Smooth Ambiguity Aversion Toward Small Risks and Continuous-Time Recursive Utility

Costis Skiadas*

first version: August 2008; this revision: July 2011.

Abstract

In a continuous-time setting with Brownian and Poissonian uncertainty, this paper formulates recursive utility under two smooth certainty equivalent (CE) types that have been proposed as representations of ambiguity aversion. For a smooth second-order expected utility CE based on the formulation of Klibanoff, Marinacci, and Mukerji (*Econometrica*, 2005), it is argued that the corresponding continuous-time recursive utility reduces to Kreps-Porteus utility, that is, recursive utility with an expected utility CE. For a smooth divergence CE, based on a formulation of Maccheroni, Marinacci, and Rustichini (*Econometrica*, 2006), the following conclusions are drawn. Under only Brownian uncertainty, the corresponding continuous-time recursive utility again reduces to the Kreps-Porteus case. Under Poissonian uncertainty, the same conclusion can be drawn if and only if the divergence CE is of the entropic type. A non-entropic divergence CE results in a new class of continuous-time smooth recursive utilities that price Brownian and Poissonian risks differently.

*Kellogg School of Management, Department of Finance, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. I am grateful for helpful discussions with Nabil Al-Najjar, Lars Hansen, Soohun Kim, Peter Klibanoff, Fabio Maccheroni, Mark Machina, Massimo Marinacci, Jianjun Miao, Sujoy Mukerji, Dimitris Papanikolaou and Viktor Todorov. I am especially thankful for the feedback I have received from Christian Hellwig and Mark Schroder. I'm responsible for any errors. The latest version of this article can be downloaded at www.kellogg.nwu.edu/faculty/skiadas/home.htm

1 Introduction

A significant class of economic models is based on the assumption that agents can act frequently on information that takes the form of a sequence of (conditionally) small risks.[1] The associated mathematical setup typically postulates an information structure that is generated by Brownian motion or other Lévy processes in continuous time. A continuous-time model can be viewed as an idealized representation of a sequence of discrete models that in the limit produce similar numerical results by virtue of the central limit theorem and its functional extensions. The main point is that the information structure allows quantitative approximations that imply significant simplifications. This paper discusses simplifications of this type as they relate to two classes of smooth utility functions that have been proposed as representations of ambiguity aversion.

A choice entails ambiguity (or Knightian uncertainty) if it is difficult to ascribe precise probabilities to the possible outcomes, a notion commonly attributed to Knight (1921). Ambiguity aversion in the sense of the well-known Ellsberg (1961) experiments has motivated a variety of proposed preference structures, perhaps best known of which is the non-smooth maxmin utility representation[2] of Gilboa and Schmeidler (1989). This paper focuses instead on two more recently proposed smooth utility functional forms: second-order expected utility and divergence preferences. The term second-order expected utility refers to the utility representation of Klibanoff, Marinacci, and Mukerji (2005), which also appears in Ergin and Gul (2009) and Nau (2006), building on earlier insights by Segal (1987, 1990). Divergence preferences were proposed by Maccheroni, Marinacci, and Rustichini (2006a), as an extension of the entropic variational preferences made prominent in macroeconomics by the work of Hansen, Sargent and their coauthors (see, for example, Hansen and Sargent (2001, 2007)).

We are interested in dynamic choice with information that is generated by random walks converging[3] to Brownian motion and/or Poisson processes as the frequency goes to infinity. We refer to an increment of such a random walk over a short time interval as a Brownian or Poissonian small risk, depending on the limiting behavior. Roughly speaking, a Brownian small risk involves a high probability of a small change, while a Poissonian small risk involves a low probability of a large change. These are natural special cases to start with, since it is well understood that Brownian motion and Poisson processes are the building blocks of all Lévy processes.[4] The decision maker only entertains priors that are consistent with the assumption of the Brownian/Poisson structure, but is uncertain about the drift of each Brownian motion and the arrival rate of each Poisson process.

[1] Important examples include Merton's contributions in Finance, collected in Merton (1990)), the approach of Holmstrom and Milgrom (1987) in the theory of optimal contracting, the analysis of Sannikov (2007) in the theory of repeated games with imperfect monitoring, as well as numerous other instances of micro and macro economic modeling. Textbook expositions involving continuous-time models include Dixit and Pindyck (1994), Duffie (2001), and Stokey (2009).

[2] Dynamic versions are given by Epstein and Schneider (2003) and Chen and Epstein (2002).

[3] The convergence results alluded to in this introduction can be found in Billingsley (1999).

[4] That is, all processes with stationary and independent increments that are continuous in probability (meaning there is no lump-sum resolution of uncertainty). Protter (2004) gives an excellent introduction to Lévy processes.

For example, a Brownian drift could represent a market risk premium associated with a source of risk whose volatility can be easily measured, while the Poisson arrival rate could correspond to the expected arrival time of a rare event (as in the literature on asset pricing with disasters represented by Rietz (1988), Barro (2006), Gabaix (2008) and others).

To define dynamic preferences, we take as our benchmark the utility of Kreps and Porteus (1978), that is, recursive utility with an expected-utility certainty equivalent (CE), and we modify the CE to correspond to second-order expected utility or divergence preferences. The resulting recursive utilities are essentially those formulated by Klibanoff, Marinacci, and Mukerji (2007) and Maccheroni, Marinacci, and Rustichini (2006b), with a broader specification of preferences over deterministic consumption, an aspect that is peripheral to this paper's main insights. In fact, the reader may choose to focus on the familiar parametric specification of Epstein and Zin (1991) and Weil (1989), replacing the expected-utility CE with the appropriate ambiguity-averse CE. We will argue that in the continuous-time limit with Brownian/Poisson information (which excludes lump-sum resolutions of uncertainty), smooth recursive utility with a second-order expected utility CE is indistinguishable from recursive utility with an expected-utility CE, as formulated in continuous time by Duffie and Epstein (1992). Moreover, the same is true of smooth recursive utility with a divergence CE given Brownian risks but not so given Poissonian risks, resulting in source-dependent uncertainty aversion that reflects the type of each source of risk.

Methodologically, we will proceed as follows. We will first consider a single node of a binomial tree, parameterized so that the tree converges to Brownian motion or a Poisson process as the time-period length h goes to zero. (In the continuous-time limit, h corresponds to a time infinitesimal dt.) A mathematically rigorous but elementary argument will establish that each CE of interest has the claimed approximation with an error term that is of smaller order of magnitude than h. The same approximations will also be verified heuristically, directly in continuous time, as implications of Ito's lemma. Given these CE approximations, we will apply the same type of heuristic argument Duffie and Epstein (1992) used to formulate a corresponding continuous-time recursive utility. Such an argument is compelling and has indeed been the standard in the literature of continuous-time recursive utility so far.[5] While important for the development of the subject, a mathematically rigorous convergence theory is, by necessity, focused on mathematical complexities and is not necessarily enlightening from an economic applications point of view.

The rest of this paper is organized in four sections and two appendices. Section 2 defines the CEs of interest. Section 3 presents the paper's essential ideas with CE approximations in a single-period model with Brownian or Poissonian uncertainty. Section 4 presents the continuous-

[5] Following this revision, I received a brand-new working paper by Kraft and Seifried (2011) that gives a first mathematically complete theory of convergence of Kreps-Porteus utility to its Duffie-Epstein Brownian limit. The convergence proof relies on the technical conditions assumed by Duffie and Epstein in their existence/uniqueness proof (which precludes the Epstein-Zin-Weil parameterization, unless log-consumption is bounded), and does not work with the type of CEs considered in this paper or with Poisson uncertainty. Nevertheless, the paper by Kraft and Seifried is an important first step toward a full convergence theory for recursive utility and illustrates the type of associated technical complexity alluded to in this Introduction.

time theory. Section 5 concludes the main part of the paper with a discussion of the results. Appendix A explains how Ito's lemma can be used to heuristically simplify the CEs studied in this paper. Appendix B contains mathematically rigorous proofs of the main approximation results of Section 3.

2 Certainty Equivalents

In a single-period setting, this section defines the certainty equivalents of interest: expected utility, second-order expected utility, and divergence. We postulate a finite state space Ω, with at least two elements, and a reference probability $P : 2^{\Omega} \to [0,1]$ that assigns a positive mass $P(\omega) = P(\{\omega\})$ to each state $\omega \in \Omega$. The expectation operator relative to P is denoted $\mathbb{E}$. Given any other probability Q on 2^{Ω}, the corresponding expectation operator is denoted $\mathbb{E}^{Q}$, and the *density* dQ/dP is defined as the random variable that takes the value $Q(\omega)/P(\omega)$ at state $\omega \in \Omega$. Therefore, $\mathbb{E}^{Q}x = \mathbb{E}[x dQ/dP]$ for every $x \in \mathbb{R}^{\Omega}$.

Fixed through the entire paper is a constant $\ell \in [-\infty, 1)$ serving as a lower bound on consumption, with typical values in applications being $\ell = -\infty$ or $\ell = 0$. A *certainty equivalent (CE)* is an increasing continuous function of the form $\upsilon : (\ell,\infty)^{\Omega} \to (\ell,\infty)$ with the property $\upsilon(\alpha\mathbf{1}) = \alpha$ for all $\alpha \in (\ell,\infty)$ (where $\mathbf{1}$ is the element of $(\ell,\infty)^{\Omega}$ identically equal to one).

In this paper, a *prior* is any probability Q that is equivalent to the reference probability P, which in the current setting means that Q assigns a positive mass to every state. This restriction on priors is imposed for expositional simplicity; it is in no way essential to the paper's arguments. The set of all priors is denoted Π.

2.1 Expected Utility CE

In this paper, a *von Neumann-Morgenstern (vNM) index* is any strictly increasing, continuous function of the form $u : (\ell,\infty) \to \infty$. Differentiability assumptions will be key in our approximations. For $n = 1, 2$, we use the notation

$$C^{n}_{\text{vNM}} = \text{ set of } n \text{ times continuously differentiable vNM indices.} \tag{1}$$

A pair of a prior Q and a vNM index u define the *expected utility certainty equivalent (EU CE)* $\upsilon = u^{-1}\mathbb{E}^{Q}u$, given more explicitly as

$$\upsilon(U) = u^{-1}\left(\mathbb{E}^{Q}u(U)\right), \quad U \in (\ell,\infty)^{\Omega}.$$

Recall that for any $Q \in \Pi$, two vNM indices u and $\tilde{u}$ define the same EU CE with prior Q if and only if they are related by a positive affine transformation (meaning that $\tilde{u} = \alpha u + \beta$ for some $\alpha \in (0,\infty)$ and $\beta \in \mathbb{R}$).

2.2 Second-Order Expected Utility CE

Consider an agent contemplating an EU CE with a given vNM index u, under each one of the possible priors $Q^1, \ldots, Q^S$. The agent is uncertain about which prior to use. In this sense, there is now a new source of uncertainty, represented by the new state-space $\{1, \ldots, S\}$, on which we postulate a probability represented by the weights $\pi^1, \ldots, \pi^S \in (0,1)$, where $\sum_s \pi^s = 1$. We refer to these weights as the agent's prior on priors. The *second-order EU CE* υ is defined in terms of another vNM index φ by

$$\upsilon(U) = \varphi^{-1}\left(\sum_{s=1}^{S} \varphi\left(u^{-1}\mathbb{E}^{Q^s} u(U)\right)\pi^s\right), \quad U \in (\ell, \infty)^{\Omega}. \tag{2}$$

The utility function $\varphi \circ \upsilon$ is of the type formulated by Klibanoff, Marinacci, and Mukerji (2005) (as can be easily seen by setting $\phi = \varphi \circ u^{-1}$).

The second-order EU CE (2) can be thought of as a representation of source-dependent risk aversion. Risk aversion toward payoffs that are contingent on the state $s \in \{1, \ldots, S\}$ is determined by φ. Conditionally on the state s, risk-aversion toward payoffs that are contingent on the state $\omega \in \Omega$ is determined by u. If $u = \varphi$, then $\upsilon = u^{-1}\mathbb{E}^{Q}u$, where Q is the compound probability

$$Q = \sum_{s=1}^{S} Q^s \pi^s. \tag{3}$$

If $u \neq \varphi$, priors and the prior on these priors cannot be compounded.

2.3 Divergence CE

The second extension of an EU CE we consider corresponds to smooth divergence preferences, which are within the axiomatic setting of Maccheroni, Marinacci, and Rustichini (2006a). We use the term *divergence index* to mean any strictly convex differentiable function of the form $\varphi : (0, \infty) \to \mathbb{R}$ such that

$$\varphi(1) = \varphi'(1) = 0 \quad \text{and} \quad \varphi'(\infty) = \infty. \tag{4}$$

Analogously to (1), we introduce the notation

$$C^n_{\text{div}} = \text{ set of } n \text{ times continuously differentiable divergence indices.}$$

We define a (smooth) *divergence CE* in terms of a vNM index u, a reference prior Q and a divergence index φ by

$$\upsilon(U) = \inf_{Q \in \Pi} u^{-1}\left(\mathbb{E}^{Q} u(U) + \mathbb{E}\varphi\left(\frac{dQ}{dP}\right)\right), \quad U \in (\ell, \infty)^{\Omega}. \tag{5}$$

The lower values φ takes the more risk/uncertainty averse the CE υ becomes, since it assigns lower values to risky outcomes.

Recall that Π is the set of all probabilities that assign a (strictly) positive weight to each state. We will focus on the case in which the infimum is achieved by some $Q \in \Pi$, a case that is fully characterized in Proposition 1 below. In the following section, we will see that the infimum is always achieved as a minimum for sufficiently small Brownian risks, and it is always achieved for Poissonian risks with negative jumps.

