^{-\frac{1}{\theta_H+1}}. \quad (\text{S.11})$$

If Equations (S.10) and (S.11) are both satisfied, then we have found an equilibrium.

Note that Equation (S.11) implies that $\mathbb{E}[M\lambda] = \beta$. In the two-state version of the model used in the calibration in Section 4.4 with $Z = \{z^l, z^h\}$, it can be shown that a partial converse is also true: If $\mathbb{E}[M\lambda] = \beta$ then Equation (S.11) is satisfied for some $\alpha \in \mathbb{R}$. This observation permits the numerical problem for this case to be reduced to solving an equation of a single variable. Specifically, we adopt the following numerical procedure:

1. Check for an equilibrium as in case 1 as follows: Define M as in Equation (S.6) and check if Equation (S.7) is satisfied. If that fails, then proceed to step 2.
2. Check for an equilibrium as in case 2 as follows: Define M as in Equation (S.8) and check if Equation (S.9) is satisfied. If that fails, then proceed to step 3.
3. Find an equilibrium as in case 3 as follows: Since the solution to Equation (S.10) is only pinned down up to a constant, look for a solution of the form

$$\hat{M}(z) = \begin{cases} p & \text{if } z = z^l \\ 1 & \text{if } z = z^h, \end{cases}$$

for $p \geq 1$. Then let

$$M(z) = \frac{\beta}{\mathbb{E}[\hat{M}\lambda]} \hat{M}(z).$$

Since M is a scalar multiple of \hat{M} it satisfies Equation (S.10), and by construction it satisfies $\mathbb{E}[M\lambda] = \beta$. Thus it satisfies Equation (S.11) for some $\alpha \in \mathbb{R}$. Finally, it can be shown that if there is not an equilibrium as in case 1 or 2, then we necessarily have $\alpha \in (0, 1)$ and hence the market clearing condition is satisfied. Thus we have found a heterogeneous-type equilibrium.

S.4.3 Asset Returns

In this section, we derive the formulas for asset returns for the calibration in Section 4.4 using the equilibrium pricing kernel. By the definition, the period t price of any random asset paying x_{t+1} in period $t+1$ is $p_t = \mathbb{E}_t[M_{t,t+1}x_{t+1}]$. The risk-free rate is therefore given by

$$1 = \mathbb{E}[M_{t,t+1}R_t^f] \quad \text{or} \quad R_t^f = \frac{1}{\mathbb{E}_t[M_{t,t+1}]}.$$

Since the equilibrium pricing kernel is stationary and satisfies $M_{t,t+1} = M(z_{t+1})$, the risk-free rate is constant and is given by the unconditional expectation $R^f = 1/\mathbb{E}[M]$.

Consider now an asset that has a dividend growth given by $d_{t+1} = \lambda^d(z_{t+1})d_t$. The price of this asset satisfies

$$p_t^d = \mathbb{E}_t[M_{t,t+1}(p_{t+1}^d + d_{t+1})].$$

The price/dividend ratio therefore satisfies

$$\frac{p_t^d}{d_t} = \mathbb{E}_t \left[M_{t,t+1} \left(\frac{p_{t+1}^d}{d_{t+1}} \frac{d_{t+1}}{d_t} + \frac{d_{t+1}}{d_t} \right) \right].$$

Since the state is i.i.d. across time and the equilibrium pricing kernel is given by $M_{t,t+1} = M(z_{t+1})$ at every period t , there is a constant price/dividend ratio p that solves this equation:

$$p = \mathbb{E}[M\lambda^d(p + 1)] \implies \frac{p}{p + 1} = \mathbb{E}[M\lambda^d].$$

The return on this asset is given by

$$R_{t+1} = \frac{p_{t+1}^d + d_{t+1}}{p_t^d} = \frac{\frac{p_{t+1}^d}{d_{t+1}} \frac{d_{t+1}}{d_t} + \frac{d_{t+1}}{d_t}}{\frac{p_t^d}{d_t}}.$$

The return is therefore time invariant and satisfies

$$R(z) = \lambda^d(z) \left(\frac{p + 1}{p} \right) = \frac{\lambda^d(z)}{\mathbb{E}[M\lambda^d]}.$$

S.5 Existence of a Value Function

The value function V is included explicitly in the definition of the ORA representation. However, it may be desirable to obtain such a value function from the other parameters (u, Φ, β) of the representation. Using similar techniques to [Epstein and Zin \(1989\)](#), the following result shows that this is possible.⁵⁸

