ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 86, No. 4 (July, 2018), 1347–1382

DYNAMIC MIXTURE-AVERSE PREFERENCES

TODD SARVER

Department of Economics, Duke University

Econometrica, Vol. 86, No. 4 (July, 2018), 1347–1382

DYNAMIC MIXTURE-AVERSE PREFERENCES

TODD SARVER
Department of Economics, Duke University

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. We describe some useful parametric examples of our representation and provide comparative statics that tightly link decreases in risk aversion to larger sets of feasible risk attitudes. We then present several applications of the model. In an insurance problem, mixture-averse preferences can produce a marginal willingness to pay for insurance coverage that increases in the level of existing coverage. In investment decisions, our model can generate endogenous heterogeneity in equilibrium stock market participation, even when consumers have identical preferences. Finally, we demonstrate that our model can address the Rabin paradox even in the presence of reasonable levels of background risk.

KEYWORDS: Mixture aversion, optimal risk attitude, insurance deductible, stock market participation, Rabin paradox.

1. INTRODUCTION

INTERTEMPORAL PREFERENCES OF INDIVIDUALS ARE of central importance in many economic interactions. They play a key role in 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]

Todd Sarver: todd.sarver@duke.edu

I am grateful to David Ahn, Philipp Sadowski, the editor, and four anonymous referees for detailed feedback that led to significant improvements in this paper. I also thank Simone Cerreia-Vioglio, Chris Chambers, Eddie Dekel, David Dillenberger, Jiayun Dong, 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 numerous seminar participants for helpful comments and discussions. This paper was previously circulated under the titles "Optimal Reference Points and Anticipation" and "Mixture-Averse Preferences and Heterogeneous Stock Market Participation."

[1]In 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.

 https://doi.org/10.3982/ECTA12687

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, our mixture aversion axiom has a connection 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. However, our axiom is not simply a manifestation of the certainty effect, as it applies more broadly to mixtures of any lotteries, not just at or near certainty. For example, an individual up for promotion may be uncertain of the exact job characteristics of the new position. In this case, even if the promotion is received for certain, the individual faces a lottery over different payoff-relevant outcomes; additional uncertainty about whether or not the promotion will be received can then be formalized as a mixture of lotteries. Mixture aversion implies that the individual places a premium on securing the promotion with certainty even when residual risk remains (the outcome conditional on being promoted or not).

The notion that individuals dislike mixtures of lotteries is not without precedent. As an illustration, consider the well-documented case of 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 (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, in contrast, are consistent with such behavior: Since risk of insurance nonpayment can be formalized as a mixture of the degenerate lottery representing certain coverage with the (nontrivial) lottery representing risk of loss without insurance, aversion to probabilistic insurance is almost directly implied by our axiom.

Our main result establishes that mixture-averse preferences can be represented using an extension of dynamic expected utility in which the individual can optimally select her risk attitude subject to a constraint or cost. After exploring several examples and developing comparative statics for the representation, we show that our model can generate realistic and qualitatively novel predictions in a number of applications, including demand for insurance and participation in equity markets. In the remainder of the Introduction, we provide an overview of the model, our applications, and the related literature.

[2]There 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 nonpayment from an insurance policy.

[3]For example, in one version of the probabilistic insurance problem, Kahneman and Tversky (1979) showed that the preferences of the majority of subjects were inconsistent with expected-utility theory. Their argument can easily be extended to show that their experimental results are 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 Supplemental Material (Sarver (2018))).

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 the 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.[5] The representation for mixture-averse preferences has a value function of the form

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

In this recursion, the random variables c_t for current consumption and w_{t+1} for future wealth 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

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

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 later used in our applications.

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},

$$\sup_{\phi \in \Phi} \mathbb{E}_t\big[\phi\big(\mathcal{V}(w_{t+1})\big)\big] \le \mathbb{E}_t\Big[\sup_{\phi \in \Phi} \phi\big(\mathcal{V}(w_{t+1})\big)\Big] = \mathbb{E}_t\big[\mathcal{V}(w_{t+1})\big].$$

In fact, one significant new feature of the ORA representation is its ability to generate high levels of risk aversion for small gambles yet more moderate attitudes toward increases in risk when exposure is already large. For example, when current risk exposure is small, the individual may be able to choose a transformation ϕ that gives high utility near the (almost) certain outcome, but is extremely sensitive to gains and losses, and she

[4]The 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.

[5]The assumption of i.i.d. returns is made simply for expositional convenience in the Introduction. No such independence assumptions are imposed in our general framework or main results.

may therefore be very reluctant to take on any additional risk on the margin. Alternatively, when current risk exposure is already large, the individual may be able to choose a different transformation that provides lower utility for low-risk allocations, but is also less sensitive to gains and losses, and she may therefore be more willing to accept additional risk on the margin. For instance, some of the examples explored in Section 3.2 can be interpreted as if the individual can decrease her level of risk aversion by incurring some fixed mental cost (i.e., choose a transformation that is less concave but also shifted downward). In this way, aversion to marginal increases in risk could actually decrease with exposure. This simple feature, which plays a central role in our applications, is surprisingly difficult to generate with existing models, as we discuss in more detail in the next section.

Section 4 contains several applications of our model. In Section 4.1, we apply the ORA representation to a simple insurance problem. 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. In Section 4.2, we apply our model to financial markets and show that it can generate endogenous heterogeneity in risk exposure, with one segment of an ex ante identical population holding significantly greater risk in equilibrium. In the companion paper Sarver (2017), we expanded our analysis and built on this feature of the model to help 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. Section 4.3 illustrates how the ORA representation can provide an explanation for the Rabin paradox that does not rely on first-order risk aversion and that is robust to moderate levels of background risk. Finally, Section 4.4 relates the model to the experimental evidence that most violations of expected-utility theory occur near the boundary of the probability simplex.

Proofs and additional results are contained in the Appendix, and some discussions and supporting results are further relegated to the Supplemental Material (Sarver (2018)).

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. In the context of the simple consumption-savings problem described above, the value function for EZKP utility takes the form

$$\mathcal{V}(w_t) = \max_{c_t, w_{t+1}} \{ u(c_t) + \beta h^{-1}(\mathbb{E}_t[h(\mathcal{V}(w_{t+1}))]) \},$$

for some strictly increasing function h. As emphasized by Epstein and Zin (1989) and Weil (1989, 1990), this model permits a separation between risk aversion and the elasticity of intertemporal substitution that is not possible for standard *time-separable* expected utility.[6] Despite its usefulness, EZKP utility is still unable to resolve a number of anomalies

[6]Time-separable expected utility refers to the standard model that is separable with respect to both states and time, that is, the special case of EZKP utility where h is the identity function.

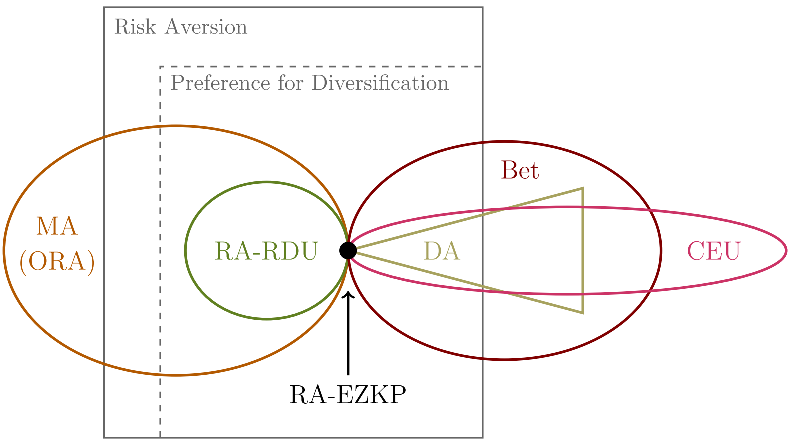


FIGURE 1.—Relationship between mixture-averse (MA) preferences (ORA representation) and other recursive risk preferences: risk-averse Epstein–Zin–Kreps–Porteus expected utility (RA-EZKP), risk-averse rank-dependent utility (RA-RDU), betweenness (Bet), disappointment aversion (DA), cautious expected utility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

associated with expected utility, such as the Allais and Rabin paradoxes. In order to overcome these limitations and to help address the equity premium puzzle and other related puzzles in finance, a large literature has developed that studies various non-expected-utility theories within the recursive framework of Epstein and Zin (1989).

Figure 1 illustrates the relationship between the ORA representation and the recursive formulation of several of these prominent non-expected-utility theories: rank-dependent utility (Quiggin (1982), Yaari (1987), Segal (1989)), betweenness (Chew (1983), Dekel (1986)), disappointment aversion (Gul (1991)), and cautious expected utility (Cerreia-Vioglio, Dillenberger, and Ortoleva (2015)).[7] The relationship between these preferences and mixture-averse preferences can be established using results for their static counterparts from Wakker (1994) and Grant, Kajii, and Polak (2000), which we review in more detail in Section S.2 of the Supplemental Material (Sarver (2018)). The connection between mixture-averse preferences and EZKP preferences is established in Section 3.2, where we show that risk-averse EZKP utility (concave h) can be mapped into a special case of the ORA representation; in addition, we provide an axiomatic analysis of the EZKP representation and characterize the overlap with mixture aversion in Appendix A.2.