Proposition 1 *Let $U_{\max} = \max_{\omega \in \Omega} U(\omega)$. The infimum in* (5) *is achieved as a minimum by some $Q \in \Pi$ if and only if*

$$\mathbb{E}\varphi'^{-1}\left(\varphi'(0+) + u(U_{\max}) - u(U)\right) < 1, \tag{6}$$

in which case the minimizing Q is given by $dQ/dP = \varphi'^{-1}(\alpha - u(U))$, where the scalar α uniquely solves

$$\mathbb{E}\varphi'^{-1}(\alpha - u(U)) = 1 \quad \text{and} \quad \alpha > \varphi'(0+) + u(U_{\max}). \tag{7}$$

Example 2 (Entropic CE) *Define the functions $\psi : \mathbb{R} \to \mathbb{R}$ and $\tilde{\psi} : (0, \infty) \to \mathbb{R}$ by*[6]

$$\psi(x) = \theta\left(1 - \exp\left(-\frac{x}{\theta}\right)\right) \quad \text{and} \quad \tilde{\psi}(y) = \theta(y \log y - y + 1), \tag{8}$$

where $\theta \in (0, \infty)$ is a given parameter. The divergence CE (5) *is defined to be* entropic *if $\varphi = \tilde{\psi}$ for some θ (necessarily equal to $\varphi''(1)$). In this case, $\varphi'(0+) = -\infty$ and therefore condition* (6) *is satisfied for any U. The corresponding minimizing probability Q is given by Proposition 1 through the density $dQ/dP = \exp\left(-\theta^{-1}u(U)\right)/\mathbb{E}\exp\left(-\theta^{-1}u(U)\right)$. Computing the corresponding minimum proves the well-known*[7] *identity*

$$\upsilon = (\psi \circ u)^{-1}\, \mathbb{E}(\psi \circ u). \tag{9}$$

In conclusion, an entropic CE is an expected utility CE; the divergence CE (5) *with $\varphi = \tilde{\psi}$ modifies the EU CE $u^{-1}\mathbb{E}u$ by an exponential concavification of the vNM index u, thus increasing risk aversion. We will show later (under a smoothness assumption) that the entropic CE is the only divergence CE that is also an expected utility CE, even in an approximate sense.*

Example 3 (Quadratic Divergence) *Suppose that $\varphi(y) = (\theta/2)(y-1)^2$ for some $\theta \in (0, \infty)$. This case is covered by Theorem 24 of Maccheroni, Marinacci, and Rustichini (2006a). We briefly review the main conclusion as a corollary of Proposition 1. Assume the validity of Condition* (6), *which in this context reduces to*

$$u(U) - \mathbb{E}u(U) < \theta. \tag{10}$$

Equation (7) *results in $\alpha = \mathbb{E}[u(U)]$, and therefore the minimizing prior Q is given by $dQ/dP = 1 - (u(U) - \mathbb{E}u(U))/\theta$. Calculating the minimum using the preceding expression results in*

$$\upsilon(U) = u^{-1}\left(\mathbb{E}u(U) - \frac{1}{2\theta}\, Var[u(U)]\right).$$

Condition (10) *guarantees that U is valued within the range where the right-hand side defines a strictly increasing function.*

[6] Note the convex duality: $\tilde{\psi}(y) = \max_x\{\psi(x) - xy\}$ and $\psi(x) = \min_{y>0}\{\tilde{\psi}(y) + xy\}$.
[7] See, for example, Donsker and Varadhan (1975).

3 CE Approximations for Small Risks

The paper's main ideas are established in this section with the simplest nontrivial instance of the uncertainty model, where the state space consists of only two states. Despite the simplicity of the two-state model, this section's arguments go to the core of the matter, since every one-dimensional Brownian motion and Poisson process is the limit of a random walk whose steps look exactly like one of the models analyzed here. We use $h \in (0,1)$ to parameterize the time length of the single period; preferences are expressed at time zero without knowledge of the state, which is revealed at time h. In a dynamic extension, this section's model would correspond to a single node of a binomial tree. Multiple stochastically independent binomial trees of this type, properly normalized, converge to the continuous-time model to follow, as the period length h goes to zero. We are therefore interested in approximations that are valid for small values of h. In expressing these approximations, we write $o(h)$ to represent some function $\varepsilon : (0,1) \to \mathbb{R}$ such that $\varepsilon(h)/h \to 0$ as $h \to 0$. Every instance of little oh can represent a potentially different function ε.

3.1 Unified Risk-Source Structure

We will introduce two uncertainty models, corresponding to Brownian and Poissonian risks, parameterized by the period length h. Each model is specified by the state space $\Omega = \{0,1\}$, which does not depend on h, and a pair of a probability $P : 2^{\{0,1\}} \to (0,1)$ and a random variable $B : \{0,1\} \to \mathbb{R}$ that can vary with h. We therefore interpret the pair (P,B) as a whole family $\left\{(P_h, B_h) : h \in \left(0,\bar{h}\right)\right\}$, for a sufficiently small $\bar{h} > 0$, even though the dependence on h is notationally suppressed.

The pair (P,B) is normalized so that

$$\mathbb{E}B = 0 \quad \text{and} \quad \mathbb{E}\left[B^2\right] = h. \tag{11}$$

Since there are only two states, any random variable $x : \Omega \to \mathbb{R}$ has the *canonical decomposition*

$$x = \alpha h + \beta B, \quad \text{where} \quad \alpha = \frac{1}{h}\mathbb{E}x \quad \text{and} \quad \beta = \frac{\mathbb{E}[Bx]}{\mathbb{E}[B^2]} = \frac{1}{h}\mathbb{E}[Bx].$$

The parameters α and β represent, respectively, the drift and volatility of a random walk whose increments are i.i.d. copies of x.

Fixed throughout is the random variable

$$U = U_0 + \mu h + \Sigma B, \tag{12}$$

where the scalars U_0, μ and Σ are given parameters, whose value does not change with h. (Adding a term $o(h)$ to the above expression for U does not affect the results.) For now, U can be thought of as a payoff realized at time h. In the dynamic setting to follow, U will represent a continuation value.

Our objective is the computation of simplifying approximations of the CE of U, for each of last section's specifications. For this purpose, the dependence of priors and CEs on h must be clarified.

A prior Q defines the scalar ρ through the canonical decomposition

$$\frac{dQ}{dP} = 1 + \rho B, \quad \text{where} \quad \rho = \frac{1}{h}\mathbb{E}\left[B\frac{dQ}{dP}\right] = \frac{1}{h}\mathbb{E}^Q B. \tag{13}$$

Based on the last equation, we call ρ the *drift* of B under Q. (In the Poissonian case, we will see that $1+\rho$ represents an arrival rate under Q.) Since B depends on h, either Q or ρ must be allowed to vary with h in order for (13) to remain valid as h goes to zero. We assume that the drift ρ (or arrival rate $1+\rho$) does *not* vary with h, and we regard each prior Q as a family of probabilities $\left\{Q_h : h \in \left(0, \bar{h}\right)\right\}$ such that $\rho = \mathbb{E}^{Q_h} B/h$ for all h. As with P, B and U, we notationally suppress the dependence of Q on h.

Consider any prior Q and corresponding drift ρ of B. The canonical decomposition of U can be expressed as

$$U = U_0 + (\mu + \rho\Sigma)\, h + \Sigma B^Q, \quad \text{where } B^Q = B - \rho h. \tag{14}$$

The first two moments of B^Q under Q are

$$\mathbb{E}^Q B^Q = 0 \quad \text{and} \quad \mathbb{E}^Q\left[(B^Q)^2\right] = h + \rho\mathbb{E}\left[B^3\right] - (\rho h)^2. \tag{15}$$

We will see later that in the Brownian model, $\mathbb{E}\left[B^3\right] = 0$ and therefore the second moment of B^Q under Q is $h + o\,(h)$, reflecting the fact that the drift but not the volatility of U is sensitive to a change of prior. In the Poissonian model, $\mathbb{E}\left[B^3\right] = h + o\,(h)$ and the second moment of B^Q under Q is $(1+\rho)\, h + o\,(h)$, reflecting the fact that the arrival rate under Q is $1+\rho$.

The CE approximations to be derived will take the general form

$$\upsilon\,(U) = U_0 + (\mu + \rho\cdot\Sigma - \mathcal{A}\,(U_0, \Sigma, \rho))\, h + o\,(h)\,. \tag{16}$$

The parameter ρ determines a prior Q that defines the risk-neutral CE

$$\mathbb{E}^Q U = U_0 + (\mu + \rho\cdot\Sigma)\, h.$$

The term $\mathcal{A}\,(U_0, \Sigma, \rho)\, h$ is a risk/uncertainty-aversion penalty to the risk-neutral CE under the prior Q and will be given in closed form for the three CE specifications.

In order to make sense of (16), it is important to clarify how the left-hand side $\upsilon\,(U)$ depends on h for each of last section's CE specifications. As a general rule, only probabilities over Ω used to construct the CEs can depend on h – all other CE parameters (including the priors on priors in the case of a second-order EU CE) are constant as h goes to zero. The following table summarizes all the dependencies on h discussed so far.

quantities that	do *not* vary with h	*can* vary with h
uncertainty model	Ω	P, B
priors	drift ρ, arrival rate $1+\rho$	any prior Q on Ω
payoff structure	U_0, μ, Σ	U
EU CE	u	Q
second-order EU CE	$u,\ \varphi;\ \pi^1, \ldots, \pi^S$	$Q^1, \ldots, Q^S$
divergence CE	$u,\ \varphi$	a minimizing prior Q

3.2 Brownian Risk

Brownian motion can be thought of as the limit of a random walk that moves up or down by a fixed increment, following the outcome of a coin toss. The coin is tossed N times per unit of time at a steady rate (so $h = 1/N$), while the increment after each toss is normalized so that the variance of the total increment over a unit of time is one. Brownian motion is the limit as N goes to infinity. If the coin is perfectly balanced, then the random walk has zero mean and converges to standard Brownian motion. The decision maker thinks that the coin can be slightly biased but is uncertain about the extent of the bias. If the probability of heads is changed from one half to $(1/2)\left(1 + \mu N^{-1/2}\right)$, for large N, then the drift of the random walk changes from zero to μ, while the variance per unit of time remains approximately the same. For example, in an application in which the random walk represents a stock price index, the drift μ represents a risk premium and has critically important implications for portfolio choice. On the other hand, the volatility of the index is relatively easy to estimate and is treated in the ideal Brownian limit as being observable — every Brownian path contains a time series that estimates the volatility with zero statistical error.

We focus on a single increment of the random walk whose possible outcomes are

$$\begin{cases} \text{state 1:} \quad B(1) = \sqrt{h} \quad \text{with probability} \quad P(1) = 1/2, \\ \text{state 0:} \quad B(0) = -\sqrt{h} \quad \text{with probability} \quad P(0) = 1/2. \end{cases} \tag{17}$$

Donsker's theorem (see, for example, Theorem 14.1 of Billingsley (1999)) shows that a random walk whose increments are i.i.d. copies of B converges to Brownian motion as h goes to zero. The formalism of the last subsection applies. Given any scalar ρ, the prior Q such that $\mathbb{E}^Q B = \rho h$ for all sufficiently small $h > 0$ is given by

$$Q(1) = \frac{1}{2}\left(1 + \rho\sqrt{h}\right) \quad \text{and} \quad Q(0) = \frac{1}{2}\left(1 - \rho\sqrt{h}\right). \tag{18}$$

Under Q, a properly normalized random walk whose increments are i.i.d. copies of $B^Q = B - \rho h$ also converges to Brownian motion as h goes to zero.

The following result, which is proved in Appendix B, summarizes the implications of the Brownian risk assumption for smooth versions of the three CE specifications of Section 2. The coefficient of absolute risk aversion of a vNM index u is denoted

$$a^u = -\frac{u''}{u'}. \tag{19}$$

Theorem 4 (Brownian CE Approximations) *Suppose* (P, B) *is defined by* (17). *The following approximations are valid for any* $u \in C^2_{vNM}$.

(a) (Expected Utility CE) Suppose there exists a constant ρ *such that* $\mathbb{E}^Q B = \rho h$ *for all sufficiently small* h. *Then*

$$u^{-1}\mathbb{E}^Q u(U) = U_0 + \left(\mu + \rho\Sigma - \frac{1}{2}a^u(U_0)\Sigma^2\right)h + o(h). \tag{20}$$

(b) (Second-Order Expected Utility CE) Suppose υ *is a second-order EU CE as defined in Section 2.2 for some* $\varphi \in C^1_{vNM}$ *and fixed prior* π *on the priors* $\{Q^1, \ldots, Q^S\}$. *If there exist constants* $\rho^1, \ldots, \rho^S$ *such that* $\mathbb{E}^{Q^s} B = \rho^s h$ *for all sufficiently small* h, *then*

$$\upsilon(U) = u^{-1}\mathbb{E}^Q u(U) + o(h), \quad \text{where } Q = \sum\nolimits_s Q^s \pi^s.$$

(c) (Divergence CE) Suppose υ *is a divergence CE as defined in Section 2.3 for some divergence index* $\varphi \in C^2_{div}$, *and let* $\theta = \varphi''(1)$. *Then*

$$\upsilon(U) = (\psi \circ u)^{-1}\mathbb{E}(\psi \circ u)(U) + o(h), \quad \text{where} \quad \psi(x) = \theta\left(1 - \exp\left(-\frac{x}{\theta}\right)\right). \tag{21}$$

Part (a) is the familiar Arrow-Pratt CE approximation (Arrow (1965, 1970) and Pratt (1964)) applied to the payoff approximation (12) relative to the prior Q.

Part (b) is essentially a corollary of the first part. Equation (20) applied to each Q^s implies that we can approximate the second-order EU CE, up to $o(h)$, by using a linear approximation of φ, in which case the terms φ^{-1} and φ in the CE definition cancel out. Since φ becomes irrelevant to this approximation, we can set it equal to u, allowing priors and the prior on priors to be compounded. Therefore, the second-order EU CE is approximately equal to the corresponding expected utility CE obtained by setting $\varphi = u$ and compounding priors.

Part (c) of Theorem 4 can be understood in terms of Example 2, stating that for an entropic CE, approximation (21) holds as an exact relationship (that is, without the $o(h)$ term). To see how, use restrictions (4) on φ in a second-order Taylor expansion to compute

$$\mathbb{E}\varphi\left(\frac{dQ}{dP}\right) = \mathbb{E}\varphi(1 + \rho B) = \frac{\theta}{2}\rho^2 h + o(h). \tag{22}$$

To first order, any smooth divergence index φ such that $\varphi''(1) = \theta$ results in approximately the same divergence CE value $\upsilon(U)$, which must therefore be approximately equal to its value for the entropic case. The rigorous proof in Appendix B deals carefully with some subtleties regarding the interchange of the order of approximation and minimization, and does not rely on the result of Example 2.