Proposition S.4 *Suppose $\beta \in (0, 1)$ and $u : C \rightarrow \mathbb{R}$ is a continuous and nonconstant function. Let $[\hat{a}, \hat{b}] = u(C)$ where $-\infty < \hat{a} < \hat{b} < +\infty$,⁵⁹ and let $a = \frac{\hat{a}}{1-\beta}$, $b = \frac{\hat{b}}{1-\beta}$. Let Φ be any collection of continuous and nondecreasing functions $\phi : [a, b] \rightarrow \mathbb{R}$ that satisfies $\sup_{\phi \in \Phi} \phi(x) = x$ for all $x \in [a, b]$. Then, there exists a bounded and lower semicontinuous function $V : D \rightarrow [a, b]$ that satisfies Equation (3).*

⁵⁸As with most other recursive non-expected-utility models, it is not possible to apply the standard techniques from [Blackwell \(1965\)](#) to prove existence of a value function.

⁵⁹Since C is compact and connected and u is continuous, $u(C)$ is a closed and bounded interval in \mathbb{R} .

For the ORA 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, Proposition S.4 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 $[a, b]$. In particular, if $u \geq 0$ ($u \leq 0$, respectively) then it suffices to define each ϕ on \mathbb{R}_+ (\mathbb{R}_- , respectively).

There are two noticeable gaps in Proposition S.4. 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. There are similar limitations to the existence results in Epstein and Zin (1989). Since resolving these issues is not central to the analysis in this paper, obtaining a stronger version of this result is left as an open question for future research. However, it is worth noting that in the case of homothetic preferences, as in the application in Section 4, it is possible to ensure both uniqueness and continuity of the value function using recent results from Marinacci and Montrucchio (2010) (see also Proposition S.3 and Footnote 56).

S.6 Proofs

S.6.1 Proof of Theorem S.1

Theorem S.1 will be proved by means of a separation argument. Since $\geq_c = \geq_{\langle \mathcal{C} \rangle}$, it is without loss of generality to assume that \mathcal{C} is a closed convex cone in the space of continuous functions $C(X)$ that contains the constant functions. Let $ca(X)$ denote the set of all signed (countably-additive) Borel measures of bounded variation on the compact metric space X . Consider the following subset of $ca(X)$:

$$K \equiv \left\{ \mu \in ca(X) : \int \phi(x) d\mu(x) \geq 0 \text{ for every } \phi \in \mathcal{C} \right\}. \quad (\text{S.12})$$

Note that K is a cone in $ca(X)$. In addition, since the constant functions identically equal to 1 and -1 are both in \mathcal{C} , $\mu(X) = 0$ for all $\mu \in K$. The following lemma makes some other simple observations about K that will be used in the proof of the proposition.

Lemma S.1 *The set K defined in Equation (S.12) is a weak* closed convex cone in $ca(X)$, and for any $\mu, \eta \in \Delta(X)$,*

$$\mu \geq_c \eta \iff \mu - \eta \in K.$$

Proof: For any $\phi \in \mathcal{C}$, the set

$$K_\phi \equiv \left\{ \mu \in ca(X) : \int \phi(x) d\mu(x) \geq 0 \right\}$$

is weak* closed and convex. Since $K = \bigcap_{\phi \in \mathcal{C}} K_\phi$ is the intersection of closed and convex sets, it is also closed and convex. The equivalence in the displayed equation follows directly from the definition of $\geq_{\mathcal{C}}$. \blacksquare

Continuing the proof of Theorem S.1, it is immediate that 2 implies 1. To prove that 1 implies 2, it suffices to show that for any $\mu \in \Delta(X)$ and any $\alpha < W(\mu)$, there exists a function $\phi_{\mu,\alpha} \in \mathcal{C}$ 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 $\eta \in \Delta(X)$. Then, letting

$$\Phi = \{\phi_{\mu,\alpha} : \mu \in \Delta(X), \alpha < W(\mu)\},$$

it follows directly that

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

Fix any $\mu \in \Delta(X)$ 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:

$$\text{epi}(W) = \{(\eta, t) \in \Delta(X) \times \mathbb{R} : t \geq W(\eta)\}$$

Since W is convex with a convex domain $\Delta(X)$, $\text{epi}(W)$ is a convex subset of $ca(X) \times \mathbb{R}$. Moreover, as a weak* lower semicontinuous function with a weak* closed domain, it is a standard result that $\text{epi}(W)$ is a closed subset of $ca(X) \times \mathbb{R}$.⁶⁰ Now, define a set $K_{\mu,\alpha}$ as follows:

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

By Lemma S.1, $K_{\mu,\alpha}$ is a closed and convex subset of $ca(X) \times \mathbb{R}$. Establishing the following claim allows the separating hyperplane theorem to be applied.⁶¹

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

Proof of claim: First, note that $K_{\mu,\alpha} \cap \text{epi}(W) = \emptyset$. To see this, take any $(\eta, t) \in K_{\mu,\alpha}$. Then, by definition, $t = \alpha$ and $\eta - \mu \in K$. If $\eta \notin \Delta(X)$, then it is trivial that $(\eta, t) \notin \text{epi}(W)$. Alternatively, if $\eta \in \Delta(X)$, then Lemma S.1 implies $\eta \geq_{\mathcal{C}} \mu$. In this case $W(\eta) \geq W(\mu) > \alpha = t$, so again $(\eta, t) \notin \text{epi}(W)$. Thus, $K_{\mu,\alpha}$ and $\text{epi}(W)$ are disjoint, closed, and convex sets.

⁶⁰The set $ca(X) \times \mathbb{R}$ is endowed with the product topology generated by the weak* topology on $ca(X)$ and the Euclidean topology on \mathbb{R} .

⁶¹Although $\text{epi}(W)$ and $K_{\mu,\alpha}$ 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.

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

$$\begin{aligned} B_1 &\equiv \text{epi}(W) \cap (\Delta(X) \times [\underline{W}, W(\mu)]) = \{(\eta, t) \in \Delta(X) \times \mathbb{R} : W(\mu) \geq t \geq W(\eta)\} \\ B_2 &\equiv \text{epi}(W) \cap (\Delta(X) \times [W(\mu), +\infty)) = \{(\eta, t) \in \Delta(X) \times \mathbb{R} : t \geq \max\{W(\eta), W(\mu)\}\}. \end{aligned}$$

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

Continuing the proof of Theorem S.1, note that $ca(X) \times \mathbb{R}$ 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 $F : ca(X) \rightarrow \mathbb{R}$ and $\lambda \in \mathbb{R}$ such that

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

For any $(\eta, t) \in \text{epi}(W)$ and $\nu \in K$, we have $(\mu + \nu, \alpha) \in K_{\mu,\alpha}$ and therefore $(\eta - \mu - \nu, t - \alpha) \in \text{epi}(W) - K_{\mu,\alpha}$. Thus

$$F(\eta) + \lambda t < F(\mu) + F(\nu) + \lambda \alpha, \quad \forall (\eta, t) \in \text{epi}(W), \forall \nu \in K. \quad (\text{S.13})$$

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 (S.13), conclude that for all $\eta \in \Delta(X)$,

$$F(\eta) + \lambda W(\eta) < F(\mu) + \lambda \alpha \implies W(\eta) > -\frac{F(\eta)}{\lambda} + \frac{F(\mu)}{\lambda} + \alpha.$$

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

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

In addition,

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

The final step in the proof is to show that $\phi_{\mu,\alpha} \in \mathcal{C}$. Fix any $\nu \in K$, and note that $r\nu \in K$ for all $r > 0$. Therefore, Equation (S.13) implies

$$\begin{aligned} F(\mu) + \lambda W(\mu) &< F(\mu) + F(r\nu) + \lambda\alpha \\ &= F(\mu) + rF(\nu) + \lambda\alpha. \end{aligned}$$

For this to be true for every $r > 0$, it must be that $F(\nu) \geq 0$. That is, $\int \psi(x) d\nu(x) \geq 0$ for all $\nu \in K$. Thus there is no $\nu \in ca(X)$ such that

$$\int \psi(x) d\nu(x) < 0 \quad \text{and} \quad \int \phi(x) d\nu(x) \geq 0, \quad \forall \phi \in \mathcal{C}.$$

An infinite-dimensional version of Farkas' Lemma (Corollary 5.84 in [Aliprantis and Border \(2006\)](#)) therefore implies $\psi \in \mathcal{C}$. Since \mathcal{C} is a convex cone that contains all constant functions, conclude also that $\phi_{\mu,\alpha} \in \mathcal{C}$. This completes the proof.