To highlight some of the significant new features of our model, Figure 1 also illustrates two properties of risk preferences that will be especially relevant for our applications. The first is risk aversion, meaning monotonicity with respect to second-order stochastic dominance. The second is preference for diversification. When risk is modeled using consumption-valued random variables rather than lotteries, preference for diversification (or preference for hedging) refers to quasiconcavity of preferences with respect to state-wise convex combinations of these random variables.[8] Preference for diversification is

[7]Some notable applications of these recursive risk preferences to finance and macroeconomics include: a parameterized special case of rank-dependent utility by Epstein and Zin (1990); the disappointment aversion model by Bekaert, Hodrick, and Marshall (1997) and Ang, Bekaert, and Liu (2005); other special cases of betweenness preferences by Epstein and Zin (2001) and Routledge and Zin (2010).

[8]We 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

often considered a desirable property since it enhances analytic tractability, and it is arguably a plausible descriptive property of preferences in some circumstances. However, it turns out that several applications of our model, including deriving some intuitive properties of demand for insurance (Section 4.1) and obtaining endogenous heterogeneity in asset market participation (Section 4.2), rely crucially on relaxing preference for diversification. At the same time, it is often desirable (for both realism and technical simplicity) to maintain risk aversion in such applications. Mixture-averse preferences are unique among those depicted in Figure 1 in their ability to separate these two properties. As we explain in greater detail in Section S.3 of the Supplemental Material (Sarver (2018)), results from Dekel (1989) and Chew, Karni, and Safra (1987) imply that none of the other preferences depicted in Figure 1 can relax preference for diversification without also violating risk aversion.

2. PRELIMINARIES

2.1. *Framework*

For any topological space X, let $\triangle(X)$ denote the set of all (countably-additive) Borel probability measures on X, endowed with the topology of weak convergence (or weak* topology). This topology is metrizable if X is a separable metrizable space. 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.[9] A one-period consumption lottery is simply an element of $\triangle(C)$. The space of two-period temporal lotteries is $\triangle(C \times \triangle(C))$, the space of three-period temporal lotteries is $\triangle(C \times \triangle(C \times \triangle(C)))$, 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 $C \times \triangle(D)$. Epstein and Zin (1989) showed that such a space is well-defined.[10] Intuitively, a 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 $m \in \triangle(D)$. The primitive of the axiomatic model is a binary relation $\succsim$ on the space D.

quasiconvexity of preferences in probabilities (see Section 3.1) and is compatible with (but does not require) 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 sufficient to ensure the optimality of fully diversified portfolios (e.g., if asset returns are i.i.d.).

[9] While the axiomatic analysis will be restricted to compact spaces, the applications in Sections 4 will permit 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 extension would result in additional technical complications and add little to the behavioral insights of the current analysis.

[10] See 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.

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 $W : \triangle([a, b]) \to \mathbb{R}$ that satisfies $W(\delta_x) = x$ for all $x \in [a, b]$ and is monotone with respect to first-order stochastic dominance.

In a static setting involving monetary gambles, the certainty equivalent is a familiar object: It gives the sure amount of money that the individual likes the same as a given lottery. In our recursive setting, the idea is the same but the certainty equivalent is instead applied to distributions over continuation values, yielding a *risk-adjusted continuation value*—the deterministic continuation value that the individual likes equally well. For any measurable value function $V : D \to [a, b]$ and any probability $m \in \triangle(D)$, let $m \circ V^{-1}$ denote the distribution (on $[a, b]$) of continuation values induced by m.[11]

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

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

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.[12] 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.

3. DYNAMIC MIXTURE-AVERSE PREFERENCES

3.1. *Main Axiom and Representation Result*

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

AXIOM 1—Mixture Aversion: *For any $c, c' \in C$ and $m, m' \in \triangle(D)$,*

$$\left(c, \frac{1}{2}m + \frac{1}{2}m'\right) \succsim (c', m) \quad \Longrightarrow \quad (c, m') \succsim \left(c', \frac{1}{2}m + \frac{1}{2}m'\right).$$

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

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

Axiom 1 puts structure on an 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 weight assigned to the lotteries m and m', this axiom implies that the benefit of increasing the weight of m' in the mixture from zero to one-half is (weakly) less than the benefit of increasing the weight from one-half to one. For example, increasing the probability of a future promotion from 0% to 50% may be less valuable (measured in terms of current effort) than increasing the probability from 50% to 100%.[13]

One interpretation of this pattern in choice is that an individual may 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.

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

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

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

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

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:

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

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.4 of the Supplemental Material (Sarver (2018)) 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

[13]Axiom 1 also relates to concepts from traditional demand theory: Consider a consumer maximization problem with a hypothetical budget constraint over current consumption and the probabilities of future outcomes. Then mixture aversion implies that current consumption is an inferior good. Formally, Axiom 1 is equivalent to the definition of $(\triangledown_i^\lambda, \triangle_i^\lambda)$-quasisubmodularity that Quah (2007) used to characterize inferior goods, taking $\lambda = 1/2$ and i equal to the consumption dimension of $C \times \triangle(D)$. Note that this property is stronger than the traditional definition of quasisubmodularity studied by Milgrom and Shannon (1994), which holds trivially for any function that is additively separable in c and m.

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 parametric examples of the ORA representation 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*:

1. *The relation $\succsim$ satisfies Axiom 1.*
2. *The certainty equivalent W in the EZ representation of $\succsim$ is convex in probabilities.*[16]
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.[17] 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 condition 3 comes from standard duality results. Since W is convex, it can be expressed as the supremum of some collection of affine functions (Aliprantis and Border (2006, Theorem 7.6)). Since any affine function on $\triangle([a, b])$ can be given an expected-utility representation, this implies there exists a collection Φ of continuous functions $\phi : [a, b] \to \mathbb{R}$ such that, for any $\mu \in \triangle([a, b])$,

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

[14]The 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.

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

[16]That is, $W(\alpha\mu + (1-\alpha)\eta) \leq \alpha W(\mu) + (1-\alpha)W(\eta)$ for all $\mu, \eta \in \triangle([a, b])$ and $\alpha \in [0, 1]$.

[17]Formally, 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 Cerreia-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.

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 $(c, m) \in D$,

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

The only missing step in this sketch is showing that the collection Φ contains only *nondecreasing* functions. Since a certainty equivalent W is by definition monotone with respect to FOSD, the following proposition establishes this property. In addition, the second part of the proposition characterizes the relationship between the set of transformations and SOSD monotonicity of the certainty equivalent, which provides a simple set of sufficient conditions for risk aversion of an ORA representation.[18]

PROPOSITION 1: *Suppose $W : \triangle([a, b]) \to \mathbb{R}$ 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* (5) *for some collection Φ of nondecreasing continuous functions $\phi : [a, b] \to \mathbb{R}$.*
2. *W is monotone with respect to SOSD if and only if it satisfies Equation* (5) *for some collection Φ of nondecreasing and concave continuous functions $\phi : [a, b] \to \mathbb{R}$.*

Proposition 1 is a special case of a more general local expected-utility result for any stochastic order that we present in Section S.1 of the Supplemental Material (Sarver (2018)). In that section, we also discuss how our results complement existing local expected-utility results from Machina (1982) and the literature that followed (see Cerreia-Vioglio, Maccheroni, and Marinacci (2017) for a recent generalization). In short, this literature requires (Fréchet or Gateaux) differentiability of the function W. Since our representation is in general not differentiable, our results are instead based on convexity.

3.2. *Parametric Examples*

In this section, we describe several special cases of the optimal risk attitude certainty equivalent. These parametric examples will be useful for illustrating the connections and departure points from previous models, and they will be used in the comparative statics and applications sections.

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 will determine sensitivity to risk. 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:

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

[18]For example, if the function u in the ORA representation is also concave, then concavity of the set of transformations Φ implies risk aversion with respect to consumption.

The following examples satisfy these conditions.