Finally, note that the approximation of part (a) can be applied to (21) together with the identity $a^{\psi \circ u} = a^u + \theta^{-1}u'$. Adding robustness to the EU CE $u^{-1}\mathbb{E}u$ through a divergence index φ is approximately equivalent to increasing the risk-aversion coefficient $a^u(U_0)$ by $u'(U_0)/\theta$.

3.3 Poissonian Risk

A Poisson process is also the limit of a random walk, but with quite different characteristics than Brownian motion. Consider an urn that contains a large number $N-1$ of black balls and a single white ball. The decision maker is certain that all balls in the urn but one are black, but is uncertain about the total number N of balls. A ball is drawn from the urn, with replacement, at a steady rate. Every time a black ball is selected, nothing happens. Every time the white ball is selected, the random walk jumps up by one unit. This unit jump can be thought of as the indicator of an event that in applications could, for example, represent a default event or a natural disaster. The Poisson process is the limiting case as N goes to infinity, provided the rate at which a ball is drawn is also increased with N so that the expected number of event arrivals per unit of time remains constant (so that $h = 1/N$ given unit arrival rate). The decision maker is uncertain about the arrival rate, which is statistically hard to estimate given time series in which the event appears a small number of times.

We focus on a single increment of a conditionally demeaned version of the random walk just described informally, assuming unit arrival rate. The possible outcomes are

$$\begin{cases} \text{state 1:} \quad B(1) = 1 - h - \varepsilon_h & \text{with probability} \quad P(1) = h + \varepsilon_h, \\ \text{state 0:} \quad B(0) = -h - \varepsilon_h & \text{with probability} \quad P(0) = 1 - h - \varepsilon_h, \end{cases} \tag{23}$$

where $\varepsilon_h = o(h)$. While any such choice of ε_h yields the same results, for concreteness, we set $\varepsilon_h = 2^{-1} - h - \sqrt{4^{-1} - h}$, which implies that $\mathbb{E}\left[B^2\right] = h$, and therefore that the formalism of Section 3.1 applies exactly. A random walk whose increments are i.i.d. copies of B converges to a compensated Poisson process with unit arrival rate as h goes to zero. A rigorous version of this statement and its proof can be found in Billingsley (1999) (Example 12.3). The assumption of a unit arrival rate can be thought of as the definition of the unit of time and therefore entails no loss of generality.

Given any scalar ρ, the prior Q such that $\mathbb{E}^Q B = \rho h$ for all sufficiently small $h > 0$ is given by

$$Q(1) = 1 - Q(0) = (1+\rho)h + \varepsilon_h. \tag{24}$$

The distribution of $B^Q = B - \rho h$ under Q is

$$B^Q = \begin{cases} 1 - (1+\rho)h - \varepsilon_h & \text{with } Q\text{-probability } (1+\rho)h + \varepsilon_h, \\ -(1+\rho)h - \varepsilon_h & \text{with } Q\text{-probability } 1 - (1+\rho)h - \varepsilon_h. \end{cases} \tag{25}$$

Up to a $o(h)$ adjustment, this is the same distribution as (23) once h is rescaled to $(1+\rho)h$, which means that a change of prior in this setting is essentially equivalent to a change of the unit of time. Under Q, a random walk whose increments are i.i.d. copies of B^Q converges, as h goes to zero, to a compensated Poisson process with arrival rate $1+\rho$.

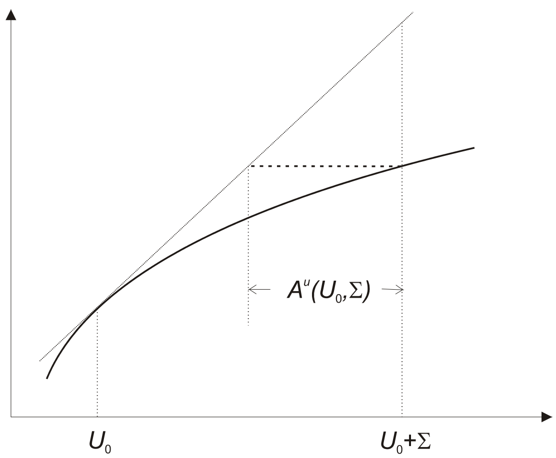

Figure 1: Graphical representation of $A^u(U_0, \Sigma)$. The curved solid line is part of the graph of u, while the slanted straight line is tangent to the graph at U_0.

Theorem 5 below summarizes the implications of Poissonian risk for the three CE specifications of Section 2. For its statement, several pieces of new notation are now introduced. In the context of Poissonian risk, the role of a coefficient of absolute risk aversion of a vNM index u is assumed by the quantity

$$A^u(U_0, \Sigma) = \Sigma - \frac{u(U_0 + \Sigma) - u(U_0)}{u'(U_0)}, \tag{26}$$

which is represented graphically in Figure 1. While $a^u(U_0)$ is a local measure of risk aversion toward risks taking values near U_0, $A^u(U_0, \Sigma)$ is a measure of risk aversion toward risks that take the value $U_0 + \Sigma$ with a small probability, and the value U_0 otherwise. Both a^u and A^u are invariant to positive affine transformations of u. The inequality $A^u \geq 0$ is equivalent to the gradient inequality for u, and therefore concavity of u is equivalent to the nonnegativity of A^u. It is also worth noting parenthetically that if $u \in C^2_{\text{vNM}}$, then

$$A^u(U_0, \Sigma) = \frac{1}{2}a^u(U_0)\Sigma^2 + o(\Sigma^2),$$

as can be seen by taking a second-order Taylor series expansion of u in (26). In our current context, however, Σ is not assumed to be small and u need not have a second derivative.

For simplicity, we focus on the case in which the minimization problem defining a divergence CE has an interior solution for all sufficiently small h. Proposition 1 (or a simple direct calculation) shows that the latter condition is equivalent to membership of the reference (U_0, Σ) to the set

$$D = \{(U_0, \Sigma) \in (\ell, \infty) \times \mathbb{R} : U_0 + \Sigma > \ell \text{ and } u(U_0 + \Sigma) - u(U_0) < -\varphi'(0+)\}. \tag{27}$$

If either $\Sigma \leq 0$ or $\varphi'(0+) = -\infty$ (as in the entropic case), then $(U_0, \Sigma) \in D$ for all $U_0 > \ell - \Sigma$. If $\varphi'(0+) > -\infty$ and Σ is positive and sufficiently high, a minimizing probability does assign zero probability to the positive jump, which is the case being excluded here.

Given any function $\zeta : (-\infty, -\varphi'(0+)) \to \mathbb{R}$, we define the notation

$$A_\zeta^u(U_0, \Sigma) = \Sigma - \frac{\zeta(u(U_0 + \Sigma) - u(U_0))}{u'(U_0)}, \tag{28}$$

which extends the notation in (26), since $A^u = A^u_{\text{identity}}$. In fact, the last identity can be viewed as a limiting case (for $\theta = \infty$) of the more interesting identity

$$A^{\psi \circ u} = A^u_\psi, \quad \text{where} \quad \psi(u) = \theta\left(1 - \exp\left(-\frac{u}{\theta}\right)\right). \tag{29}$$

We are now in a position to state the promised smooth CE approximations under Poissonian risk. The proof can be found in Appendix B.

Theorem 5 (Poissonian CE Approximations) *Suppose (P, B) is defined by (23). The following approximations are valid for any $u \in C^1_{vNM}$.*

(a) (Expected Utility CE) Suppose there exists a constant ρ such that $\mathbb{E}^Q B = \rho h$ for all sufficiently small h. Then

$$u^{-1}\mathbb{E}^Q u(U) = U_0 + (\mu + \rho\Sigma - A^u(U_0, \Sigma)(1 + \rho))h + o(h).$$

(b) (Second-Order EU CE) The statement of part (b) of Theorem 4 is valid in the current context.

(c) (Divergence CE) Suppose υ is a divergence CE as defined in Section 2.3 for some $\varphi \in C^1_{div}$. If $(U_0, \Sigma) \in D$, then

$$\upsilon(U) = U_0 + \left(\mu - A^u_\zeta(U_0, \Sigma)\right)h + o(h), \quad \textit{where} \quad \zeta(x) = \min_{y \in (0,\infty)} \{xy + \varphi(y)\}. \tag{30}$$

Finally, there exists some prior W and $w \in C^1_{vNM}$ such that

$$\upsilon(U) = w^{-1}\mathbb{E}^W w(U) + o(h) \quad \textit{for all } (U_0, \Sigma) \in D \tag{31}$$

if and only if $W = P$ and $\varphi(y) = \theta(y \log y - y + 1)$ for some $\theta > 0$, in which case Example 2 applies and $\zeta = \psi$.

The expected utility CE approximation of part (a) extends the Arrow-Pratt approximation to Poissonian small risks. Given part (a), the argument that gives part (b) is essentially the same as in the Brownian case. A smooth second-order EU CE is therefore approximately an expected utility CE given a small risk, whether the risk is Brownian or Poissonian.

Part (c) of Theorem 5 gives this paper's first instance of a CE approximation that is *not* consistent with expected utility. In Example 2, we saw that an entropic divergence CE is an expected utility CE, which is consistent with identity (29). Theorem 5(c) gives a strong converse: *If a smooth divergence CE can be approximated by a smooth expected utility CE, then it must be entropic.*

3.4 Divergence and Source-Dependent Risk Aversion

The different ways in which a divergence index affects risk aversion toward Brownian and Poissonian risks is more interesting within a model that allows both types of risk, resulting in source-dependent risk aversion. This section outlines a straightforward extension of our earlier results that captures this essential idea; a more general version will be stated in the continuous-time setting.

The CE v is assumed to be a divergence CE relative to an arbitrary given prior R :

$$v(U) = \inf_{Q \in \Pi} u^{-1}\left(\mathbb{E}^Q u(U) + \mathbb{E}^R \varphi\left(\frac{dQ}{dR}\right)\right),$$

where $\varphi \in C^2_{\text{div}}$. This differs from our earlier definition (5), where we assumed the normalization $R = P$ (implying zero drift in the Brownian case and a unit arrival rate in the Poissonian case). Theorem 4(c) remains valid, with $\mathbb{E}^R$ in place of $\mathbb{E}$ and only straightforward modifications to the proof. Theorem 5(c) extends to this case with a simple change of time units—what was approximation $v(U) = \mathbb{E}U - A^u_\zeta(U_0, \Sigma) h + o(h)$ now becomes $v(U) = \mathbb{E}^R U - A^u_\zeta(U_0, \Sigma)(1+\rho) h + o(h)$, where $1 + \rho$ is the arrival rate implied by R.

The risk source model is now extended to be the product of the Brownian and Poissonian models. Let (P_1, B_1) be the Brownian risk model of Section 3.2, let (P_2, B_2) be the Poissonian risk model of Section 3.3, and consider the model (P, B), where $P = P_1 \times P_2$ and $B = (B_1, B_2)'$, defined on $\Omega = \{0,1\} \times \{0,1\}$. The formalism of Section 3.1 has an easy extension to this setting. Let $\rho = (\rho_1, \rho_2)' = h^{-1}\mathbb{E}^R B$. The parameter ρ_1 (resp. $1 + \rho_2$) represents the drift (resp. arrival rate) under R of a random walk whose steps are i.i.d. copies of B_1 (resp. B_2), which in the continuous-time limit becomes a Brownian motion (resp. compensated Poisson process). The risky payoff is $U = U_0 + \mu h + \Sigma' B$, where the parameters $U_0, \mu \in \mathbb{R}$ and $\Sigma \in \mathbb{R}^2$ do not depend on h, and $(U_0, \Sigma_2) \in D$ (which is necessary and sufficient for an interior minimizing prior). Since Taylor series approximations are additive, approximation (16) is valid with

$$\mathcal{A}(U_0, \Sigma, \rho) = \frac{1}{2} a^{\psi \circ u}(U_0)(\Sigma_1)^2 + A^u_\zeta(U_0, \Sigma_2)(1 + \rho_2), \quad (U_0, \Sigma_2) \in D, \tag{32}$$

where ψ and ζ are defined in (29) and (30), respectively.

Let us decompose the divergence index as $\varphi(\cdot) = \theta\phi(\cdot)$, where $\theta = \varphi''(1)$. Note that ϕ can be any twice continuously differentiable function from $(0, \infty)$ to $[0, \infty)$ that satisfies $\phi(1) = \phi'(1) = 0$, $\phi'(\infty) = \infty$ and $\phi''(1) = 1$. The parameter θ, but not ϕ, modifies the first term of $\mathcal{A}$, which represents an adjustment for exposure to the Brownian risk source B_1. With the latter fixed through the choice of θ, the parameter ϕ can be used to modify the second term in $\mathcal{A}$, which represents an adjustment for exposure to the Poissonian risk source B_2. It is worth noting that ρ enters $\mathcal{A}$ only through its second component ρ_2. This is because ρ_1 changes only the drift of the Brownian risk and not the volatility Σ_1. On the other hand, exposure to Poissonian risk is measured by both the jump size Σ_2 and the arrival rate $(1 + \rho_2)$. In the entropic case (see Example 2), ϕ is entirely determined by θ, and the flexibility in representing source-dependent risk aversion is sacrificed, reflecting the fact that the entropic CE is in fact an EU CE.

4 Continuous-Time Recursive Utility

With the rigorous CE approximation results over a single step of a binomial tree completed, in this section we turn to the formulation of a corresponding continuous-time recursive utility given Brownian/Poisson information. The first subsection presents the stochastic setting. The second subsection summarizes the mathematical form of a continuous-time recursive utility, with arguments that parallel those in Duffie and Epstein (1992) and Skiadas (2008), extended to include jumps. The core material is given in the third subsection, where the general structure is specialized to correspond to the conditional CEs of interest, using expressions that are analogous to last section's CE approximations. The associated calculations using Ito's lemma are given in Appendix A. The presentation assumes only an intuitive grasp of stochastic calculus with jumps at the level of the appendices of Duffie (2001)—the advanced theory can be found in Applebaum (2004) and Jacod and Shiryaev (2003). For the reader comfortable with stochastic calculus, the heuristic arguments of Appendix A offer a shortcut to understanding relative to the rigorous proofs of Appendix B for the single-period model. Conversely, the reader unfamiliar with stochastic calculus can leverage the understanding of the single-period model to interpret continuous-time heuristics.