S.6.2 Proof of Proposition S.1

Fix any w, L, π and any $y, y' \in [0, L]$. By definition, the individual is indifferent between insurance coverage y at premium $P(y)$ and insurance coverage y' at premium $P(y')$:

$$\begin{aligned} W \left(\begin{array}{cc} w - P(y) - L + y & \pi \\ w - P(y) & 1 - \pi \end{array} \right) &= W \left(\begin{array}{cc} w - P(y') - L + y' & \pi \\ w - P(y') & 1 - \pi \end{array} \right) \\ &= W \left(\begin{array}{cc} w - L & \pi \\ w & 1 - \pi \end{array} \right). \end{aligned}$$

Preference for diversification therefore implies that for any $\alpha \in [0, 1]$,

$$W \left(\begin{array}{cc} w - \alpha P(y) - (1 - \alpha)P(y') - L + \alpha y + (1 - \alpha)y' & \pi \\ w - \alpha P(y) - (1 - \alpha)P(y') & 1 - \pi \end{array} \right) \geq W \left(\begin{array}{cc} w - L & \pi \\ w & 1 - \pi \end{array} \right).$$

Since W respects FOSD, this implies $P(\alpha y + (1 - \alpha)y') \geq \alpha P(y) + (1 - \alpha)P(y')$, so P is concave. Thus the individual has a nonincreasing marginal willingness to pay for insurance coverage.

S.6.3 Proof of Proposition S.3

Note that for any function $f : \mathbb{R}_+^Z \rightarrow \mathbb{R}$ and any $\alpha \in \mathbb{R}$, the certainty equivalent in Equation (S.4) satisfies $\mathcal{R}(f + \alpha) = \mathcal{R}(f) + \alpha$. Therefore, the value function operator defined by the right side of Equation (S.3) satisfies the conditions of the theorem of [Blackwell](#)

(1965), which ensures the existence and uniqueness of the value function $\mathcal{V}(w_t; M)$ for any $M : Z \rightarrow \mathbb{R}_{++}$.

We will verify that the functional form $\mathcal{V}(w_t; M) = \Lambda(M) + \log(w_t)$ satisfies the recursion in Equation (S.3). Assuming this form of the value function and substituting into Equation (S.5) gives

$$\mathcal{V}_\theta(w_t) = \max_{\substack{c_t \in \mathbb{R}_+ \\ w_{t+1} \in \mathbb{R}_+^Z}} \left\{ (1 - \beta) \log(c_t) + \beta \log \left(\mathbb{E}[w_{t+1}^{-\theta}]^{-\frac{1}{\theta}} \right) - \beta \tau(\theta) + \beta \Lambda(M) \right\}, \quad (\text{S.14})$$

where the maximization is subject to Equation (S.2). Applying standard Lagrangian optimization techniques to this problem gives the following maximizers:

$$\begin{aligned} c_{\theta,t} &= (1 - \beta)w_t, \\ w_{\theta,t+1}(z) &= \beta w_t \mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-1} M(z)^{-\frac{1}{\theta+1}}. \end{aligned}$$

Note that these solutions imply

$$\begin{aligned} \mathbb{E}[w_{\theta,t+1}^{-\theta}]^{-\frac{1}{\theta}} &= \beta w_t \mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-1} \mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-\frac{1}{\theta}} \\ &= \beta w_t \mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-\frac{\theta+1}{\theta}}. \end{aligned}$$

Substituting these solutions into Equation (S.14) therefore yields

$$\begin{aligned} \mathcal{V}_\theta(w_t; M) &= (1 - \beta) \log((1 - \beta)w_t) + \beta \log(\beta w_t) \\ &\quad + \beta \log \left(\mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-\frac{\theta+1}{\theta}} \right) - \beta \tau(\theta) + \beta \Lambda(M) \\ &= \beta \Lambda(M) + (1 - \beta) \log(1 - \beta) + \beta \log(\beta) \\ &\quad + \beta \log \left(\mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-\frac{\theta+1}{\theta}} \right) - \beta \tau(\theta) + \log(w_t). \end{aligned}$$

This will establish that \mathcal{V}_θ has the form claimed in the statement of the proposition, once it is verified that \mathcal{V} takes the form assumed above. To see that the latter is true, note that

$$\begin{aligned} \mathcal{V}(w_t; M) &= \max_{\theta \in \Theta} \mathcal{V}_\theta(w_t; M) \\ &= \beta \Lambda(M) + (1 - \beta) \log(1 - \beta) + \beta \log(\beta) \\ &\quad + \beta \max_{\theta \in \Theta} \left\{ \log \left(\mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-\frac{\theta+1}{\theta}} \right) - \tau(\theta) \right\} + \log(w_t). \end{aligned}$$

Therefore, the assumed form $\mathcal{V}(w_t; M) = \Lambda(M) + \log(w_t)$ satisfies the recursion in Equa-

tion (S.3) if

$$\Lambda(M) = \log(1 - \beta) + \frac{\beta}{1 - \beta} \log(\beta) + \frac{\beta}{1 - \beta} \max_{\theta \in \Theta} \left\{ \log \left(\mathbb{E} \left[M^{\frac{\theta}{\theta+1}} \right]^{-\frac{\theta+1}{\theta}} \right) - \tau(\theta) \right\}.$$

We have thus established all of the claims in the statement of the proposition.