EXAMPLE 1—Smooth Transformation: For $\gamma \in \mathbb{R}$ and $\theta > 0$, consider the parameterized function

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

EXAMPLE 2—Kinked Transformation: For $\gamma \in \mathbb{R}$ and $\theta \in [0, 1]$, consider the parameterized function

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

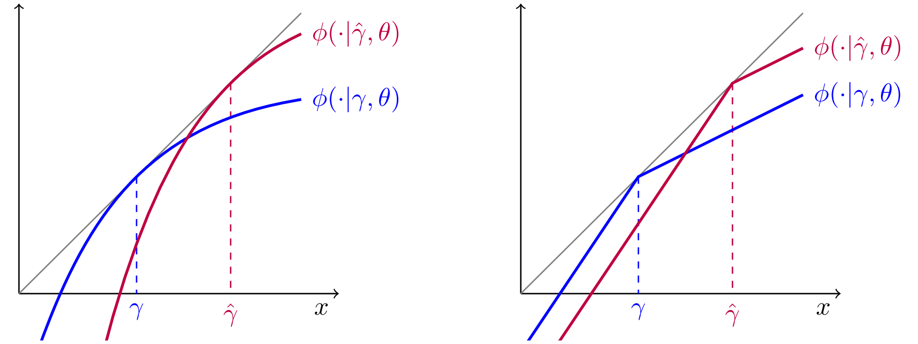
In the simplest class of examples, θ is a fixed parameter and the individual optimizes only over γ. The set of transformations is therefore $\Phi = \{\phi(\cdot|\gamma, \theta) : \gamma \in \Gamma\}$, which implies the certainty equivalent can be written as

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

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 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 related models. 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

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



(a) Example 1: smooth risk attitude transformation function

(b) Example 2: kinked risk attitude transformation function

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

for some increasing and concave function φ that satisfies $\varphi(0) = 0$ and $\varphi(x) \leq x$.[19] 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. We demonstrate later in this section that this connection is one particular instance of a more general relationship between recursive expected utility and the ORA representation. 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).[20]

Some of the most novel implications of our model arise 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 (cf. Corollary 2), but at some psychological cost. Formally, let Θ be any subset of real numbers, let $\tau : \Theta \to \mathbb{R}$ be a "cost" function that satisfies $\inf_{\theta \in \Theta} \tau(\theta) = 0$, and define a collection of transformations by

$$\Phi = \{\phi(\cdot|\gamma, \theta) - \tau(\theta) : \gamma \in \Gamma, \theta \in \Theta\}. \tag{10}$$

The resulting certainty equivalent can be written as

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

The applications explored in Section 4 involve ORA representations with certainty equivalents taking the form in Equation (11). This class of examples permits the novel feature of the model discussed in the Introduction: Aversion to marginal increases in risk may decrease with exposure. One particular special case of interest will be a simple extension of Epstein–Zin–Kreps–Porteus (EZKP) expected utility where local risk aversion, parameterized by θ, is tailored to the specific risk being faced. The following proposition shows how EZKP utility, as well as this extension, can be expressed as a special case of the ORA representation, provided the transformation h is concave (see also Appendix A.2 for an axiomatic analysis of EZKP utility).

PROPOSITION 2: *Suppose $h : [a, b] \to \mathbb{R}$ is differentiable,[21] concave, and satisfies $h' > 0$. Then for any measurable function $V : D \to [a, b]$ and $m \in \triangle(D)$,*

$$h^{-1}\big(\mathbb{E}_m[h(V)]\big) = \max_{\gamma \in [a,b]} \mathbb{E}_m\big[\phi(V|\gamma)\big],$$

where $\phi(\cdot|\gamma)$ is defined for $x, \gamma \in [a, b]$ by

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

[19] See also Gollier and Muermann (2010) for a related utility representation.

[20] Moreover, we show in Proposition S.1 of the Supplemental Material (Sarver (2018)) that the certainty equivalent defined by Equations (8) and (9) is a special 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 of the measure μ, and $g(\alpha) = (1 + \theta)\alpha$ for $\alpha \leq 1/2$ and $g(\alpha) = (1 - \theta)\alpha + \theta$ for $\alpha \geq 1/2$.

[21] Differentiability is only assumed for expositional simplicity. If h is not differentiable at a point γ, then $h'(\gamma)$ in Proposition 2 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.

Applying this formula to the special case of exponential functions $h_\theta(x) = -\exp(-\theta x)$, the following corollary demonstrates the mapping between ORA utility and generalized EZKP utility with optimization over θ.[22]

COROLLARY 1: *For* $\gamma \in \mathbb{R}$ *and* $\theta \in \Theta \subset \mathbb{R}_{++}$, *define* $\phi(x|\gamma, \theta)$ *as in Equation* (7), *and suppose* $\tau : \Theta \to \mathbb{R}$ *satisfies* $\inf_{\theta \in \Theta} \tau(\theta) = 0$. *Then, any ORA representation* (V, u, Φ, β) *with* Φ *defined as in Equation* (10) *can be equivalently expressed as*

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

The connection formalized in Corollary 1 makes it easy to compare our model to the benchmark of EZKP utility (the case where $\Theta = \{\theta\}$), which will be particularly useful in some of the applications in Section 4.

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.[23]

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 $\phi : [a, b] \to \mathbb{R}$ (where $a = \min V$ and $b = \max V$) such that, for any $m \in \triangle(D)$,

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

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

[23]Another 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 Supplemental Material (Sarver (2018)), we apply a result from Cerreia-Vioglio, Maccheroni, and Marinacci (2017) to characterize a set of transformations Φ that is minimal in the sense of admitting the smallest possible set of expected-utility *preferences*.

The ORA representation is *maximal* if $\Phi = \Phi^*$.[24]

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* $f : \Phi_1^* \to \Phi_2^*$ *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.[25]

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

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

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* $f : \Phi_1^* \to \Phi_2^*$ *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.

[24]An 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).

[25]It is immediate from the construction in Epstein and Zin (1989) that $C^{\mathbb{N}}$ can be embedded as a subset of D. Formally, for every $(c_0, c_1, c_2, \dots) \in C^{\mathbb{N}}$, there exists a sequence $\{m_i\}_{i \in \mathbb{N}}$ such that $m_i = \delta_{(c_i, m_{i+1})}$. We abuse notation slightly and denote (c_0, m_1) by $(c_0, c_1, c_2, \dots)$.

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 of 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 2: *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.*[26]
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). *If* $\theta_1 \geq \theta_2$, *then* $\succsim_1$ *is more risk averse than* $\succsim_2$ (*and conversely provided* $\succsim_1 \neq \succsim_2$).
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* $\succsim_2$.

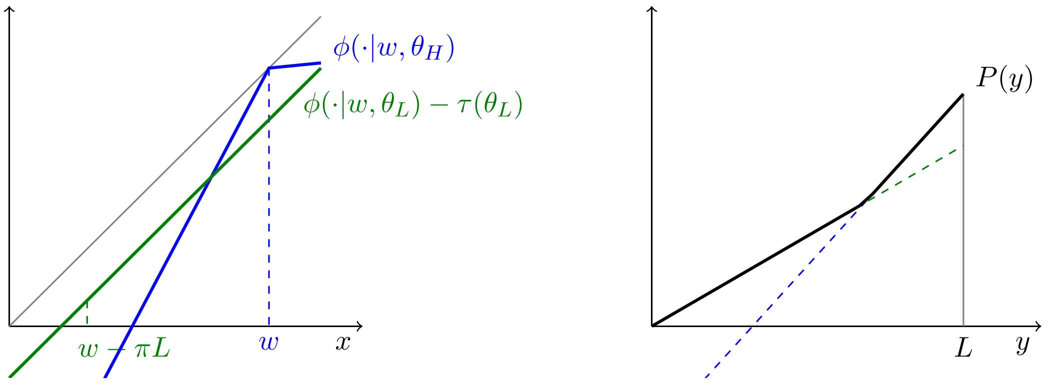
4. APPLICATIONS AND EXPERIMENTAL EVIDENCE

This section contains several applications of our model and some discussion of links to experimental evidence. Recall that one significant new feature of mixture-averse preferences is that they can allow aversion to marginal increases in risk to decrease with exposure. Some of the implications of this feature of the model are explored in these applications. In Section 4.1, we show that mixture-averse preferences can permit the marginal willingness to pay for additional insurance coverage to actually increase with the existing level of coverage, which may help to explain the high prices some consumers pay to decrease their insurance deductibles. In Section 4.2, we show that our model can generate endogenous heterogeneity in equilibrium stock market participation, with one segment of the population holding significantly greater risk. In Section 4.3, we use mixture-averse preferences to illustrate how the Rabin paradox can be resolved without relying on first-order risk aversion, and we find that moderate levels of background risk can actually improve the fit of the model. In Section 4.4, we show that our model provides a simple explanation for the experimental evidence that most violations of expected-utility theory occur near the boundary of the probability simplex.

4.1. *Demand for Insurance*

In this section, we consider how an individual's marginal willingness to pay for additional insurance coverage changes with her existing level of coverage. When an insurance policy 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.

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



(a) Transformation functions when $\gamma = w$ **(b)** Reservation price $P(y)$

FIGURE 3.—Reservation price for insurance from Example 3.