4.1 Stochastic Setting

Given is a probability space $(\Omega, \mathcal{F}, P)$ on which are defined d mutually stochastically independent processes forming the column vector $B = (B^1, \ldots, B^k, B^{k+1}, \ldots, B^d)'$, where

- B^i is a standard Brownian motion for $i = 1, \ldots, k$, and
- B^i is a compensated Poisson process with unit arrival rate for $i = k+1, \ldots, d$.

The last statement means that there exist independent Poisson processes $N^{k+1}, \ldots, N^d$ such that $\mathbb{E}N_t^i = t$ and $B_t^i = N_t^i - t$ for every time t and process $i \in \{k+1, \ldots, d\}$.

The underlying filtration $\{\mathcal{F}_t : t \in [0,T]\}$, where $T > 0$ is a given finite time horizon, is defined as the smallest filtration such that B_t is $\mathcal{F}_t$-measurable for every time t (and each $\mathcal{F}_t$ contains the P-null events, a technicality that can be ignored). We assume that $\mathcal{F} = \mathcal{F}_T$, without loss of generality. Intuitively, uncertainty is represented by the possible paths of B. At time t the agent observes the realized path of B up to time t. Given that information, conditional uncertainty resolved by time $t + dt$ is spanned by the stochastically and linearly independent factors $dB_t^1, \ldots, dB_t^d$, representing infinitesimal risks, some of which are Brownian and some Poissonian. The approximations of Section 3, which become exact relationships in the limit, will apply with dt in place of h, and dB^i in place of B.

A process X is *adapted* if X_t is $\mathcal{F}_t$-measurable for every time t. We will not enter into the technical definition of a *predictable* process X, but we think of the concept heuristically as the condition that X_t is $\mathcal{F}_{t-}$-measurable. For any process X whose paths have left limits, we use the heuristic notation $dX_t = X_{t+dt} - X_{t-}$, where X_{t-} denotes the left limit of X at t.

Conditional expectation given time-t information $\mathcal{F}_t$ (resp. $\mathcal{F}_{t-}$) is denoted $\mathbb{E}_t$ (resp. $\mathbb{E}_{t-}$). A *volatility process* is any d-dimensional predictable process σ such that $\int_0^T \sigma_t' \sigma_t dt < \infty$ with probability one. A local martingale M can be thought of as an adapted process whose instantaneous increments have conditionally zero mean: $\mathbb{E}_{t-} dM_t = 0$. Given an integrability condition, such an increment dM_t can be expressed as a linear combination of the instantaneous linear factors $dB_t^1, \ldots, dB_t^d$. This intuition is formalized by a martingale representation theorem, which in the current context states that a process M is a locally square-integrable martingale if and only if there exists some volatility process σ such that

$$M_t = M_0 + \int_0^t \sigma_u' dB_u, \quad \text{or equivalently} \quad dM_t = \sigma_t' dB_t. \tag{33}$$

A *prior* is any probability on $\mathcal{F}$ that is equivalent to P (meaning that it defines the same null events as P). Associated with a prior Q are a martingale ξ^Q, a d-dimensional predictable process ρ^Q, and a d-dimensional adapted process B^Q, defined by

$$\xi_t^Q = \mathbb{E}_t \left[\frac{dQ}{dP} \right], \quad \frac{d\xi_t^Q}{\xi_{t-}^Q} = \rho_t^{Q\prime} dB_t \quad \text{and} \quad B_t^Q = B_t - \int_0^t \rho_t^Q dt. \tag{34}$$

(Note that the ratio $\xi_{t+dt}^Q / \xi_{t-}^Q = 1 + \rho_t^{Q\prime} dB_t$ is the analog of the ratio $dQ/dP = 1 + \rho B$ in the single period model, where Q and P represent transition probabilities over a single binomial step. Here Q and P represent probabilities over entire paths.) We let Π denote the set of every prior Q such that ρ^Q is a volatility process.[8] Girsanov's theorem implies that for any $Q \in \Pi$,

$$B^Q \text{ is a local martingale under } Q. \tag{35}$$

Expectation relative to a prior Q is denoted $\mathbb{E}^Q$. Heuristically, we have $\mathbb{E}_{t-}^Q dB_t^Q = 0$.

Priors in this setting correspond to beliefs about the drift process of each Brownian motion and the arrival rate process of each Poisson process, all of which can be path dependent. For the Brownian factors, Lévy's characterization of Brownian motion implies that $B^{Q1}, \ldots, B^{Qk}$ are independent standard Brownian motions under the probability Q, and therefore ρ^{Qi} is the drift of Brownian motion B^{Qi} under Q. For the Poissonian factors, we note that

$$B_t^{Qi} = N_t^i - \int_0^t \left(1 + \rho_s^{Qi}\right) ds, \quad i = k+1, \ldots, d, \tag{36}$$

which in combination with (35) implies that $1 + \rho^{Qi}$ is the arrival rate process of the point process N^i under the probability Q.

[8] The conditional density process ξ^Q can be recovered from ρ^Q by the formula

$$\log \xi_t^Q = \sum_{i=1}^k \int_0^t \left(\rho_s^{Q\prime} dB_s^i - \frac{1}{2} \left(\rho_s^{Qi}\right)^2 ds \right) + \sum_{i=k+1}^d \int_0^t \left(\log\left(1 + \rho_s^{Qi}\right) dB_s^i + \left(\log\left(1 + \rho_s^{Qi}\right) - \rho_s^{Qi} \right) ds \right),$$

as can be verified by an application of Ito's lemma.

4.2 Recursive Utility as a BSDE

As in the single-period analysis, we assume consumption is valued in the interval (ℓ, ∞), for some $\ell \in [-\infty, 1)$. Ignoring technical integrability conditions, a *consumption plan* is any (ℓ, ∞)-valued adapted process c, where c_t represents a consumption rate if $t < T$, and c_T represents a lump-sum terminal consumption. The consumption plan that is identically equal to one (including unit terminal lump-sum consumption) is denoted $\mathbf{1}$. For each consumption plan c, we will define a corresponding utility process $U(c)$, which is normalized so that $U(s\mathbf{1}) = s\mathbf{1}$ for any $s \in (\ell, \infty)$. The interpretation is that at time t, conditionally on time-t information, the agent is indifferent between the plans c and $U_t(c)\mathbf{1}$.

We henceforth fix a reference consumption plan c and we simplify the notation for the corresponding utility process by writing U instead of $U(c)$. Heuristically, we assume that the process U is computed by a backward-in-time recursion of the form[9]

$$U_{t-} = \Phi\left(dt, c_t, \upsilon_t\left(U_{t+dt}\right)\right), \quad U_T = c_T, \tag{37}$$

where $\upsilon_t\left(U_{t+dt}\right)$ is a time-t conditional CE of U_{t+dt}, to be specified later on, and Φ is an intertemporal aggregator. In most applications,

$$\Phi\left(dt, c, \upsilon\right) = u^{-1}\left(\left(1 - e^{-\beta dt}\right) u\left(c\right) + e^{-\beta dt} u\left(\upsilon\right)\right), \tag{38}$$

for some $u \in C^2_{\mathrm{vNM}}$ (as defined in section 2.1) and impatience parameter $\beta \in [0, \infty)$. In general, Φ is assumed to have continuous partial derivatives, denoted Φ_{dt}, Φ_c and Φ_υ, with Φ_c and Φ_υ taking values in $(0, \infty)$, reflecting preference monotonicity. Moreover, Φ is assumed to satisfy the consistency condition $\Phi(0, c, \upsilon) = \upsilon$. For practical purposes, the functional form (38) can be assumed throughout.

To make mathematical sense of the heuristic recursion (37), suppose the processes μ and Σ represent, respectively, the drift and volatility of U :

$$dU_t = \mu_t dt + \Sigma_t' dB_t = \left(\mu_t + \rho_t^Q \cdot \Sigma_t\right) dt + \Sigma_t' dB_t^Q, \tag{39}$$

for any prior Q, with B^Q defined in (34). Note that $dU_t = U_{t+dt} - U_{t-}$ includes any time-t jump. Since $\mathbb{E}_{t-}^Q dB_t^Q = 0$, the second equation in (39) implies that the risk-neutral conditional CE of U_{t+dt} under the prior Q is

$$\mathbb{E}_{t-}^Q U_{t+dt} = U_{t-} + \left(\mu_t + \rho_t^Q \cdot \Sigma_t\right) dt. \tag{40}$$

[9] Decision-theoretic foundations for recursive utility have been developed in a number of papers, including Kreps and Porteus (1978), Selden (1978), Johnsen and Donaldson (1985), Chew and Epstein (1989), Epstein and Zin (1989), Skiadas (1998), Wang (2003), Hayashi (2005), and Klibanoff and Ozdenoren (2007). Kreps and Porteus (1978) emphasized the role of preferences for the timing of resolution of uncertainty (an issue that is cast more broadly in Skiadas (1998) as preferences over filtrations). In our context, the information tree is fixed, and as a consequence the axioms that lead to recursive utility with an arbitrary conditional CE are quite simple; they can be found in Chapter 6 of Skiadas (2009).

Reflecting the discrete-time approximation (16) , the continuous-time conditional CEs of interest in this paper can be expressed as

$$\upsilon_t\left(U_{t+dt}\right) = U_{t-} + \left(\mu_t + \rho_t^Q \cdot \Sigma_t - \mathcal{A}\left(U_{t-}, \Sigma_t, \rho_t^Q\right)\right) dt, \tag{41}$$

where the function $\mathcal{A}$ represents a risk/ambiguity aversion adjustment to the risk-neutral conditional CE (40) . This claim is explained in the following subsection, where an explicit expression for $\mathcal{A}$ is given for each of the CE types introduced in the first part of this paper.

Let us now expand the right-hand side of the heuristic recursion (37) in a first-order Taylor approximation with respect to the arguments dt and $\upsilon_t\left(U_{t+dt}\right)$, using expression (41), to obtain

$$U_{t-} = \Phi\left(0, c_t, U_{t-}\right) + \Phi_{dt}\left(0, c_t, U_{t-}\right) dt + \Phi_{\upsilon}\left(0, c_t, U_{t-}\right)\left(\mu_t + \rho_t^Q \cdot \Sigma_t - \mathcal{A}\left(U_{t-}, \Sigma_t, \rho_t^Q\right)\right) dt.$$

Given the normalization $U_{t-} = \Phi\left(0, c_t, U_{t-}\right)$, the preceding equation can be rearranged to

$$-\mu_t = f\left(c_t, U_{t-}\right) + \rho_t^Q \cdot \Sigma_t - \mathcal{A}(U_{t-}, \Sigma_t, \rho_t^Q), \quad \text{where} \quad f\left(c_t, U_{t-}\right) = \frac{\Phi_{dt}\left(0, c_t, U_{t-}\right)}{\Phi_{\upsilon}\left(0, c_t, U_{t-}\right)}. \tag{42}$$

In particular, if Φ is given by (38), we obtain

$$f\left(c, \upsilon\right) = \beta \frac{u\left(c\right) - u\left(\upsilon\right)}{u'\left(\upsilon\right)}. \tag{43}$$

Combining (39) and (42) , we have transformed the heuristic recursive specification (37) of the utility process U to the backward stochastic differential equation (BSDE):

$$dU_t = -\left(f\left(c_t, U_t\right) + \rho_t^Q \cdot \Sigma_t - \mathcal{A}(U_{t-}, \Sigma_t, \rho_t^Q)\right) dt + \Sigma_t' dB_t, \quad U_T = c_T. \tag{44}$$

Given the terminal value U_T, a solution to the BSDE consists of an adapted pair (U, Σ) such that (44) holds. We also refer to the process U as a solution to the BSDE if (U, Σ) is a BSDE solution for some Σ. The fixed-point nature of BSDE (44) means that special restrictions must be imposed on the primitives to guarantee the existence and uniqueness of a solution. BSDE existence and uniqueness results were obtained by Pardoux and Peng (1990), Duffie and Epstein (1992), Barles, Buckdahn, and Pardoux (1997), Pardoux (1997), Pardoux, Pradeilles, and Rao (1997), and many others since, albeit, under assumptions that are violated in common homothetic applications, including the continuous-time version of the widely used parametrization of Epstein and Zin (1989). Existence and uniqueness results on the latter were developed by Duffie and Lions (1996) and Schroder and Skiadas (1999), results that still require generalization, for example, to include jumps.

4.3 Smooth Ambiguity Aversion in Continuous Time

A general form of continuous-time recursive utility has been expressed as BSDE (44), with the function $\mathcal{A}$ defined heuristically in the CE expression (41). This section specifies the function $\mathcal{A}$ for smooth conditional CEs corresponding to second-order expected utility and divergence preferences, and characterizes the instances in which the resulting specification is equivalent to one with an expected utility conditional CE. The claimed expressions are based on formal applications of Ito's lemma, as explained in Appendix A, although the irrelevance of the function φ for a smooth second-order EU CE will be seen to be even simpler than that. This section's CE expressions are analogous to corresponding expressions in Theorems 4 and 5, applied conditionally on time t information, with the infinitesimal time interval dt in place of h, and dB_t in place of B.

4.3.1 Expected Utility CE

We first establish the continuous-time version of the recursive utility of Kreps and Porteus (1978) in what amounts to a variant of the argument of Duffie and Epstein (1992), extended to include Poisson jumps. We assume that

$$\upsilon_t (U_{t+dt}) = u^{-1} \mathbb{E}_{t-}^{Q} u (U_{t+dt}), \tag{45}$$

for some $Q \in \Pi$ and $u \in C^2_{\text{vNM}}$ (or just C^1_{vNM} if there is no Brownian risk). Recall that the risk-aversion functions a^u and A^u are defined in (19) and (26), respectively. The first parts of Theorems 4 and 5 suggest that the CE expression (41), and therefore BSDE (44), is satisfied with

$$\mathcal{A}(U_{t-}, \Sigma_t, \rho_t) = \frac{a^u (U_{t-})}{2} \sum_{i=1}^{k} \left(\Sigma_t^i\right)^2 + \sum_{i=k+1}^{d} A^u \left(U_{t-}, \Sigma_t^i\right) \left(1 + \rho_t^i\right). \tag{46}$$

A confirmation of this claim is given in section A.1 as an application of Ito's formula.