S.6.4 Proof of Proposition S.4

As in the statement of the proposition, let $[\hat{a}, \hat{b}] = u(C)$ and $a = \frac{\hat{a}}{1-\beta}$, $b = \frac{\hat{b}}{1-\beta}$. Let L denote the space of all lower semicontinuous functions from D to $[a, b]$:

$$L \equiv \{f : D \rightarrow [a, b] : f \text{ is lower semicontinuous}\}.$$

Define an operator T on L by

$$Tf(c, m) = u(c) + \beta \sup_{\phi \in \Phi} \int_D \phi(f(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}),$$

for $(c, m) \in D$.

The first step is to show that $Tf \in L$ for all $f \in L$, and hence $T : L \rightarrow L$. Fix any $f \in L$. Since f is bounded by a and b and each ϕ is nondecreasing, it follows that

$$\phi(a) \leq \int_D \phi(f(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}) \leq \phi(b), \quad \forall m \in \Delta(D), \phi \in \Phi.$$

Taking the supremum of each expression and using the property $\sup_{\phi \in \Phi} \phi(x) = x$ gives

$$a \leq \sup_{\phi \in \Phi} \int_D \phi(f(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m}) \leq b.$$

Since $(1 - \beta)a \leq u(c) \leq (1 - \beta)b$ for all $c \in C$, this implies $a \leq Tf \leq b$. Next, the lower semicontinuity of f implies that $\phi \circ f$ is lower semicontinuous for all $\phi \in \Phi$, since each ϕ is continuous and nondecreasing. This in turn implies that the mapping $m \mapsto \int_D \phi(f(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m})$ is lower semicontinuous (see Theorem 15.5 in [Aliprantis and Border \(2006\)](#)). It is a standard result that the supremum of any collection of lower semicontinuous functions is lower semicontinuous. Together with the continuity of u , conclude that Tf is lower semicontinuous. Hence $Tf \in L$ for all $f \in L$.

The proof is completed by showing that T has a fixed point $V \in L$. To show the existence of a fixed point, first construct a sequence as follows: Let $V_1(c, m) = a$ for all $(c, m) \in D$, and let $V_{n+1} = TV_n$ for all $n \in \mathbb{N}$. Since each $\phi \in \Phi$ is nondecreasing, it follows immediately that T is monotone: $f \leq g$ implies $Tf \leq Tg$. Note that $V_1 \leq V_2$, since $V_1 \leq g$ for any $g \in L$ by definition. Thus $V_2 = TV_1 \leq TV_2 = V_3$. Proceeding by induction, $V_n \leq V_{n+1}$ for all $n \in \mathbb{N}$. Since $\{V_n\}$ is an increasing sequence of bounded

functions, it converges pointwise to some function $V : D \rightarrow [a, b]$. Moreover, since V is equal to the supremum of the collection of lower semicontinuous functions $\{V_n : n \in \mathbb{N}\}$, it is lower semicontinuous. Hence $V \in L$.

The last step is to show $TV = V$. Since $V_n \leq V$ for all n , monotonicity of the operator T implies $V_{n+1} = TV_n \leq TV$. Taking limits gives $V \leq TV$. To establish the opposite inequality, note first that for any $m \in \Delta(D)$ and $\phi \in \Phi$, the mapping $f \mapsto \int_D \phi(f(\hat{c}, \hat{m})) dm(\hat{c}, \hat{m})$ is continuous in the product topology (i.e., the topology of pointwise convergence) by the dominated convergence theorem. This implies the mapping $f \mapsto Tf(c, m)$ is lower semicontinuous for all $(c, m) \in D$. Thus $V_n \rightarrow V$ implies

$$TV(c, m) \leq \liminf_n TV_n(c, m) = \liminf_n V_{n+1}(c, m) = V(c, m),$$

for all $(c, m) \in D$. Hence $TV = V$, which completes the proof.