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 $\Theta = \{\theta_L, \theta_H\}$, where $0 = \theta_L < \theta_H < 1$ and $0 = \tau(\theta_H) < \tau(\theta_L)$.[27] 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

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

where the second equality follows because w is the median outcome of these lotteries and hence $\gamma = w$ is optimal as noted previously in Section 3.2. Figure 3 illustrates the function $P(y)$ in the case where the parameters are such that θ_L is optimal for $y = 0$, that is, when $\theta_H \pi L \geq \tau(\theta_L)$.[28]

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.

[27]For ease of exposition, this example focuses on the certainty equivalent applied in a static decision problem. It could equivalently be formulated using risk about future wealth within a (dynamic) ORA representation with $u(c) = c$.

[28]For ease of exposition, we have assumed that the only nonlinearity is the kink in the transformation $\phi(x|\gamma,\theta)$ 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 $u(c)$ rather than the linearity implicitly assumed in this stylized example (cf. Section 4.3). The implication would be that marginal willingness to pay for insurance is strictly decreasing in the level of coverage, except at those levels at which the optimal θ is changing, where there will be a discrete increase in marginal willingness to pay.

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 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 Supplemental Material (Sarver (2018)), marginal willingness to pay for additional insurance coverage must be nonincreasing in the level of coverage for any risk preference that satisfies preference for diversification (quasiconcavity in random variables). Moreover, using results from Dekel (1989) and Chew, Karni, and Safra (1987), we also discuss how most non-expected-utility preferences used in the literature cannot relax preference for diversification without also violating monotonicity with respect to second-order stochastic dominance.[29] Thus, for such preferences, risk aversion implies a decreasing marginal willingness to pay for insurance coverage. It seems overly restrictive that these two properties should be so tightly linked, especially in the context of insurance decisions where risk aversion plays such a prominent role. In contrast, since all of the transformation functions in Example 3 are concave, the preference respects second-order stochastic dominance (see Proposition 1).

4.2. *Heterogeneous Stock Market Participation*

In a companion paper Sarver (2017), we used mixture-averse preferences to study asset allocation decisions in the dynamic stochastic general equilibrium of a calibrated economy. We showed that the special case of the optimal risk attitude representation described in Corollary 1 can be used to provide a partial resolution to the stock market participation puzzle. This puzzle concerns the difficulty of explaining with standard models the non-participation or limited participation in equity markets by many households, including a nontrivial fraction of wealthy households.[30] In this section, we provide a brief illustration of how our model can partially address this puzzle by generating endogenous heterogeneity in equilibrium risk exposure, even when consumers have identical preferences. This mechanism enables the model to produce large cross-sectional variation in equity holdings without assuming large (or even any) cross-sectional variation in risk preferences.[31]

To illustrate as simply as possible how heterogeneity in risk exposure can arise endogenously in our model, we will restrict attention to a static environment with consumption in only a single period. However, we should emphasize that our illustration is representative of risk preferences over uncertain future wealth in the recursive infinite-horizon model used in Sarver (2017). Figure 4 illustrates the indifference curves of the risk preferences from Corollary 1 (equivalently, Equations (7) and (11)) applied to random consumption allocations $c = (c(z^l), c(z^h))$ when there are two states, z^l and z^h, that occur with equal probability. In this figure, there are two feasible values of the parameter θ, with $\theta_L < \theta_H$ and $0 = \tau(\theta_H) < \tau(\theta_L)$.

Assume there is a continuum of identical consumers in the economy, each endowed with the allocation e. For preferences that are quasiconcave in random variables (i.e.,

[29]It should be noted that Dekel (1989, Proposition 1) constructed an example (outside the general classes of preferences illustrated in Figure 1) that exhibits risk aversion but violates preferences for diversification.

[30]See, for example, Mankiw and Zeldes (1991), Haliassos and Bertaut (1995), and Heaton and Lucas (2000). For recent surveys, see Campbell (2006) and Guiso and Sodini (2013).

[31]For a complete analysis, as well as an extensive discussion of alternative approaches to generating heterogeneous stock market participation that have been explored in the literature, the reader is referred to Sarver (2017).

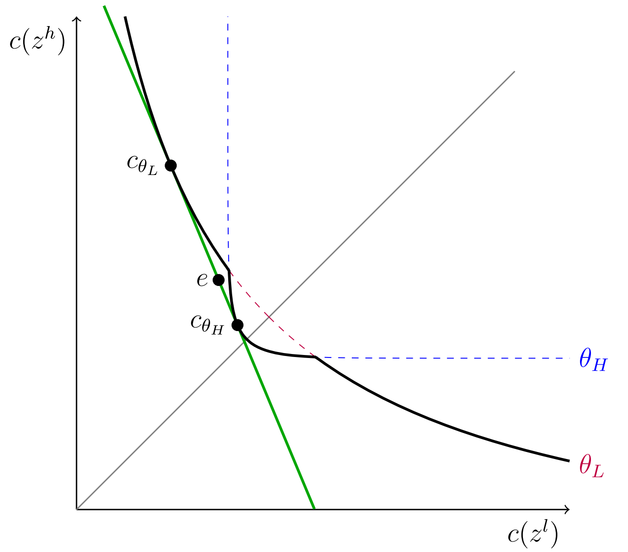


FIGURE 4.—Risk preferences and endogenous heterogeneity in general equilibrium. Type θ_L consumers choose allocation c_{θ_L} and type θ_H consumers choose allocation c_{θ_H}. The fraction of consumers selecting each type is determined by the market-clearing condition.

those that satisfy preference for diversification), homogeneity of preferences and endowments would imply the existence of an equilibrium in which each agent consumes her endowment. In contrast, although the preferences illustrated in Figure 4 are risk averse (respect SOSD), they do not satisfy preference for diversification, and it is clear from the figure that there is no equilibrium in which $c = e$ for each consumer. Instead, in the equilibrium of our model, some consumers select type θ_L and choose the allocation c_{θ_L} while other consumers select type θ_H and choose the (much lower risk) allocation c_{θ_H}.[32] 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.

4.3. *Rabin Paradox With Background Risk*

One difficulty of expected utility is that it requires either unrealistically low levels of risk aversion for small risks or else unrealistically high levels of risk aversion for large risks. The simplest illustration of this problem comes in the form of the Rabin paradox (Rabin (2000)): If a risk-averse expected-utility maximizer rejects an even-odds gamble that could lose \$100 or gain \$101 (as is overwhelmingly the case empirically) at all wealth levels, then this individual would also reject any even-odds gamble that could lose \$10,000, regardless of the size of the potential gain.

It is well known that non-expected-utility preferences can resolve the Rabin paradox. For instance, models that exhibit first-order risk aversion due to a kink in the local utility function near certainty (e.g., the disappointment aversion preferences of Gul (1991) or

[32]Since there is a continuum of consumers, the theorem of Aumann (1966) ensures the existence of an equilibrium in our economy.

the kinked transformation example of the ORA representation from Equation (8)) can be highly risk averse for small gambles without being overly risk averse for large gambles. Nonetheless, these models are still subject to an extended version of the Rabin paradox: Safra and Segal (2008) (see also Barberis, Huang, and Thaler (2006)) showed that in the presence of sufficiently large background risk, non-expected-utility preferences also cannot generate reasonable levels of risk aversion for small gambles without imposing an absurdly high degree of risk aversion for larger gambles.

The goal of this section is not to circumvent these negative results, as the Safra and Segal (2008) result for background risk applies also to the ORA representation. Rather, we have the more modest goal of using our model to make two useful observations about the Rabin paradox: First, we will illustrate that first-order risk aversion is not required to obtain descriptively accurate attitudes toward both small and large gambles. Second, and more significant, we observe that introducing a *moderate* (rather than large) amount of independent background risk can actually *improve* the fit of our model. Thus, while a large amount of background risk is known to have a negative impact on the ability of essentially any risk preference to generate reasonable attitudes toward both small and large gambles, we will show that the impact of more moderate levels of background risk will depend finely on the specifics of the model.

Table I describes the gains needed for an individual to be indifferent between accepting and rejecting a 50–50 gamble for various possible loss values when initial wealth is \$300,000. Our table includes five utility specifications: two cases of (EZKP) expected utility, two cases of the ORA representation that are new to this paper, and one case of rank-dependent utility (RDU). For each case, we will assume that one-time (atemporal) gambles are evaluated by applying the specified certainty equivalent to a value function for wealth that takes the form $\mathcal{V}(w) = \log(w)$.[33] The exact parameterizations used in these specifications are as follows:

1. The first four specifications use the certainty equivalent defined by Equations (7) and (11). Recall that by Proposition 2 (and Corollary 1), these preferences can be equivalently expressed as

$$\sup_{\theta\in\Theta}\left\{-\frac{1}{\theta}\log\bigl(\mathbb{E}\bigl[\exp\bigl(-\theta\mathcal{V}(w)\bigr)\bigr]\bigr)-\tau(\theta)\right\}=\sup_{\theta\in\Theta}\bigl\{\log\bigl(\mathbb{E}\bigl[w^{-\theta}\bigr]^{-\frac{1}{\theta}}\bigr)-\tau(\theta)\bigr\}.$$

Specifications EZKP1 and EZKP2 assume $\Theta = \{\theta_H\}$, and gambles are therefore evaluated using expected-utility preferences with a coefficient of relative risk aversion of $\theta_H + 1$. Specifications ORA1 and ORA2 assume $\Theta = \{\theta_L, \theta_H\}$, where $\theta_L < \theta_H$ and $0 = \tau(\theta_H) < \tau(\theta_L) \equiv \tau_L$.