4.3.2 Second-Order Expected-Utility CE

In the second parts of Theorems 4 and 5, it was shown that a smooth second-order EU CE over small Brownian or Poissonian risks is approximately equal to an EU CE with the compound prior. The relationship should be exact in the continuous time limit, meaning that the utility BSDE for a smooth second-order EU CE should be specified as in the last subsection, with Q being the compound prior. There are two parts to this claim: the function φ does not appear in the utility BSDE, and the prior Q is the compound prior. Let us now clarify these two parts.

The irrelevance of the function φ can be made in greater generality than the second-order EU formulation, as follows. Suppose that for some $\varphi \in C^1_{\text{vNM}}$,

$$\upsilon_t (U_{t+dt}) = \varphi^{-1} \left(\sum\nolimits_{s=1}^{S} \varphi \left(\upsilon_t^s (U_{t+dt})\right) \pi_t^s \right), \tag{47}$$

where $\pi^1, \ldots, \pi^S$ are $(0,1)$-valued predictable processes such that $\sum_s \pi^s = \mathbf{1}$, and $\upsilon^1, \ldots, \upsilon^S$ are conditional CEs that can be expressed as

$$\upsilon_t^s (U_{t+dt}) = U_{t-} + (\mu_t + \mathcal{D}_t^s (U_{t-}, \Sigma_t))\, dt, \quad s = 1, \ldots, S. \tag{48}$$

(As always, μ and Σ refer to the dynamics (39) of U.) A first-order Taylor expansion gives

$$\varphi (\upsilon_t^s (U_{t+dt})) = \varphi (U_{t-}) + \varphi' (U_-) (\mu_t + \mathcal{D}_t^s (U_{t-}, \Sigma_t))\, dt.$$

Multiplying the last equation by π_t^s, adding up over s, applying φ^{-1} on both sides and taking another first-order expansion results in

$$\upsilon_t (U_{t+dt}) = U_{t-} + (\mu_t + \mathcal{D}_t (U_{t-}, \Sigma_t))\, dt, \quad \text{where} \quad \mathcal{D}_t = \sum_{s=1}^{S} \mathcal{D}_t^s \pi_t^s. \tag{49}$$

A corresponding BSDE can be established as in Section 4.2. *The function φ is not part of equation* (49) *or the corresponding BSDE and is therefore irrelevant.*

The second-order EU CE definition specializes the preceding formulation by postulating priors $Q^1, \ldots, Q^S$ and a $u \in C^2_{\text{vNM}}$ such that

$$\upsilon_t^s (U_{t+dt}) = u^{-1} \mathbb{E}_{t-}^{Q^s} u (U_{t+dt}), \quad s = 1, \ldots, S,$$

and by requiring that the weights π_t^s are updated using Bayes' rule:

$$\pi_t^s = \frac{\xi_{t-}^s \pi_0^s}{\sum_{i=1}^{S} \xi_{t-}^i \pi_0^i}, \quad \text{where} \quad \xi_t^s = \mathbb{E}_t \left[\frac{dQ^s}{dP} \right].$$

By the EU CE analysis of the last subsection, equations (48) are satisfied with

$$\mathcal{D}_t^s (U_{t-}, \Sigma_t) = \rho_t^{Q^s} \cdot \Sigma_t - \mathcal{A}(U_{t-}, \Sigma_t, \rho_t^{Q^s}), \quad s = 1, \ldots, S, \tag{50}$$

where the function $\mathcal{A}$ is defined in equation (46). The compound prior is defined by

$$Q = \sum_{s=1}^{S} Q^s \pi_0^s \quad \text{and therefore} \quad \xi^Q = \sum_{s=1}^{S} \xi^s \pi_0^s. \tag{51}$$

In addition to the irrelevance of φ, we can now claim that υ *is an EU CE with prior Q and vNM index u*. To verify this claim, we first note that the second equation in (51) together with the Bayes formula defining π_t^s results in

$$\frac{d\xi_t^Q}{\xi_{t-}^Q} = \sum_{s=1}^{S} \frac{d\xi_t^s}{\xi_{t-}^s} \pi_t^s \quad \text{and therefore} \quad \rho_t^Q = \sum_{s=1}^{S} \rho_t^{Q^s} \pi_t^s.$$

Since in equation (50) the dependence of $\mathcal{A}(U_{t-}, \Sigma_t, \rho_t^{Q^s})$ on ρ^{Q^s} is linear (as stated in (46)), we conclude that

$$\mathcal{D}_t (U_{t-}, \Sigma_t) = \sum_{s=1}^{S} \mathcal{D}_t^s (U_{t-}, \Sigma_t) \pi_t^s = \rho_t^Q \cdot \Sigma_t - \mathcal{A}(U_{t-}, \Sigma_t, \rho_t^Q).$$

Therefore, the CE expression (49) becomes

$$\upsilon_t (U_{t+dt}) = U_{t-} + \left(\mu_t + \rho_t^Q \cdot \Sigma_t - \mathcal{A}(U_t, \Sigma_t, \rho_t^Q) \right) dt = u^{-1} \mathbb{E}_{t-}^Q u (U_{t+dt}).$$

4.3.3 Divergence CE

Finally, we formulate the continuous-time version of the divergence conditional CE of Section 3.4, extended to include any finite number of Brownian and Poissonian risk sources.

We postulate a reference prior R. For each Brownian risk source $i \in \{1, \ldots, k\}$, the corresponding reference drift term is ρ^{Ri}, and for each Poissonian risk source $i \in \{k+1, \ldots, d\}$, the corresponding reference arrival rate is $1 + \rho^{Ri}$. For any other prior Q, let the positive R-martingale $\xi^{Q/R}$ and the predictable process $\rho^{Q/R}$ be defined by[10]

$$\xi_t^{Q/R} = \mathbb{E}_t^R \left[\frac{dQ}{dR} \right] \quad \text{and} \quad \frac{d\xi_t^{Q/R}}{\xi_{t-}^{Q/R}} = \rho_t^{Q/R\prime} dB_t^R.$$

The other primitives needed to define the conditional CE are $u \in C^2_{\text{vNM}}$ and $\varphi \in C^2_{\text{div}}$ (see section 2.3). If there is no Brownian risk, it is sufficient to assume that $u \in C^1_{\text{vNM}}$ and $\varphi \in C^1_{\text{div}}$.

The continuous-time counterpart of the CE formulation of Section 2.3 is

$$\upsilon_t (U_{t+dt}) = \inf_{Q \in \Pi} u^{-1} \left(\mathbb{E}_{t-}^Q u (U_{t+dt}) + \mathbb{E}_{t-}^R \varphi \left(\frac{\xi_{t+dt}^{Q/R}}{\xi_{t-}^{Q/R}} \right) \right). \tag{52}$$

The interior-solution condition

$$u \left(U_{t-} + \Sigma_t^i \right) - u (U_{t-}) < -\varphi' (0+), \quad i = k+1, \ldots, d, \tag{53}$$

is assumed throughout. Of course the condition is automatically satisfied if there are no Poisson terms $(k = d)$.

Let us recall the notation

$$\theta = \varphi'' (1), \quad \psi (u) = \theta \left(1 - \exp \left(-\frac{u}{\theta} \right) \right), \quad \zeta (x) = \min_{y \in (0,\infty)} \{ xy + \varphi (y) \}, \tag{54}$$

as well as the definition of the function A_ζ^u in equation (28). The discussion of Section 3.4, and expression (32) in particular, suggests that the conditional CE expression (41) and corresponding utility BSDE (44) hold with

$$\mathcal{A}(U_{t-}, \Sigma_t, \rho_t) = \frac{a^{\psi \circ u} (U_{t-})}{2} \sum_{i=1}^{k} \left(\Sigma_t^i \right)^2 + \sum_{i=k+1}^{d} A_\zeta^u \left(U_{t-}, \Sigma_t^i \right) \left(1 + \rho_t^i \right). \tag{55}$$

This is confirmed in Section A.2 as a formal application of Ito's lemma, along with the following

[10] The change-of-measure formula for conditional expectations and the integration-by-parts formula for semimartingales can be used to show the relationships:

$$\xi^{Q/R} = \frac{\xi^Q}{\xi^R}; \quad \rho^{Q/Ri} = \rho^{Qi} - \rho^{Ri}, \quad i = 1, \ldots, k; \quad 1 + \rho_t^{Q/Ri} = \frac{1 + \rho^{Qi}}{1 + \rho^{Ri}}, \quad i = k+1, \ldots, d.$$

claims:

- If there is no Brownian risk $(k = d)$, then the above divergence utility specification is equivalent to one with an expected-utility CE, and is therefore within the Duffie-Epstein class of Section 4.3.1.
- If there is Poissonian risk ($k < d$), then the above divergence utility specification is equivalent to one with an expected-utility CE if and only if the divergence is of the entropic type: $\varphi(y) = \theta(y \log y - y + 1)$, in which case $\zeta = \psi$ and $A^u_\zeta = A^{\psi \circ u}$.

The first claim follows immediately by comparing (55) to (46), while the second claim requires a considerably more involved argument that is the topic of Section B.3.

Example 6 (Quadratic Divergence) *Suppose u is the identity function and the CE is of the divergence type as specified above with* $\varphi(y) = \theta(y-1)^2/2$ *and therefore* $\zeta(x) = x - x^2/(2\theta)$. *Assuming the validity of the interior-solution condition* (53)*, which in this case states that* $\Sigma^i_t < \theta$ *for* $i > k$, *expression* (55) *reduces to*

$$\mathcal{A}(U_{t-}, \Sigma_t, \rho_t) = \frac{1}{2\theta}\left(\sum_{i=1}^{k}\left(\Sigma^i_t\right)^2 + \sum_{i=k+1}^{d}\left(\Sigma^i_t\right)^2\left(1+\rho^i_t\right)\right).$$

The pricing and portfolio theory in Schroder and Skiadas (2008) shows the tractability advantages of this specification in the presence of Poisson jumps.

5 Concluding Remarks

The paper's conclusions are dependent on the type of the underlying information structure. Relative to a lump-sum resolution of uncertainty at a deterministic or predictable date, second-order expected utility and divergence preferences are in general quantitatively distinguishable from expected utility and each other. The paper's results only concern Brownian and Poissonian risk sources. Smooth divergence preferences are quantitatively similar to expected utility with respect to Brownian risks, but not Poissonian risks (except in the entropic case), and smooth second-order expected utility is quantitatively similar to expected utility with respect to both Brownian and Poissonian risks. The latter conclusion is potentially problematic for papers, such as Ju and Miao (2009) and Collard, Mukerji, Sheppard, and Tallon (2011), that seek to attribute risk premia to ambiguity aversion represented by smooth second-order expected utility. On the other hand, the dependence of the quantitative effect of uncertainty aversion on the type of the risk source can be a useful modeling tool, as it leads to source-dependent risk premia.

It is worth emphasizing that this paper's arguments hinge on the assumption that the function φ, representing ambiguity aversion in the CE specification, remains the same as the frequency is taken to infinity, and the CE specification is not modified based on the nature of the underlying risk source. These assumptions are analogous to the commonly adopted conventions that the risk aversion parameter for Epstein-Zin-Weil utility and more generally the von Neumann-Morgenstern index u of Kreps-Porteus utility are taken to be fixed as one transitions from discrete time to the Duffie-Epstein continuous-time limit and across stochastic settings. (For example, discrete-time recursive utility with a second-order expected utility CE reduces to Kreps-Porteus utility exactly when $\varphi = u$.) The general idea is that φ should capture ambiguity aversion as a fixed aspect of preferences that applies in every uncertainty environment, just as u captures risk aversion in the Kreps-Porteus specification. Recall (from Section 4.2) that we have normalized utility to correspond essentially to the payment rate of an equivalent perpetuity.[11] This device allows us to think of the continuation utility one period ahead as a state-contingent perpetuity that can be embedded in the same static setting, no matter what the frequency, thus anchoring risk aversion and ambiguity aversion.

In response to an earlier version of this paper, Hansen and Sargent (2009) verified the irrelevance of a fixed φ in a specific continuous-time context and proposed an alternative frequency-dependent parameterization of φ, finely tuned to preserve the effect of ambiguity aversion in the Brownian limit of recursive utility with a second-order expected utility CE. While in the present paper ambiguity aversion is kept constant as the frequency increases, the Hansen-Sargent parameterization implies that ambiguity aversion goes to infinity as the frequency goes to infinity, thus resulting in a different continuous-time limit, which as the authors point out, is not smooth in the level of the continuation utility. In other words, the limiting conditional CE over infinitesimal Brownian risks implied by the Hansen-Sargent formulation is a very different CE type than the smooth second-order expected utility CE that has been analyzed in this paper. Yet another modification of second-order expected utility motivated by an early version of the present paper was proposed by Gindrat and Lefoll (2010). The fact remains, however, that the simplest and most direct continuous-time interpretation of a smooth CE derived from a static model of second-order expected utility or divergence preferences is not quantitatively distinguishable from expected utility when small Brownian risks are involved (as well as small Poissonian risks in the case of second-order expected utility).

[11] Over a finite horizon, a unit perpetuity is represented as a unit annuity followed by a lump-sum payment that can be thought of as a claim in a unit perpetuity. Alternatively, one can talk of equivalent annuities. As the time-period goes to zero, the difference in duration of these annuities from the perspective of a fixed date becomes negligible.

A Appendix: Continuous-Time CE Expressions

This appendix verifies claims made in Sections 4.3.1 and 4.3.3 using a formal application of Ito's rule, as stated below. A more complete discussion of Ito's lemma in the context of Lévy processes is given by Applebaum (2004).

Consider any process of the form

$$dX_t = \mu_t^Q dt + \sigma_t' dB_t^Q,$$

where σ is a volatility process, and μ^Q is a *drift* process, meaning that it is predictable and satisfies $P\left[\int_0^t \left|\mu_u^Q\right| du < \infty\right] = 1$. For any twice continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$, Ito's rule states that

$$df(X_t) = \alpha_t^Q dt + \beta_t' dB_t^Q. \tag{56}$$

where

$$\begin{aligned} \alpha_t^Q &= f'(X_t)\mu_t^Q + \frac{f''(X_t)}{2}\sum_{i=1}^k \left(\sigma_t^i\right)^2 + \sum_{i=k+1}^d \left(f\left(X_{t-} + \sigma_t^i\right) - f(X_{t-}) - f'(X_{t-})\sigma_t^i\right)\left(1 + \rho_t^{Qi}\right), \\ \beta_t^i &= f'(X_{t-})\sigma_t^i, \quad i = 1, \ldots, k; \quad \beta_t^i = f\left(X_{t-} + \sigma_t^i\right) - f(X_{t-}), \quad i = k+1, \ldots, d. \end{aligned}$$

Since B^Q is a local martingale under Q, we have the heuristic expression

$$\mathbb{E}_{t-}^Q f(X_{t+dt}) = f(X_{t-}) + \alpha_t^Q dt, \tag{57}$$

which we utilize below in transforming heuristic conditional CE expressions to corresponding expressions for $\mathcal{A}$ in BSDE (44).