2. Specification RDU1 instead uses the certainty equivalent defined by Equations (8) and (9) for a parameter θ_H. As we show in Proposition S.1 in the Supplemental Material (Sarver (2018)), these preferences can be equivalently expressed as the following special case of rank-dependent utility:

$$\int \log(w)\, d(g\circ F)(w),$$

[33]This assumption is roughly equivalent to assuming log utility for consumption within each period. For example, for ORA representations with $u(c) = \log(c)$ and shift-invariant certainty equivalents like those specified in this section, it is standard that the value function for wealth takes the log form within any consumption-saving-investment problem where the stochastic process driving the economy is i.i.d. See Epstein and Zin (1989, Section 5) for general results along these lines, and see Sarver (2017) for specific results for these cases of the ORA representation.

TABLE I

CALIBRATION RESULTS: EU, ORA, AND RDU MODELS[a]

	Risk-Preference Model				
	EZKP1	EZKP2	ORA1	ORA2	RDU1
θ_H	3.00	25.00	25.00	145.00	0.20
θ_L	–	–	3.00	3.00	–
τ_L	–	–	0.02	0.02	–
Panel A: Binary 50–50 Gambles					
Loss	Gain that leads to indifference for initial wealth $300,000				
$100	100.13	100.87	100.87	105.12	150.06
$400	402.14	414.37	414.37	497.15	601.00
$1,000	1,013.51	1,094.95	1,094.95	2,023.89	1,506.27
$5,000	5,357.20	8,995.81	8,995.81	18,991.43	7,659.35
$10,000	11,539.60	∞	26,396.79	26,396.79	15,650.25
$20,000	27,302.60	∞	45,692.48	45,692.48	32,710.18
Panel B: Binary 50–50 Gambles With Background Risk					
Loss	Gain that leads to indifference for random initial wealth $300,000 ± $7,000				
$100	100.13	100.89	100.89	105.25	100.03
$400	402.15	414.56	414.56	500.07	400.54
$1,000	1,013.55	1,096.33	1,096.33	2,081.18	1,003.36
$5,000	5,358.24	9,098.40	9,098.40	6,814.09	5,085.24
$10,000	11,544.43	∞	22,003.17	13,123.15	11,990.94
$20,000	27,329.83	∞	40,223.31	29,257.73	28,855.27

[a]An individual is indifferent between accepting and rejecting a 50–50 gamble with the loss listed in the left column and the gain listed in the table. Results include two specifications of EZKP expected utility, two additional specifications of the ORA representation, and one specification of rank-dependent utility.

where F is the cumulative distribution of wealth, and g is defined by

$$g(\alpha) = \begin{cases} (1+\theta_H)\alpha & \text{for } \alpha \leq 1/2, \\ (1-\theta_H)\alpha + \theta_H & \text{for } \alpha > 1/2. \end{cases}$$

Panel A lists the compensating gains when initial wealth is deterministic. Specifications EZKP1 and EZKP2 provide a parametric illustration of the Rabin paradox. Both specifications impose too little risk aversion for small-stake gambles. The second improves slightly over the first for small gambles, but is excessively risk averse for large gambles: The individual will reject any gamble in which she would lose over $10,000 (3.3% of wealth) with even odds, regardless of the size of the possible gain that could be won.[34] Specification ORA1 generates identical risk attitudes to EZKP2 for small risks, but generates much more reasonable attitudes toward larger gambles. Specification ORA2 further increases the parameter θ_H to 145, which implies the individual would reject a gamble that

[34]Similar 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.

could lose \$100 or gain \$105 with even odds. Note, however, that it maintains the same attitudes toward large gambles as ORA1. Neither of these specifications exhibits first-order risk aversion, yet both generate significant risk aversion for gambles on the scale of \$100. Finally, specification RDU1, which exhibits first-order risk aversion, is perhaps the best suited for generating high risk aversion for small gambles and moderate risk aversion for larger gambles.

Panel B lists the compensating gains in the presence of independent background risk. Specifically, the individual may gain or lose \$7,000 with equal odds, independent of whether she accepts or rejects the gambles in the table (and independently distributed). It is known that for CRRA expected-utility preferences, the introduction of independent background risk only serves to increase effective risk aversion (see Gollier and Pratt (1996)). Therefore, as expected, the compensating gains for specifications EZKP1 and EZKP2 are slightly larger than in Panel A, although the change is relatively minor.

For specifications ORA1 and ORA2, the effect of background risk is similar to what was observed for expected utility for small gambles where θ_H is optimal. Thus the individual will continue to be very averse to small gambles. Interestingly, the effect reverses for large gambles and the compensating gain actually decreases. To understand this result, consider a sample entry from Panel A: Absent background risk, an individual with utility given by ORA2 is indifferent between accepting or rejecting an even-odds gamble that could lose \$20,000 or gain \$45,692.48. However, since introducing background risk has a more significant negative impact on baseline utility absent the gamble (where θ_H is optimal) than on utility with the gamble (where θ_L is optimal), the gamble becomes relatively more attractive and the required compensating gain in Panel B is thereby reduced to \$29,257.73. Thus the introduction of moderate background risk improves the fit of ORA1 and ORA2 by *increasing* the gain required for the individual to accept a gamble with a small loss (albeit only slightly) while *decreasing* the gains required to accept gambles with large losses.[35]

In contrast, introducing background risk significantly diminishes the performance of specification RDU1. The compensating gains for the smallest losses of \$100 and \$400 decrease dramatically, to within \$1 of the respective losses. For the intuition behind this result, recall from Section 3.2 that an optimal value of γ for this kinked transformation function is at the median outcome, that is, \$300,000. Thus, regardless of the value of θ_H, the only concavity of the local utility function near \$293,000 and \$307,000 comes from the log value function. The implication is that in the presence of this background risk, small gambles of magnitude less than \$7,000 are evaluated as if the individual has expected-utility preferences with a CRRA of 1.

Our observation about the potential fragility of explanations of the Rabin paradox based solely on first-order risk aversion is not new. For example, Barberis, Huang, and Thaler (2006) made a similar observation regarding the negative impact of background risk on the ability of disappointment-averse preferences to address this puzzle.[36] What is novel is our proposed resolution of the issue. Some papers have suggested narrow framing of risks, so that background risk is evaluated separately from the risk in the prospective gamble (see Freeman (2015) for recent results in this vein). Some degree of narrow

[35]The scale of background risk is, of course, important for this conclusion. If background risk becomes so large that θ_H is optimal even when these gambles are declined, then the compensating gains will be identical to those of specification EZKP1, that is, roughly risk neutral for small gambles.

[36]It should be noted that Barberis, Huang, and Thaler (2006, pp. 1076–1077), imposed background risk of significantly greater magnitude than we do in this section. They assumed background risk with a standard deviation of roughly 17% of wealth, whereas our results invoke background risk with a standard deviation of roughly 2.3% of wealth; the conclusions they drew were commensurately more extreme.

framing arises naturally in recursive non-expected-utility models such as the ORA representation, since a gamble is evaluated together with other risk resolving in that same time period but separately from risk resolving in different periods. In particular, if we take the time period in our model to be one year, then the relevant simultaneous risk is the variation of wealth *within each year*. We chose background risk with a standard deviation of roughly 2.3% of wealth to match estimates that place the standard deviation of annual aggregate U.S. consumption growth between 1% and 3.6%, depending on the range of years considered (Mehra and Prescott (1985), Campbell (1999)).[37]

Importantly, it has been argued that such narrow framing of risk across time is not sufficient to resolve the Rabin paradox, unless the bracketing is so narrow that gambles and background risk resolve in different periods. For example, Barberis, Huang, and Thaler (2006) suggested that narrow framing by year is not sufficient to resolve the puzzle, and they proposed that, in addition, individuals must frame the Rabin gambles separately from other risks to wealth. In contrast, Table I shows that specifications ORA1 and ORA2 perform well without need for any additional assumptions of narrow framing of different types of risk within each year.