A.1 Expected Utility Conditional CE

Suppose the conditional CE υ_t takes the expected-utility form (45), for some $Q \in \Pi$ and $u \in C_{\text{vNM}}^2$ (or just C_{vNM}^1 if there is no Brownian risk). Applying Ito's lemma in the way it was applied in (57), we obtain

$$\mathbb{E}_{t-}^Q u(U_{t+dt}) = u(U_{t-}) + u'(U_{t-})\left(\mu_t + \rho_t^Q \cdot \Sigma_t - \mathcal{A}(U_{t-}, \Sigma_t, \rho_t^Q)\right) dt, \tag{58}$$

where $\mathcal{A}$ is given by (46). Further applying u^{-1} on both sides of equation (58) and using Ito's lemma (in a trivial sense) results in

$$u^{-1}\mathbb{E}_{t-}^Q u(U_{t+dt}) = U_{t-} + \left(\mu_t + \rho_t^Q \cdot \Sigma_t - \mathcal{A}\left(U_{t-}, \Sigma_t, \rho_t^Q\right)\right) dt.$$

A.2 Divergence Conditional CE

Adopting the setting and notation of Section 4.3.3, we now confirm the claims made in that section. Analogously to the derivation of equation (58), Ito's lemma implies that

$$\mathbb{E}_{t-}^R \varphi\left(1 + \rho_t^{Q/R\prime} dB_t^R\right) = \sum_{i=1}^k \frac{\theta}{2}\left(\rho^{Qi} - \rho^{Ri}\right)^2 dt + \sum_{i=k+1}^d \varphi\left(\frac{1+\rho^{Qi}}{1+\rho^{Ri}}\right)\left(1 + \rho_t^{Ri}\right) dt.$$

Combining the preceding expression with (58) results in

$$u^{-1}\left(\mathbb{E}_{t-}^{Q}u\left(U_{t+dt}\right)+\mathbb{E}_{t-}^{R}\varphi\left(\frac{\xi_{t+dt}^{Q/R}}{\xi_{t-}^{Q/R}}\right)\right) = U_{t-}+\mu_t dt + \sum_{i=1}^{k}\mathcal{C}_t\left(\rho_t^{Qi}\right)dt+\sum_{i=k+1}^{d}\mathcal{J}_t\left(\rho_t^{Qi}\right)dt, \tag{59}$$

where

$$\begin{aligned}\mathcal{C}_t\left(\rho_t^i\right) &= \rho_t^i\Sigma_t^i-\frac{a^u\left(U_{t-}\right)}{2}\left(\Sigma_t^i\right)^2+\frac{\theta}{2u'\left(U_{t-}\right)}\left(\rho_t^i-\rho_t^{Ri}\right)^2,\\ \mathcal{J}_t\left(\rho_t^i\right) &= \rho_t^i\Sigma_t^i-A^u\left(U_{t-},\Sigma_t^i\right)\left(1+\rho_t^i\right)+\frac{1}{u'\left(U_{t-}\right)}\varphi\left(\frac{1+\rho_t^i}{1+\rho_t^{Ri}}\right)\left(1+\rho_t^{Ri}\right).\end{aligned}$$

The last two terms are minimized separately, noting that $\mathcal{C}_t$ is quadratic and $\mathcal{J}_t$ is strictly convex. The assumed inequality (53) is equivalent to the condition $\mathcal{J}_t'\left(-1+\varepsilon\right)<0$ for some sufficiently small $\varepsilon>0$, which is necessary and sufficient for $\mathcal{J}_t$ to be minimized by some ρ^i such that $1+\rho^i$ is strictly positive. It follows that the right-hand side of (59) is minimized by the value ρ^Q, where the Brownian terms are given by

$$\rho_t^{Qi}=\rho_t^{Ri}-\frac{u'\left(U_{t-}\right)}{\theta}\Sigma_t^i,\quad i=1,\ldots,k,$$

and the Poissonian terms are given by

$$1+\rho_t^{Qi}=\left(1+\rho_t^{Ri}\right)\varphi'^{-1}\left(u\left(U_{t-}\right)-u\left(U_{t-}+\Sigma_t^i\right)\right),\quad i=k+1,\ldots,d.$$

Substituting the minimizing value of ρ^Q in (59) results in

$$\upsilon_t\left(U_{t+dt}\right)=U_{t-}+\left(\mu_t+\rho_t^R\cdot\Sigma_t-\frac{a^{\psi\circ u}\left(U_{t-}\right)}{2}\sum_{i=1}^{k}\left(\Sigma_t^i\right)^2-\sum_{i=k+1}^{d}A_\zeta^u\left(U_{t-},\Sigma_t^i\right)\left(1+\rho_t^{Ri}\right)\right)dt, \tag{60}$$

where ψ and ζ are defined in (54). Expression (60) is the same as (41) with $Q=R$ and the function $\mathcal{A}$ given by equation (55), as claimed.

If there is no Poissonian risk $(k=d)$, then expression (60) reduces to an EU CE with prior R and vNM index $\psi\circ u$, and therefore the corresponding recursive utility is within the class of continuous-time Kreps-Porteus utilities.

Finally, suppose there is Poissonian risk $(0\leq k<d)$. The question is: Can the divergence CE (60) be expressed as an EU CE relative to some prior and smooth vNM index? We argue that there exists some $W\in\Pi$ and $w\in C_{\text{vNM}}^2$ such that the CE (60) takes the expected utility form

$$\upsilon_t\left(U_{t+dt}\right)=U_{t-}+\left(\mu_t+\rho_t^W\cdot\Sigma_t-\frac{a^w\left(U_{t-}\right)}{2}\sum_{i=1}^{k}\left(\Sigma_t^i\right)^2-\sum_{i=k+1}^{d}A^w\left(U_{t-},\Sigma_t^i\right)\left(1+\rho_t^{Wi}\right)\right)dt \tag{61}$$

for all values of (U_{t-}, Σ_t) satisfying condition (53) if and only if

$$W = R \quad \text{and} \quad \varphi(x) = \theta(x \log x + x - 1) \quad (\text{where } \theta = \varphi''(1)). \tag{62}$$

The "if" part is immediate, since (62) implies that $\zeta = \psi$ and $A^u_\zeta = A^{\psi \circ u}$. Conversely, suppose that for some prior W and $w \in C^2_{\text{vNM}}$, equation (61) is true for all values of (U_{t-}, Σ_t) satisfying (53), that is, for all values of (U_{t-}, Σ_t) such that $\left(U_{t-}, \Sigma^i_t\right) \in D$ for every Poissonian factor i, where D is defined in (27). Isolating any such factor $i \in \{k+1, \ldots, d\}$, the equality of the conditional CEs (60) and (61) implies that

$$\rho^{Ri}_t \Sigma^i_t - A^u_\zeta\left(U_{t-}, \Sigma^i_t\right)\left(1 + \rho^{Ri}_t\right) = \rho^{Wi}_t \Sigma^i_t - A^w\left(U_{t-}, \Sigma^i_t\right)\left(1 + \rho^{Wi}_t\right),$$

for all values of $\left(U_{t-}, \Sigma^i_t\right)$ in D. An application of Lemma 7 of Section B.3 shows that then (62) must hold. This argument is necessarily heuristic, since the conditional CE expressions are heuristic to start with. A fully rigorous version of the same ideas is expressed by Theorem 5(c), as is evident from the theorem's proof.

B Appendix: Proofs

B.1 Proof of Proposition 1

Let $\Omega = \{0, 1, \ldots, n\}$, where $U(0) = U_{\max} \equiv \max_\omega U(\omega)$. A probability $Q \in \Pi$ is identified with a vector $q \in \mathbb{R}^n_{++}$, where $Q(\omega) = q_\omega$ for $\omega = 1, \ldots, n$ and $Q(0) = 1 - \sum_{\omega=1}^n q_\omega$. In particular, P is identified with the vector $p \in \mathbb{R}^n_{++}$. Then $V(q) \equiv \mathbb{E}^Q u(U) + \mathbb{E}\varphi(dQ/dP)$ can be written as

$$V(q) = u(U_{\max}) + \sum_{\omega=1}^n \left\{ q_\omega\left(u(U_\omega) - u(U_{\max})\right) + p_\omega \varphi\left(\frac{q_\omega}{p_\omega}\right)\right\} + p_0 \varphi\left(\frac{1 - \sum_{\omega=1}^n q_\omega}{p_0}\right).$$

The strictly convex function V is minimized at $q \in \mathbb{R}^n_{++}$ if and only if $\partial V(q)/\partial q_\omega = 0$ for every $\omega \in \{1, \ldots, n\}$, a condition that is easily shown to be equivalent to

$$q_\omega = p_\omega \varphi'^{-1}(\alpha - u(U_\omega)), \quad \omega = 1, \ldots, n, \quad \text{where} \quad \alpha = u(U_{\max}) + \varphi'\left(\frac{1 - \sum_{\omega=1}^n q_\omega}{p_0}\right). \tag{63}$$

The last term is greater than $\varphi'(0+)$, and $\varphi'^{-1}\left(\varphi'(0+) + u(U_{\max}) - u(U(0))\right) = 0$. Therefore, adding up over ω shows that condition (63) implies condition (6), and therefore the latter is necessary for the existence of a minimizing (strictly positive) prior.

Conversely, suppose that condition (6) is satisfied and define the function

$$f(\alpha) = \mathbb{E}\varphi'^{-1}(\alpha - u(U)), \quad \alpha > \varphi'(0+) + u(U_{\max}).$$

Clearly, f is strictly increasing and continuous. Because of condition (6), $f(\alpha)$ takes values below one as α approaches $\varphi'(0+) + u(U_{\max})$. On the other hand, by the definition of a divergence index,

$\phi'(\infty) = \infty$, and therefore $f(\alpha)$ takes values greater than one for sufficiently large α. This proves the existence of a unique value of $\alpha > \varphi'(0+) + u(U_{\max})$ such that $f(\alpha) = 1$, which is exactly condition (7). By construction, $\alpha - u(U) > \varphi'(0+)$ and therefore the vector q is well-defined by the first equation of condition (63). Finally, the definition of q and the identities $f(\alpha) = 1$ and $U_{\max} = U(0)$ can easily be shown to imply the last equation of condition (63). This proves condition (63), which we have seen is sufficient for the optimality of the prior defined by q.

B.2 Proof of Theorems 4 and 5

We begin with some preliminaries to be used in the proof of both theorems.

B.2.1 Taylor Approximations

For easy reference, we quickly review[12] the Taylor approximation error estimates we will be using. Consider any positive integer n (we'll only need $n = 1$ or 2), any interval $[-\varepsilon, \varepsilon] \subseteq \mathbb{R}$, where $\varepsilon > 0$, and any n times continuously differentiable real-valued function f on an open interval that includes $[-\varepsilon, \varepsilon]$. Let $f^{(i)}$ denote the i^{th} derivative of f, with $f^{(0)} = f$. Let also f_n be the n-degree polynomial such that $f^{(i)}(0) = f_n^{(i)}(0)$ for $i = 0, \ldots, n$. (We will only need $f_1(x) = f(0) + f'(0)x$ and $f_2(x) = f_1(x) + f''(0)x^2/2$.) Then

$$f = f_n + R_n, \quad \text{where } R_n(x) = \int_0^x \left(f^{(n)}(t) - f^{(n)}(0)\right) \frac{(x-t)^{n-1}}{(n-1)!} dt. \tag{64}$$

(Proof: Integration by parts implies $R_{n-1}(x) = f^{(n)}(0)x^n/n! + R_n(x)$. Now use induction in n.) It follows that there exists a continuous function $r_n : [-\varepsilon, \varepsilon] \to [0, \infty)$ such that[13]

$$r_n(0) = 0 \quad \text{and} \quad |R_n(x)| \leq r_n(x)x^n \text{ for all } x \in [-\varepsilon, \varepsilon]. \tag{65}$$

B.2.2 Normalization

We introduce a normalization that entails no loss of generality and simplifies notation.

Suppose u and $\tilde{u}$ are two vNM indices such that $\tilde{u} = \alpha u + \beta$ for some $\alpha \in (0, \infty)$ and $\beta \in \mathbb{R}$. Then $u^{-1}\mathbb{E}^Q u = \tilde{u}^{-1}\mathbb{E}^Q \tilde{u}$ for any prior Q, $a^u = a^{\tilde{u}}$ and $A^u = A^{\tilde{u}}$. Moreover, if $\tilde{\varphi} = \alpha\varphi$, the pair (u, φ) defines the same divergence CE as the pair $(\tilde{u}, \tilde{\varphi})$, and $A^u_\zeta = A^{\tilde{u}}_{\tilde{\zeta}}$, where $\tilde{\zeta}(x) = \min_{y>0}\{xy + \tilde{\varphi}(y)\}$.

[12] This material is of course standard and can be found, for example, in Apostol, Calculus, 2nd ed., Wiley 1967.

[13] Note that under the stronger assumption that f is $n+1$ times differentiable, integration by parts allows us to restate the expression for $R_n(x)$ in (64) in the more familiar form

$$R_n(x) = \int_0^x f^{(n+1)}(t) \frac{(x-t)^n}{n!} dt.$$

If $f^{(n+1)}$ is bounded on $[-\varepsilon, \varepsilon]$ (for example, if it is continuous), then there exists a constant K such that $|R_n(x)| \leq Kx^{n+1}$ for all $x \in [-\varepsilon, \varepsilon]$. In other words, we can set $r_n(x) = Kx$ in (65).

It is then clear that to prove Theorems 4 and 5, it is sufficient to prove them with $u(\cdot)$ replaced with $(u(\cdot) - u(U_0))/u'(U_0)$. There is, therefore, no loss of generality in assuming the normalization

$$u(U_0) = 0 \quad \text{and} \quad u'(U_0) = 1, \tag{66}$$

which is assumed throughout the remainder of this appendix.