4.4. *Expected Utility Away From Certainty*

The special cases of our model that were used in the applications in Sections 4.1, 4.2, and 4.3 are consistent with experimental evidence that finds that the majority of the violations of the expected-utility axioms occur near the boundary of the probability simplex (see Harless and Camerer (1994) for a detailed discussion and an aggregation of numerous earlier studies). For example, taking $\Theta = \{\theta_L, \theta_H\}$ for $\theta_L < \theta_H$ and $0 = \tau(\theta_H) < \tau(\theta_L)$ in the extension of expected utility described in Corollary 1 generates this pattern: θ_L will be optimal for lotteries involving significant risk, whereas the individual will switch to θ_H for lotteries near the vertices of the simplex. Violations of the independence axiom can occur because indifference curves have different slopes in the θ_L and θ_H optimality regions, but independence holds within each region when considered in isolation.

Similarly, the certainty equivalent defined by Equations (8) and (11) with the parameters used in Section 4.1 is also consistent with this evidence. For lotteries involving significant risk, $\theta_L = 0$ will be optimal, which implies expected-utility preferences in this region of the probability simplex. For lotteries with little risk, $\theta_H > 0$ will be optimal, and in this region preferences instead conform to rank-dependent utility (see Proposition S.1 in the Supplemental Material (Sarver (2018))) and exhibit first-order risk aversion.

These special cases of our representation are similar in spirit to, and share 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 preferences considered in this paper are continuous and respect first-order stochastic dominance. The optimal risk attitude representation is therefore able to accommodate much of the experimental evidence that motivated these models, while at the same time maintaining convenient properties that make it amenable to standard techniques from macroeconomics and finance.

[37]Recall that for homothetic preferences, consumption is a fixed proportion of wealth.

APPENDIX 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* $m \in \triangle(D)$ *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* $c, \hat{c}, \hat{c}' \in C$ *and* $\hat{m}, \hat{m}' \in \triangle(D)$,

$$(\hat{c}, \hat{m}) \succsim \left(\hat{c}', \hat{m}'\right) \iff (c, \delta_{(\hat{c},\hat{m})}) \succsim (c, \delta_{(\hat{c}',\hat{m}')}).$$

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* $c, c', \hat{c}, \hat{c}' \in C$ *and* $\hat{m}, \hat{m}' \in \triangle(D)$,
1. $(c, \delta_{(\hat{c},\hat{m})}) \succsim (c', \delta_{(\hat{c}',\hat{m})})$ *if and only if* $(c, \delta_{(\hat{c},\hat{m}')}) \succsim (c', \delta_{(\hat{c}',\hat{m}')})$.
2. $(c, \delta_{(\hat{c},\hat{m})}) \succsim (c', \delta_{(\hat{c},\hat{m}')})$ *if and only if* $(c, \delta_{(\hat{c}',\hat{m})}) \succsim (c', \delta_{(\hat{c}',\hat{m}')})$.

Condition 1 in Axiom 6 says that the comparison of c today and $\hat{c}$ tomorrow versus c' today and $\hat{c}'$ 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 $\hat{c}'$). 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 $\triangle(D)$. Recall that in the case of monetary gambles, FOSD roughly corresponds to increasing the probability of better monetary outcomes. The same is true 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 $(c, \delta_{(\hat{c}',\hat{m}')}) \succsim (c, \delta_{(\hat{c},\hat{m})})$.[38]

[38]When 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): $(c, \delta_{(\hat{c},\hat{m})}) \succsim (c, \delta_{(\hat{c}',\hat{m}')}) \iff (c, \alpha\delta_{(\hat{c},\hat{m})} + (1-\alpha)m) \succsim (c, \alpha\delta_{(\hat{c}',\hat{m}')} + (1-\alpha)m)$.

AXIOM 7—FOSD: *For any* $c \in C$ *and* $m, m' \in \triangle(D)$, *if for all* $(\hat{c}, \hat{m}) \in D$,

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

then $(c, m) \succsim (c, m')$.[39]

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

PROPOSITION 3: *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 $V : D \to \mathbb{R}$ that represents $\succsim$, a continuous and nonconstant function $u : C \to \mathbb{R}$, a continuous and strictly increasing function $h : [a, b] \to \mathbb{R}$ (where $a = \min V$ and $b = \max V$), and a scalar $\beta \in (0, 1)$ such that, for all $(c, m) \in D$,

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

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. Proposition 2 in Section 3.2 showed that this case of EZKP utility can be expressed as an ORA representation. In this section, we provide an axiomatic characterization of EZKP utility and further explore the connection with mixture aversion.

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.

AXIOM 8—Independence: *For any* $c \in C$, $m, m', m'' \in \triangle(D)$, *and* $\alpha \in (0, 1)$,

$$(c, m) \succ \big(c, m'\big) \quad \Longrightarrow \quad \big(c, \alpha m + (1 - \alpha)m''\big) \succ \big(c, \alpha m' + (1 - \alpha)m''\big).$$

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.

[39]Implicit 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.

PROPOSITION 4: *Suppose $\succsim$ has an EZ representation.*[40] *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 demonstrated that any EZKP representation with concave h could be expressed as an ORA representation. The second part of Proposition 4 establishes the same connection using the axioms and also shows the converse: These are the only cases of EZKP utility that satisfy mixture aversion.

APPENDIX B: PROOFS

B.1. *Proofs of Proposition 3 and Theorem 1*

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

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

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

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

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

For sufficiency, the first step is to 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 $f : C \to \mathbb{R}$, $g : C \to \mathbb{R}$, and $h : \triangle(D) \to \mathbb{R}$ such that

$$(c, \delta_{(\hat{c},\hat{m})}) \succsim (c', \delta_{(\hat{c}',\hat{m}')}) \iff f(c) + g(\hat{c}) + h(\hat{m}) \geq f(c') + g(\hat{c}') + h(\hat{m}'). \tag{B.1}$$

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

$$\begin{aligned} (\hat{c}, \hat{m}) \succsim (\hat{c}', \hat{m}') &\iff (c, \delta_{(\hat{c},\hat{m})}) \succsim (c, \delta_{(\hat{c}',\hat{m}')}) \\ &\iff g(\hat{c}) + h(\hat{m}) \geq g(\hat{c}') + h(\hat{m}'), \end{aligned} \tag{B.2}$$

and hence g and h give an additive representation for $\succsim$. In particular, the combination of Equations (B.1) and (B.2) implies

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

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

[40]Equivalently, suppose $\succsim$ satisfies Axioms 2–7 in Appendix A.1.

that

$$g(c) = \beta f(c) + \alpha_1, \quad \forall c \in C,$$
$$h(\delta_{(\hat{c},\hat{m})}) = \beta\big[g(\hat{c}) + h(\hat{m})\big] + \alpha_2, \quad \forall (\hat{c},\hat{m}) \in D. \tag{B.3}$$

Define $u_1 : C \to \mathbb{R}$ and $u_2 : \triangle(D) \to \mathbb{R}$ by $u_1(c) = g(c) + \frac{\alpha_2}{\beta}$ and $u_2(m) = h(m)$. Then, claims 1 and 2 follow directly from Equations (B.2) and (B.3).

It remains only to show that $\beta < 1$. Following a similar approach to Gul and Pesendorfer (2004), this can be established using continuity. By nontriviality, there exist $c^*, c_* \in C$ such that $u_1(c^*) > u_1(c_*)$. Fix any $m \in \triangle(D)$ and, with slight abuse of notation, define sequences $\{d_n\}$ and $\{d'_n\}$ in D as follows:[41]

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

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

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

Since this difference converges to a real number, it must be that $\beta < 1$. *Q.E.D.*

LEMMA 2: *Suppose $\succsim$ is represented by $V(c,m) = u_1(c) + u_2(m)$, where $u_1 : C \to \mathbb{R}$ and $u_2 : \triangle(D) \to \mathbb{R}$ are continuous,*[42] *and $\succsim$ satisfies stationarity. Then, $\succsim$ satisfies FOSD (Axiom 7) if and only if, for any $m, m' \in \triangle(D)$,*

$$m \circ V^{-1}\big([x,\infty)\big) \geq m' \circ V^{-1}\big([x,\infty)\big), \quad \forall x \in V(D) \quad \Longrightarrow \quad u_2(m) \geq u_2\big(m'\big). \tag{B.4}$$

PROOF: To see that FOSD implies Equation (B.4), consider any two measures $m, m' \in \triangle(D)$ such that

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

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,

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

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

[42] Continuity is not necessary for this result; measurability of u_1 and u_2 is 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 (B.4) implies the FOSD axiom is similar. *Q.E.D.*

LEMMA 3: *Suppose* $\succsim$ *is represented by* $V(c,m) = u_1(c) + u_2(m)$, *where* $u_1 : C \to \mathbb{R}$ *and* $u_2 : \triangle(D) \to \mathbb{R}$ *are nonconstant and continuous. Then,* $\succsim$ *satisfies 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,

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

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 $m, m' \in \triangle(D)$ 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

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

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,

$$u_2(m') - u_2\left(\frac{1}{2}m + \frac{1}{2}m'\right) \geq u_1(c') - u_1(c) = u_2\left(\frac{1}{2}m + \frac{1}{2}m'\right) - u_2(m)$$
$$\implies \quad \frac{1}{2}u_2(m) + \frac{1}{2}u_2(m') \geq u_2\left(\frac{1}{2}m + \frac{1}{2}m'\right).$$

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

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

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. *Q.E.D.*

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

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

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

Fix any $\mu \in \triangle(V(D))$. Let $W(\mu) = \frac{1}{\beta}u_2(m)$ for any $m \in \triangle(D)$ such that $\mu = m \circ V^{-1}$. There exists at least one such m by the preceding arguments. In addition, if $\mu = m \circ V^{-1} = m' \circ V^{-1}$ for $m, m' \in \triangle(D)$, then $u_2(m) = u_2(m')$ by Equation (B.4). Thus W is well defined, and by construction, $u_2(m) = \beta W(m \circ V^{-1})$ for all $m \in \triangle(D)$.

Proof of 1: By condition 2 in Lemma 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 $\triangle(V(D))$ that converges to some $\mu \in \triangle(V(D))$. It suffices to show that there exists a subsequence $\{\mu_{n_k}\}$ such that $W(\mu_{n_k}) \to W(\mu)$.[43] For each n, take any $m_n \in \triangle(D)$ such that $\mu_n = m_n \circ V^{-1}$. Since $\triangle(D)$ is compact and metrizable, there is a subsequence $\{m_{n_k}\}$ converging to some $m \in \triangle(D)$. By the continuity of V,

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

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,

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

Therefore, W is weak* continuous. To see that W is monotone with respect to FOSD, suppose $\mu, \eta \in \triangle(V(D))$ satisfy $\mu([x, b]) \geq \eta([x, b])$ for all $x \in [a, b] \equiv V(D)$. Take any $m, m' \in \triangle(D)$ such that $\mu = m \circ V^{-1}$ and $\eta = m' \circ V^{-1}$. Then, Equation (B.4) implies $u_2(m) \geq u_2(m')$, and hence $W(\mu) \geq W(\eta)$.

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

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

[43]If W is not continuous at a point μ, there exist $\varepsilon > 0$ 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.

and hence

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

establishing the convexity of W. *Q.E.D.*

PROOF OF PROPOSITION 3: The necessity of the axioms is straightforward. To establish sufficiency, suppose $\succsim$ satisfies Axioms 2–7. By Lemmas 1, 2, and 4, there exist a continuous function $V : D \to \mathbb{R}$, a scalar $\beta \in (0,1)$, a continuous and nonconstant function $u : C \to \mathbb{R}$, and a weak* continuous function $W : \triangle(V(D)) \to \mathbb{R}$ such that

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

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 $a, b \in \mathbb{R}$. *Q.E.D.*

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 Proposition 1 to conclude there exists a collection Φ of continuous and nondecreasing functions $\phi : [a, b] \to \mathbb{R}$ such that

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

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

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

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 $\bar{x} \in [a, b]$. *Q.E.D.*

B.2. *Proof of Proposition 2*

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.3. *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 $g : V(D) \to C^{\mathbb{N}}$ 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 and $\sup_{\phi \in \Phi} \phi(x) = x$ for all $x \in V(D)$,

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

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

The existence of a function $g : V(D) \to C^{\mathbb{N}}$ 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). *Q.E.D.*

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

$$V_1(\mathbf{c}') \geq V_1(\mathbf{c}) \quad \Longrightarrow \quad V_2(\mathbf{c}') \geq V_2(\mathbf{c}). \tag{B.5}$$

Note that we have not established that the preferences over deterministic consumption streams are 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 (B.5) 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

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

and fix any consumption streams $\mathbf{c} = (c_0, c_1, c_2, \dots)$ and $\mathbf{c}' = (c_0', c_1', c_2', \dots)$ such that $c_0 = c$, $c_t = c^*$, $c_0' = c'$, $c_t' = c_*$, and $c_\tau = c_\tau'$ for $\tau \notin \{0, t\}$. Thus $V_1(\mathbf{c}') > V_1(\mathbf{c})$. By Equation (B.5), this implies $V_2(\mathbf{c}') \geq V_2(\mathbf{c})$ or, equivalently,

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

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 (B.5). 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 $u_2 = \alpha u_1 + \lambda(1-\beta_1)$. *Q.E.D.*

Fix any $(c, m) \in D$, and let $x = V_1(c, m)$. Define $g_1 : V_1(D) \to C^{\mathbb{N}}$ and $g_2 : V_2(D) \to C^{\mathbb{N}}$ 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,

$$\begin{aligned} V_2(c, m) &\geq V_2\big(g_1(x)\big) \\ &= \alpha V_1\big(g_1(x)\big) + \lambda \\ &= \alpha V_1(c, m) + \lambda. \end{aligned}$$

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

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

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

$$\sup_{\phi \in \Phi_2} \int_D \phi\big(V_2(\hat{c}, \hat{m})\big)\, dm(\hat{c}, \hat{m}) \geq \alpha \sup_{\phi \in \Phi_1} \int_D \phi\big(V_1(\hat{c}, \hat{m})\big)\, dm(\hat{c}, \hat{m}) + \lambda. \tag{B.6}$$

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

$$\int_D \phi_1\big(V_1(\hat{c}, \hat{m})\big)\, dm(\hat{c}, \hat{m}) \leq \sup_{\hat{\phi} \in \Phi_1} \int_D \hat{\phi}\big(V_1(\hat{c}, \hat{m})\big)\, dm(\hat{c}, \hat{m}), \quad \forall m \in \triangle(D). \tag{B.7}$$

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

$$V_2(\hat{c}, \hat{m}) = V_2\Big(g_2\big(V_2(\hat{c}, \hat{m})\big)\Big) = \alpha V_1\Big(g_2\big(V_2(\hat{c}, \hat{m})\big)\Big) + \lambda. \tag{B.8}$$

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

$$\begin{aligned} &\alpha \int_D \phi_1\left(\frac{V_2(\hat{c}, \hat{m}) - \lambda}{\alpha}\right) dm(\hat{c}, \hat{m}) + \lambda \\ &\quad = \alpha \int_D \phi_1\Big(V_1\Big(g_2\big(V_2(\hat{c}, \hat{m})\big)\Big)\Big)\, dm(\hat{c}, \hat{m}) + \lambda \quad \text{(by Equation (B.8))} \\ &\quad = \alpha \int_D \phi_1\big(V_1(\hat{c}, \hat{m})\big)\, d\tilde{m}(\hat{c}, \hat{m}) + \lambda \quad \text{(change of variables)} \\ &\quad \leq \alpha \sup_{\hat{\phi} \in \Phi_1} \int_D \hat{\phi}\big(V_1(\hat{c}, \hat{m})\big)\, d\tilde{m}(\hat{c}, \hat{m}) + \lambda \quad \text{(by Equation (B.7))} \end{aligned}$$

$$\leq \sup_{\hat{\phi}\in\Phi_2} \int_D \hat{\phi}\big(V_2(\hat{c},\hat{m})\big)\,d\tilde{m}(\hat{c},\hat{m}) \quad \big(\text{by Equation (B.6)}\big)$$

$$= \sup_{\hat{\phi}\in\Phi_2} \int_D \hat{\phi}\big(V_2\big(g_2\big(V_2(\hat{c},\hat{m})\big)\big)\big)\,dm(\hat{c},\hat{m}) \quad (\text{change of variables})$$

$$= \sup_{\hat{\phi}\in\Phi_2} \int_D \hat{\phi}\big(V_2(\hat{c},\hat{m})\big)\,dm(\hat{c},\hat{m}) \quad \big(\text{by Equation (B.8)}\big).$$

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

B.4. *Proof of Proposition 4*

Suppose $\succsim$ has an Epstein–Zin representation (V, u, W, β). Since $\succsim$ also satisfies the expected-utility axioms when restricted to lotteries $m \in \triangle(D)$, there exists a continuous function $f : D \to \mathbb{R}$ 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 $m \in \triangle(D)$. 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

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

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

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, that is, h is concave. Since Axiom 1 corresponds to the convexity of the certainty equivalent by Theorem 1, this completes the proof.