B.2.3 Proof of Theorem 4

Define

$$\Delta(h) = \mu h + \Sigma B = (\mu + \rho\Sigma) h + \Sigma B^Q, \tag{67}$$

and recall that $U = U_0 + \Delta(h)$. Let $\varepsilon \in (0,1)$ be such that $U_0 - \varepsilon > \ell$. Throughout this proof, we assume that $h \in (0, \bar{h})$, where $\bar{h} > 0$ is small enough so that $h \in (0, \bar{h})$ implies $\Delta(h) \in (-\varepsilon, +\varepsilon)$. Note that $\varepsilon, \bar{h}$ and $\Delta(h)$ do not depend on ρ.

(a+) We prove part (a) as well as an error bound that will be used in part (c). Recall that the normalization (66) is assumed.

Applying the Taylor approximation (64) with $f(x) = u(U_0 + x)$ and $n = 2$, we have

$$u(U) = \Delta(h) - \frac{1}{2} a^u(U_0) \Delta(h)^2 + R_2(\Delta(h)). \tag{68}$$

Recall that $\mathbb{E}^Q B^Q = 0$ and $\mathbb{E}^Q \left[(B^Q)^2\right] = h - (\rho h)^2$ by (15). Using these facts and the second expression for $\Delta(h)$ in (67), we compute

$$\mathbb{E}^Q \Delta(h) = (\mu + \rho\Sigma) h \quad \text{and} \quad \mathbb{E}^Q[\Delta(h)^2] = \Sigma^2 h + \left(\mu^2 + 2\mu\Sigma\rho\right) h^2. \tag{69}$$

Therefore, applying $\mathbb{E}^Q$ to both sides of (68), we can write

$$\mathbb{E}^Q u(U) = \left(\mu + \rho\Sigma - \frac{1}{2} a^u(U_0) \Sigma^2\right) h + \delta(\rho, h) h, \tag{70}$$

where

$$\delta(\rho, h) = -\frac{1}{2} a^u(U_0) \left(\mu^2 + 2\mu\Sigma\rho\right) h + \frac{1}{h} \mathbb{E}^Q R_2(\Delta(h)). \tag{71}$$

We apply the bound (65) to the last term. Since the possible values of $\Delta(h)$ are $\mu h \pm \Sigma\sqrt{h}$,

$$|r_2(\Delta(h))| \leq b(h) \equiv \max\left\{\left|r_2\left(\mu h + \Sigma\sqrt{h}\right)\right|, \left|r_2\left(\mu h - \Sigma\sqrt{h}\right)\right|\right\}, \tag{72}$$

resulting in the bound $\mathbb{E}^Q |R_2(\Delta(h))| \leq b(h) \mathbb{E}^Q[\Delta(h)^2]$. Using this bound and (69) in (71), we obtain

$$\delta(\rho, h) \leq \left|\mu^2 + (2\mu\Sigma)\rho\right| \left(\frac{a^u(U_0)}{2} + b(h)\right) h + \Sigma^2 b(h). \tag{73}$$

Note that b is continuous (and therefore bounded), vanishes at zero, and does not depend on ρ. What is important for our purposes is the fact that there exist constants K_0 and K_1 such that

$$\delta(\rho, h) \leq (K_0 + K_1 \rho) h + \Sigma^2 b(h), \tag{74}$$

for all $h \in \left(0, \bar{h}\right)$ and any choice of ρ (that is consistent with the positivity of Q).

The proof of part (a) is now easily completed by a first-order Taylor series expansion of u^{-1} around zero, using (70). (Note that, by the normalization (66), $u^{-1}(0) = U_0$ and $u^{-1\prime}(0) = 1$.)

(b) Equation (20) and a first-order Taylor series expansion of φ around U_0 imply that

$$\varphi(\upsilon^s(U)) = \varphi(U_0) + \varphi'(U_0)\left(\mu + \rho^s \Sigma - \frac{1}{2} a^u(U_0) \Sigma^2\right) h + o(h), \quad s \in \{1, \ldots, S\}.$$

Therefore,

$$\upsilon(U) = \varphi^{-1}\left(\varphi(U_0) + \varphi'(U_0)\left(\mu + \rho\Sigma - \frac{1}{2} a^u(U_0) \Sigma^2\right) h + o(h)\right), \quad \rho = \sum\nolimits_s \rho^s \pi^s.$$

Note that $\rho = \mathbb{E}^Q B/h$, where $Q = \sum_s Q^s \pi^s$. Taking a first-order Taylor approximation of φ^{-1} around $\varphi(U_0)$, and using part (a), we conclude:

$$\upsilon(U) = U_0 + \left(\mu + \rho\Sigma - \frac{1}{2} a^u(U_0) \Sigma^2\right) h + o(h) = u^{-1} \mathbb{E}^Q u(U) + o(h).$$

(c) The proof of this part is more elaborate because we have to justify the interchange of the operations of minimization and approximation.

It is straightforward to check that condition (6) is satisfied for all sufficiently small h. Select $\bar{h}$ so that, by Proposition 1, a minimizing prior exists for every $h \in \left(0, \bar{h}\right)$. Given any $Q \in \Pi$ and corresponding drift $\rho = h^{-1}\mathbb{E}^Q B$, let

$$\begin{aligned} V_h(\rho) &= \mathbb{E}^Q u(U) + \mathbb{E}\varphi(dQ/dP) \\ &= u\left(U_0 + \mu h - \Sigma\sqrt{h}\right) + \frac{1}{2}\left[\left(1 + \rho\sqrt{h}\right) H_h + \varphi\left(1 + \rho\sqrt{h}\right) + \varphi\left(1 - \rho\sqrt{h}\right)\right], \end{aligned} \tag{75}$$

where

$$H_h = u\left(U_0 + \mu h + \Sigma\sqrt{h}\right) - u\left(U_0 + \mu h - \Sigma\sqrt{h}\right).$$

Let ρ_h be the value of ρ that minimizes $V_h(\rho)$. Setting the derivative of V_h at ρ_h to zero, we find

$$H_h + \varphi'\left(1 + \rho_h\sqrt{h}\right) - \varphi'\left(1 - \rho_h\sqrt{h}\right) = 0. \tag{76}$$

Since $\lim_{h \downarrow 0} H_h = 0$, it follows that

$$\lim_{h \downarrow 0} \rho_h \sqrt{h} = 0. \tag{77}$$

We now feed this limit back into the optimality condition to prove that in fact ρ_h converges. Equation (76) can be rearranged to read

$$\rho_h = -\Sigma \frac{\left(u\left(U_0 + \mu h + \Sigma\sqrt{h}\right) - u\left(U_0 + \mu h - \Sigma\sqrt{h}\right)\right)/2\Sigma\sqrt{h}}{\left(\varphi'\left(1 + \rho_h\sqrt{h}\right) - \varphi'\left(1 - \rho_h\sqrt{h}\right)\right)/2\rho_h\sqrt{h}}.$$

As $h \downarrow 0$, the numerator converges to $u'(U_0)$, which we normalized to one, and the denominator converges to $\varphi''(1) = \theta$. Therefore, as $h \downarrow 0$, ρ_h converges to $\rho_0 \equiv -\Sigma/\theta$, which is the value of ρ that minimizes the quadratic

$$\mathcal{G}(\rho) \equiv \mu + \rho\Sigma - \frac{1}{2}a^u(U_0)\Sigma^2 + \frac{\theta}{2}\rho^2. \tag{78}$$

Summarizing,

$$V_h(\rho_h) = \min_\rho V_h(\rho), \quad \mathcal{G}(\rho_0) = \min_\rho \mathcal{G}(\rho), \quad \lim_{h\downarrow 0}\rho_h = \rho_0. \tag{79}$$

In the next stage of the proof, we bring in the analysis of part (a+) to approximate V_h. A suitable approximation for the first term, $E^Q u(U)$, of $V_h(\rho)$ was derived in part (a+). We now derive an analogous approximation for the second term, $\mathbb{E}\varphi(1+\rho B)$. Given any $\varepsilon \in (0,1)$, further reduce the value of $\bar{h}$ if necessary so $\rho\sqrt{h}, \rho_h\sqrt{h} \in (-\varepsilon, \varepsilon)$ for all $h \in \left(0, \bar{h}\right)$ and $\rho \in (\rho_0 - 1, \rho_0 + 1)$. In the remainder of this proof, we assume that every instance of the pair (ρ, h) lies in $(\rho_0 - 1, \rho_0 + 1) \times \left(0, \bar{h}\right)$, and therefore ρB is valued in $(-\varepsilon, \varepsilon)$. We approximate the term $\mathbb{E}\varphi(1+\rho B)$ in the definition of $V_h(\rho)$ by applying a second-order Taylor approximation, as in Section B.2.1 with $f(x) = \varphi(1+x)$ and $n = 2$. In fact, we can directly apply approximation (70) and associated error bound (73) by making the following substitutions: Replace Q with P, and therefore the ρ with zero; replace φ with u; and replace U_0 with 1, μ with zero, and Σ with ρ. Since $\varphi(1) = 1$, $\varphi'(1) = 0$ and $\varphi''(1) = \theta$, the result is

$$\left|\mathbb{E}\varphi(1+\rho B) - \frac{\theta}{2}\rho^2 h\right| \leq \rho^2 h c\left(\rho\sqrt{h}\right),$$

for a function $c : [-\varepsilon, \varepsilon] \to \mathbb{R}_+$ that is continuous and vanishes at zero. Combining this with approximation (70) with the simplified bound (74), we obtain

$$V_h(\rho) = \mathcal{G}(\rho)h + \chi(\rho, h)h, \tag{80}$$

where the continuous function χ satisfies the bound

$$|\chi(\rho, h)| \leq (K_0 + K_1\rho)h + \Sigma^2 b(h) + \rho^2 c\left(\rho\sqrt{h}\right).$$

Again, b and c are continuous and vanish at zero. Given that ρ_h converges, we note the key facts

$$\lim_{h\downarrow 0}\chi(\rho_0, h) = 0 \quad \text{and} \quad \lim_{h\downarrow 0}\chi(\rho_h, h) = 0. \tag{81}$$

On the other hand, (79) and (80) imply the string of inequalities

$$\mathcal{G}(\rho_0)h + \chi(\rho_0, h)h = V_h(\rho_0) \geq V_h(\rho_h) = \mathcal{G}(\rho_h)h + \chi(\rho_h, h)h \geq \mathcal{G}(\rho_0)h + \chi(\rho_h, h)h.$$

The last two displays prove

$$u(\upsilon(U)) = V_h(\rho_h) = \mathcal{G}(\rho_0)\, h + o(h).$$

Given the normalization (66), a first-order Taylor expansion of u^{-1} around zero yields the conclusion

$$\upsilon(U) = U_0 + \mathcal{G}(\rho_0)\, h + o(h), \quad \rho_0 = -\frac{\Sigma}{\theta}. \tag{82}$$

Computing the value $\mathcal{G}(\rho_0)$ and using the identity $a^{\psi \circ u} = a^u + \left(a^\psi \circ u\right)(u')$ with the normalization (66), we also have

$$\mathcal{G}(\rho_0) = \mu - \frac{1}{2}\left(a^u(U_0) + \frac{1}{\theta}\right)\Sigma^2 = \mu - \frac{1}{2} a^{\psi \circ u}(U_0)\,\Sigma^2.$$

Therefore, by part (a),

$$(\psi \circ u)^{-1}\,\mathbb{E}(\psi \circ u)(U) = U_0 + \mathcal{G}(\rho_0)\, h + o(h). \tag{83}$$

The combination of equations (82) and (83) completes the proof.

B.2.4 Proof of Theorem 5

This proof is analogous to that of Theorem 4, and therefore some details will be abbreviated. Recall that we are assuming the normalization (66) , which entails no loss of generality. Therefore, $A^u(U_0, \Sigma) = \Sigma - u(U_0 + \Sigma)$ and $A^u_\zeta(U_0, \Sigma) = \Sigma - \zeta(u(U_0 + \Sigma))$. As with Theorem 4, the discussion that follows applies for $h \in \left(0, \bar{h}\right)$, for sufficient small $\bar{h} > 0$.

(a+) We prove part (a) as well as an error bound that will be used in part (c). Let

$$\alpha_h = (\mu - \Sigma)\, h - \Sigma \varepsilon_h. \tag{84}$$

The distribution of U under Q is

$$U = \begin{cases} U_0 + \Sigma + \alpha_h, & \text{with } Q\text{-probability } (1+\rho)\, h + \varepsilon_h, \\ U_0 + \alpha_h, & \text{with } Q\text{-probability } 1 - (1+\rho)\, h - \varepsilon_h, \end{cases} \tag{85}$$

We compute the expectation $\mathbb{E}^Q u(U)$, using the first-order Taylor expansions

$$u(U_0 + \Sigma + \alpha_h) = u(U_0 + \Sigma) + u'(U_0 + \Sigma)\,\alpha_h + R^1(\alpha_h) \quad \text{and} \quad u(U_0 + \alpha_h) = \alpha_h + R^0(\alpha_h),$$

where $\left|R^i(\alpha)\right| \leq \left|r^i(\alpha)\right| |\alpha|$ for continuous functions r^i that vanish at zero (see Section B.2.1). Direct computation then shows that

$$\mathbb{E}^Q u(U) = [\mu + \rho\Sigma - A^u(U_0, \Sigma)(1+\rho)]\, h + (\delta_0(h) + \delta_1(h)(1+\rho))\, h,$$

where the functions δ_i do not depend on ρ, and satisfy $|\delta_i(h)| \le b_i(h)$ for some some continuous functions b_i that vanish at zero.

A first-order Taylor expansion of u^{-1} completes the proof of part (a).

(b) The proof of this part is essentially the same as that of part (b) of Theorem 4, replacing the risk-adjustment term $a^u(U_0)\Sigma^2/2$ with $A^u(U_0,\Sigma)(1+\rho^s)$, for each prior Q^s.

(c) An application of Proposition 1 proves the existence of a minimizing prior, for any $(U_0,\Sigma)\in D$. For any given prior Q, let $\rho = \mathbb{E}^Q B/h$ and define, using the notation in (84) and (85) :

$$\begin{aligned} V_h(\rho) &= \mathbb{E}^Q u(U) + \mathbb{E}\varphi(dQ/dP) \\ &= u(U_0+\alpha_h) + ((1+\rho)h+\varepsilon_h)H_h \\ &\quad + (h+\varepsilon_h)\varphi(1+\rho(1-h-\varepsilon_h)) + (1-h-\varepsilon_h)\varphi(1+\rho(-h-\varepsilon_h)), \end{aligned}$$

where $H_h = u(U_0+\Sigma+\alpha_h) - u(U_0+\alpha_h)$. Let ρ_h be the value of ρ that minimizes $V_h(\rho)$. Setting the derivative of V_h at ρ_h to zero, we find

$$H_h + \left(1+\frac{\varepsilon_h}{h}\right)(1-h-\varepsilon_h)G_h = 0, \tag{86}$$

where

$$G_h = \varphi'(1+\rho_h(1-h-\varepsilon_h)) - \varphi'(1-\rho_h(h+\varepsilon_h)). \tag{87}$$

As h goes to zero, H_h converges to $u(U_0+\Sigma)$ and ε_h/h converges to zero. From equation (86), it follows that

$$\lim_{h\downarrow 0} G_h = -u(U_0+\Sigma). \tag{88}$$

We show that ρ_h must then also converge. Since $1+\rho B$ is strictly positive, $1+\rho_h(1-h-\varepsilon_h)>0$, which implies that $\rho_h > -2$ for all $h\in(0,\bar{h})$, provided $\bar{h}$ is sufficiently small. Since φ' is increasing, the quantity $\varphi'(1-\rho_h(h+\varepsilon_h))$ in (87) is bounded above by a constant for all $h\in(0,\bar{h})$. Because of (88), it must then be the case that the first term in (87) is also bounded above by a constant for sufficiently small h, and therefore ρ_h must also be bounded above for all sufficiently small h. Since ρ_h is bounded both below and above as h goes to zero,

$$\lim_{h\downarrow 0}\rho_h(h+\varepsilon_h) = 0. \tag{89}$$

Finally, equation (87) can be rearranged to

$$\rho_h = -1 + \rho_h(h+\varepsilon_h) + \varphi'^{-1}(G_h + \varphi'(1-\rho_h(h+\varepsilon_h))),$$

which combined with (88) and (89) implies that

$$\lim_{h\downarrow 0}\rho_h = \rho_0 \equiv -1 + \varphi'^{-1}(-u(U_0+\Sigma)).$$

Note that ρ_0 is the value of ρ that minimizes

$$\mathcal{G}_t(\rho) \equiv \mu + \rho\Sigma - A^u(U_0, \Sigma)(1+\rho) + \varphi(1+\rho).$$

Summarizing, condition (79) of the proof of Theorem 4 holds in the current context, too.

In the next stage of the proof, we bring in the analysis of part (a+) to approximate V_h. A suitable approximation for the first term, $E^Q u(U)$, of $V_h(\rho)$ was derived in part (a+). The analogous computation for $\mathbb{E}\varphi(1+\rho B)$ yields an approximation of the form

$$|\mathbb{E}\varphi(1+\rho B) - \varphi(1+\rho)h| \leq \delta(\rho, h) h,$$

where $\delta(\rho_0, h)$ and $\delta(\rho_h, h)$ converge to zero as h goes to zero. Adding up the two approximations, results in expression (80) for a function χ such that the limits (81) hold. In the proof of Theorem 4(c), these conditions together with (79) were used to prove equation (82). The exact same argument in this context yields

$$\upsilon(U) = U_0 + \mathcal{G}(\rho_0) h + o(h), \quad 1+\rho_0 = \varphi'^{-1}(-u(U_0+\Sigma)).$$

Condition (30) follows from the identity

$$\mathcal{G}(\rho_0) = \mu - A^u_\zeta(U_0, \Sigma), \quad \text{where} \quad \zeta(x) = \min_{y\in(0,\infty)} \{xy + \varphi(y)\},$$

which can be shown easily using the definitions and the observation that for $x = u(U_0+\Sigma)$, the minimum value $\zeta(x)$ is achieved by $y = \varphi'^{-1}(-x) = 1+\rho_0$.

We still must show the last claim of part (c). By part (a), the CE approximation (31) is equivalent to

$$-A^u_\zeta(U_0, \Sigma) = \Sigma\rho^w - A^w(U_0, \Sigma)(1+\rho^w), \quad (U_0, \Sigma) \in D,$$

where $\rho^w = \mathbb{E}^W B/h$. The proof is completed by Lemma 7 of the following section.

B.3 Characterization of Entropic Divergence

The following lemma is used in the proof of Theorem 5 as well as Section A.2.

Lemma 7 *Suppose D is defined in (27), and ψ, θ and ζ are defined by (54) in terms of $\varphi \in C^1_{div}$. Then the following statements are equivalent, for any $w \in C^2_{vNM}$, $\Sigma \in \mathbb{R}$ and $\rho, \rho^w \in (-1, \infty)$.*

1. *$\Sigma\rho - A^u_\zeta(U_0, \Sigma)(1+\rho) = \Sigma\rho^w - A^w(U_0, \Sigma)(1+\rho^w)$ for all $(U_0, \Sigma) \in D$.*
2. *$\rho = \rho^w$ and $\varphi(y) = \theta(y\log y - y + 1)$ for all $y > 0$.*

Proof. We show the implication ($1 \implies 2$), since the converse is a matter of simple computation. We assume that

$$w'(1) = u'(1) \quad \text{and} \quad w(1) = u(1) = 0,$$

which entails no loss of generality since A^w is invariant to a positive affine transformation of w, and A^u_ζ is invariant to adding a constant to u.

Suppose the Lemma's condition 1 is true. Defining the function

$$f(x) = \frac{1+\rho}{1+\rho^w}\zeta(x),$$

the assumed condition is equivalent to

$$\frac{f(u(U_0+\Sigma)-u(U_0))}{u'(U_0)} = \frac{w(U_0+\Sigma)-w(U_0)}{w'(U_0)}, \tag{90}$$

for all $(U_0,\Sigma) \in D$. Letting $U_0 = 1$ and $z = 1+\Sigma$ it follows that if $u(z) < -\varphi'(0+)$,

$$f(u(z)) = w(z) \quad \text{and therefore } f'(u(z))\,u'(z) = w'(z).$$

Assuming

$$x = u(U_0) < -\varphi'(0+) \quad \text{and} \quad y = u(U_0+\Sigma) - u(U_0) < -\varphi'(0+),$$

condition (90) becomes

$$f'(x)\,f(y) = f(x+y) - f(x).$$

Differentiating with respect to y and taking logarithms results in

$$\log f'(x) + \log f'(y) = \log f'(x+y).$$

Since f' is continuous, there exists a scalar a such that

$$\log f'(x) = ax. \tag{91}$$

Since $\zeta(x) = \min_{y>0}(xy+\varphi(y))$, if $x = -\varphi'(y)$, then $\zeta(x) = xy+\varphi(y)$ and $\zeta'(x) = y$ (by the envelope theorem). Using this fact in identity (91) with $x = -\varphi'(y)$, we obtain

$$\log\left(\frac{1+\rho}{1+\rho^w}\right) + \log y = -a\varphi'(y), \quad y > 0.$$

Since $\varphi'(1) = 0$, it follows that $\rho^w = \rho$. Since $\varphi''(1) = \theta$, it follows that $a\theta = -1$. Therefore, φ solves the ODE

$$\varphi(1) = 0, \quad \varphi'(y) = \theta\log(y), \quad y > 0,$$

whose unique solution is $\varphi(y) = \theta(y\log y - y + 1)$. ■

References

APPLEBAUM, D. (2004): *Lévy Processes and Stochastic Calculus*. Cambridge University Press, Cambridge, U.K.

ARROW, K. J. (1965): *Aspects of the Theory of Risk Bearing*. Yrjo Jahnssonin Saatio, Helsinki.

——— (1970): *Essays in the Theory of Risk Bearing*. North Holland, London.

BARLES, G., R. BUCKDAHN, AND E. PARDOUX (1997): "Backward Stochastic Differential Equations and Integro-Partial Differential Equations," *Stochastics and Stochastics Reports*, 60, 57–83.

BARRO, R. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121, 823–866.

BILLINGSLEY, P. (1999): *Convergence of Probability Measures*. John Wiley & Sons, New York, second edn.

CHEN, Z., AND L. EPSTEIN (2002): "Ambiguity, Risk, and Asset Returns in Continuous Time," *Econometrica*, 70, 1403–1443.

CHEW, S. H., AND L. G. EPSTEIN (1989): "The Structure of Preferences and Attitudes Toward the Timing of the Resolution of Uncertainty," *International Economic Review*, 30, 103–117.

COLLARD, F., S. MUKERJI, K. SHEPPARD, AND J.-M. TALLON (2011): "Ambiguity and the Historical Equity Premium," working paper, dept. of Economics, University of Oxford, U.K.

DIXIT, A. K., AND R. S. PINDYCK (1994): *Investment under Uncertainty*. Princeton Univ. Press, Princeton, New Jersey.

DONSKER, M. D., AND S. R. S. VARADHAN (1975): "Asymptotic Evaluation of Certain Markov Process Expectations for Large Time I," *Communications on Pure and Applied Mathematics*, 28, 1–47.

DUFFIE, D. (2001): *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, New Jersey, third edn.

DUFFIE, D., AND L. G. EPSTEIN (1992): "Stochastic Differential Utility," *Econometrica*, 60, 353–394.

DUFFIE, D., AND P.-L. LIONS (1996): "PDE Solutions of Stochastic Differential Utility," *Journal of Mathematical Economics*, 21, 577–606.

ELLSBERG, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643–669.

Epstein, L., and M. Schneider (2003): “Recursive Multiple Priors,” *Journal of Economic Theory*, 113, 1–31.

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

——— (1991): “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis,” *The Journal of Political Economy*, 99, 263–286.

Ergin, H., and F. Gul (2009): “A Theory of Subjective Compound Lotteries,” *Journal of Economic Theory*, 144, 899–929.

Gabaix, X. (2008): “Variable Rare Disasters: A Tractable Theory of Ten Puzzles in Macro-Finance,” *American Economic Review: Papers & Proceedings*, 98.

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

Gindrat, R., and J. Lefoll (2010): “Smooth Ambiguity Aversion and the Continuous-Time Limit,” working paper, University of Geneva and SFI, Switzerland.

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

——— (2007): *Robustness*. Pinceton University Press, Princeton, NJ.

——— (2009): “Robustness, Estimation, and Detection,” Working Paper, dept. of Economics, University of Chicago.

Hayashi, T. (2005): “Intertemporal Substitution, Risk Aversion and Ambiguity Aversion,” *Economic Theory*, 25, 933–956.

Holmstrom, B., and P. Milgrom (1987): “Aggregation and Linearity in the Provision of Intertemporal Incentives,” *Econometrica*, 55, 303–328.

Jacod, J., and A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin Heidelberg, second edn.

Johnsen, T. H., and J. B. Donaldson (1985): “The Structure of Intertemporal Preferences under Uncertainty and Time Consistent Plans,” *Econometrica*, 53, 1451–1458.

Ju, N., and J. Miao (2009): “Ambiguity, Learning and Asset Returns,” working paper, dept. of Economics, Boston University.

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

——— (2007): "Recursive Smooth Ambiguity Preferences," working paper, MEDS, Kellogg School of Management.

Klibanoff, P., and E. Ozdenoren (2007): "Subjective Recursive Expected Utility," *Economic Theory*, 30, 49–87.

Knight, F. H. (1921): *Risk, Uncertainty and Profit*. Houghton Mifflin, Boston and New York.

Kraft, H., and F. T. Seifried (2011): "Stochastic Differential Utility as the Continuous-Time Limit of Recursive Utility," working paper, Goethe University, Frankfurt and University of Kaiserslautern.

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

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

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

Merton, R. C. (1990): *Continuous Time Finance*. Blackwell, Malden, MA.

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

Pardoux, E. (1997): "Generalized Discontinuous Backward Stochastic Differential Equations," in *Backward Stochastic Differential Equations*, ed. by N. El Karoui, and L. Mazliak, pp. 207–219. Addison Wesley Longman, Essex, England.

Pardoux, E., and S. Peng (1990): "Adapted Solution of a Backward Stochastic Differential Equation," *Systems and Control Letters*, 14, 55–61.

Pardoux, E., F. Pradeilles, and Z. Rao (1997): "Probabilistic Interpretation of a System of Semi-Linear Parabolic Partial Differential Equations," *Annales de l'Institut Henri Poincaré*, 33, 467–490.

Pratt, J. W. (1964): "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 122–136.

Protter, P. E. (2004): *Stochastic Integration and Differential Equations*. Springer Verlag, New York, second edn.

Rietz, T. A. (1988): "The Equity Risk Premium: A Solution," *Journal of Monetary Economics*, 22, 117–131.

Sannikov, Y. (2007): "Games with Imperfectly Observable Actions in Continous Time," *Econometrica*, 75, 1285–1329.

Schroder, M., and C. Skiadas (1999): "Optimal Consumption and Portfolio Selection with Stochastic Differential Utility," *Journal of Economic Theory*, 89, 68–126.

——— (2008): "Optimality and State Pricing in Constrained Financial Markets with Recursive Utility under Continuous and Discontinuous Information," *Mathematical Finance*, 18, 199–238.

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

——— (1990): "Two-Stage Lotteries Without the Reduction Axiom," *Econometrica*, 58, 349–377.

Selden, L. (1978): "A New Representation of Preferences over 'Certain x Uncertain' Consumption Pairs: The 'Ordinal Certainty Equivalent' Hypothesis," *Econometrica*, 46, 1045–1060.

Skiadas, C. (1998): "Recursive Utility and Preferences for Information," *Economic Theory*, 12, 293–312.

——— (2008): "Dynamic Portfolio Choice and Risk Aversion," in *Handbooks in OR & MS,* Vol. 15, ed. by J. R. Birge, and V. Linetsky, chap. 19, pp. 789–843. Elsevier.

——— (2009): *Asset Pricing Theory.* Princeton Univ. Press, Princeton, NJ.

Stokey, N. L. (2009): *The Economics of Inaction: Stochastic Control Models with Fixed Costs.* Princeton Univ. Press, Princeton, New Jersey.

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

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