REFERENCES

ALIPRANTIS, C., AND K. BORDER (2006): *Infinite Dimensional Analysis* (Third Ed.). Berlin, Germany: Springer-Verlag. [1355,1374,1377]

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. [1348]

ANG, A., G. BEKAERT, AND J. LIU (2005): "Why Stocks May Disappoint," *Journal of Financial Economics*, 76, 471–508. [1351]

AUMANN, R. J. (1966): "Existence of Competitive Equilibria in Markets With a Continuum of Traders," *Econometrica*, 34, 1–17. [1364]

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. [1365,1367, 1368]

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. [1351]

BEN-TAL, A., AND M. TEBOULLE (1986): "Expected Utility, Penalty Functions, and Duality in Stochastic Nonlinear Programming," *Management Science*, 32, 1445–1466. [1357]

——— (2007): "An Old-New Concept of Convex Risk Measures: The Optimized Certainty Equivalent," *Mathematical Finance*, 17, 449–476. [1357-1359]

BRANDENBURGER, A., AND E. DEKEL (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189–198. [1352]

CAMPBELL, J. Y. (1999): "Asset Prices, Consumption, and the Business Cycle," in *Handbook of Macroeconomics*, Vol. 1, ed. by J. Taylor and M. Woodford. Amsterdam: North-Holland. [1368]

——— (2006): "Household Finance," *Journal of Finance*, 61, 1553–1604. [1363]

CERREIA-VIOGLIO, S. (2009): "Maxmin Expected Utility on a Subjective State Space: Convex Preferences Under Risk," Working paper. [1355]

CERREIA-VIOGLIO, S., D. DILLENBERGER, AND P. ORTOLEVA (2015): "Cautious Expected Utility and the Certainty Effect," *Econometrica*, 83, 693–728. [1351]

CERREIA-VIOGLIO, S., F. MACCHERONI, AND M. MARINACCI (2017): "Stochastic Dominance Analysis Without the Independence Axiom," *Management Science*, 63 (4), 1097–1109. [1356,1359]

CHETTY, R., AND A. SZEIDL (2007): "Consumption Commitments and Risk Preferences," *Quarterly Journal of Economics*, 122, 831–877. [1355]

——— (2016): "Consumption Commitments and Habit Formation," *Econometrica*, 84, 855–890. [1355]

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. [1351]

CHEW, S. H., AND L. G. EPSTEIN (1991): "Recursive Utility Under Uncertainty," in *Equilibrium Theory in Infinity Dimensional Spaces*, ed. by A. Khan and N. Yannelis. Springer Verlag. [1349,1352,1353,1360,1369,1370]

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. [1352,1363]

DEBREU, G. (1960): "Topological Methods in Cardinal Utility Theory," in *Mathematical Methods in the Social Sciences*, ed. by K. J. Arrow, S. Karlin, and P. Suppes. Stanford, California: Stanford University Press. [1353,1369,1371,1378]

DEKEL, E. (1986): "An Axiomatic Characterization of Preferences Under Uncertainty: Weakening the Independence Axiom," *Journal of Economic Theory*, 40, 304–318. [1351]

——— (1989): "Asset Demands Without the Independence Axiom," *Econometrica*, 57, 163–169. [1352,1363]

DIECIDUE, E., U. SCHMIDT, AND P. P. WAKKER (2004): "The Utility of Gambling Reconsidered," *Journal of Risk and Uncertainty*, 29, 241–259. [1368]

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. [1347,1349-1352,1354,1360,1365]

——— (1990): "'First-Order' Risk Aversion and the Equity Premium Puzzle," *Journal of Monetary Economics*, 26, 387–407. [1347,1351]

——— (2001): "The Independence Axiom and Asset Returns," *Journal of Empirical Finance*, 8, 537–572. [1351]

ERGIN, H., AND T. SARVER (2015): "Hidden Actions and Preferences for Timing of Resolution of Uncertainty," *Theoretical Economics*, 10, 489–541. [1355]

FISHBURN, P. C. (1970): *Utility Theory for Decision Making*. New York: John Wiley and Sons. [1371,1378]

FREEMAN, D. (2015): "Calibration Without Reduction for Non-Expected Utility," *Journal of Economic Theory*, 158, 21–32. [1367]

GABAIX, X., AND D. LAIBSON (2001): "The 6D Bias and the Equity-Premium Puzzle," *NBER Macroeconomics Annual*, 16, 257–312. [1355]

GOLLIER, C., AND A. MUERMANN (2010): "Optimal Choice and Beliefs With Ex Ante Savoring and Ex Post Disappointment," *Management Science*, 56, 1272–1284. [1358]

GOLLIER, C., AND J. W. PRATT (1996): "Risk Vulnerability and the Tempering Effect of Background Risk," *Econometrica*, 64, 1109–1123. [1367]

GRANT, S., A. KAJII, AND B. POLAK (2000): "Temporal Resolution of Uncertainty and Recursive Non-Expected Utility Models," *Econometrica*, 68, 425–434. [1351]

GROSSMAN, S., AND G. LAROQUE (1990): "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, 58, 25–51. [1355]

GUISO, L., AND P. SODINI (2013): "Household Finance: An Emerging Field," in *Handbook of the Economics of Finance*, Vol. 2, ed. by G. M. Constantinides, M. Harris, and R. M. Stulz. Amsterdam, The Netherlands: Elsevier. [1363]

GUL, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59, 667–686. [1351,1364]

GUL, F., AND W. PESENDORFER (2004): "Self-Control and the Theory of Consumption," *Econometrica*, 72, 119–158. [1352,1372]

HALIASSOS, M., AND C. C. BERTAUT (1995): "Why Do so few Hold Stocks?" *The Economic Journal*, 105, 1110–1129. [1363]

HARLESS, D., AND C. CAMERER (1994): "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62, 1251–1289. [1368]

HEATON, J., AND D. LUCAS (2000): "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk," *Journal of Finance*, 55, 1163–1198. [1363]

KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–292. [1348]

KREPS, D., AND E. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200. [1347,1350,1370]

——— (1979): "Temporal von Neumann-Morgenstern and Induced Preferences," *Journal of Economic Theory*, 20, 81–109. [1355]

LUCAS, R. E. (2003): "Macroeconomic Priorities," *American Economic Review*, 93, 1–14. [1366]

MACCHERONI, F. (2002): "Maxmin Under Risk," *Economic Theory*, 19, 823–831. [1355]

MACHINA, M. (1982): "'Expected Utility' Analysis Without the Independence Axiom," *Econometrica*, 50, 277–324. [1356]

——— (1984): "Temporal Risk and the Nature of Induced Preferences," *Journal of Economic Theory*, 33, 199–231. [1355,1360]

MANKIW, N. G., AND S. P. ZELDES (1991): "The Consumption of Stockholders and Nonstockholders," *Journal of Financial Economics*, 29, 97–112. [1363]

MEHRA, R., AND E. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161. [1366,1368]

MERTENS, J.-F., AND S. ZAMIR (1985): "Formulation of Bayesian Analysis for Games With Incomplete Information," *International Journal of Game Theory*, 14, 1–29. [1352]

MILGROM, P., AND C. SHANNON (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180. [1354]

NEILSON, W. S. (1992): "Some Mixed Results on Boundary Effects," *Economic Letters*, 39, 275–278. [1368]

QUAH, J. (2007): "The Comparative Statics of Constrained Optimization Problems," *Econometrica*, 75, 401–431. [1354]

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343. [1351]

RABIN, M. (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68, 1281–1292. [1364]

ROUTLEDGE, B., AND S. ZIN (2010): "Generalized Disappointment Aversion and Asset Prices," *Journal of Finance*, 65, 1303–1332. [1351]

SAFRA, Z., AND U. SEGAL (2008): "Calibration Results for Non-Expected Utility Theories," *Econometrica*, 76, 1143–1166. [1365]

SARVER, T. (2017): "Risk Attitude Optimization and Heterogeneous Stock Market Participation," Economic Research Initiatives at Duke (ERID) Working Paper No. 264, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3135964. [1350,1363,1365]

——— (2018): "Supplement to 'Dynamic Mixture-Averse Preferences'," *Econometrica Supplemental Material*, 86, https://doi.org/10.3982/ECTA12687. [1348,1350-1352,1354,1356,1358,1359,1363,1365,1368]

SCHMIDT, U. (1998): "A Measurement of the Certainty Effect," *Journal of Mathematical Psychology*, 42, 32–47. [1368]

SEGAL, U. (1988): "Probabilistic Insurance and Anticipated Utility," *Journal of Risk and Insurance*, 55, 287–297. [1348]

——— (1989): "Anticipated Utility: A Measure Representation Appraoch," *Annals of Operations Research*, 19, 359–373. [1351]

SEGAL, U., AND A. SPIVAK (1990): "First Order versus Second Order Risk Aversion," *Journal of Economic Theory*, 51, 111–125. [1362]

SYDNOR, J. (2010): "(Over)Insuring Modest Risks," *American Economic Journal: Applied Economics*, 2, 177–199. [1361]

WAKKER, P. (1994): "Separating Marginal Utility and Probabilistic Risk Aversion," *Theory and Decision*, 36, 1–44. [1351]

WAKKER, P., R. THALER, AND A. TVERSKY (1997): "Probabilistic Insurance," *Journal of Risk and Uncertainty*, 15, 7–28. [1348]

WEIL, P. (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421. [1350]

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

[1350]
YAARI, M. E. (1987): "The Dual Theory of Choice Under Risk," *Econometrica*, 55, 95–115. [1351,1358]

Co-editor Itzhak Gilboa handled this manuscript.

Manuscript received 25 July, 2014; final version accepted 25 February, 2018; available online 6 March, 2018.