Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®

ELSEVIER Stochastic Processes and their Applications 108 (2003) 155–202

stochastic processes and their applications

www.elsevier.com/locate/spa

Optimal lifetime consumption-portfolio strategies under trading constraints and generalized recursive preferences

Mark Schroder[a], Costis Skiadas[b,*]

[a]Department of Finance, Eli Broad Graduate School of Management, Michigan State University, 323 Eppley Center, East Lansing, MI 48824, USA

[b]Kellogg School of Management, Department of Finance, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA

Received 31 July 2003; received in revised form 3 September 2003; accepted 3 September 2003

Abstract

We consider the lifetime consumption-portfolio problem in a competitive securities market with essentially arbitrary continuous price dynamics, and convex trading constraints (e.g., incomplete markets and short-sale constraints). Abstract first-order conditions of optimality are derived, based on a preference-independent notion of constrained state pricing. For homothetic generalized recursive utility, we derive closed-form solutions for the optimal consumption and trading strategy in terms of the solution to a single constrained BSDE. Incomplete market solutions are related to complete markets solutions with modified risk aversion towards non-marketed risk. Methodologically, we develop the utility gradient approach, but for the homothetic case we also verify the solution using the dynamic programming approach, without having to assume a Markovian structure. Finally, we present a class of parametric examples in which the BSDE characterizing the solution reduces to a system of Riccati equations.

1. Introduction

We consider the lifetime consumption-portfolio problem for an agent with some initial wealth who can trade in a competitive securities market with essentially arbitrary continuous price dynamics, and whose portfolio, in terms of wealth proportions,

* Corresponding author. Tel.: +1-847-4672328; fax: +1-847-4915719.

E-mail addresses: schrode7@pilot.msu.edu (M. Schroder), c-skiadas@kellogg.northwestern.edu (C. Skiadas).

URLs: http://www.msu.edu/~schrode7/, http://www.kellogg.nwu.edu/faculty/skiadas/home.htm

0304-4149/$ - see front matter
doi:10.1016/j.spa.2003.09.001

is constrained to lie in a convex set at all times. Examples include incomplete markets, short-sale or borrowing constraints, and maximum investment constraints. For any utility function with a well-defined supergradient density, we develop the first-order conditions of optimality, which are shown to be sufficient and, given utility smoothness, necessary for optimality. The solution to the first-order conditions is interpreted as the solution to an unconstrained complete markets problem with properly modified expected instantaneous returns, to reflect the shadow price of constraints.

The characterization of the optimum is applied to the generalized recursive utility specification proposed by Lazrak and Quenez (2003) as a unifying extension of the usual time-additive and recursive formulations (Epstein and Zin, 1989, Duffie and Epstein, 1992b), and multiple-prior formulations (Chen and Epstein, 2002; Anderson et al., 2000; Maenhout, 1999). The utility specification allows for first and second-order risk aversion (in a dynamic version of the definition given by Segal and Spivak, 1990) that can be dependent on the source of risk, a dependence that can be thought of as reflecting the source's ambiguity (in the sense of Ellsberg, 1961). We henceforth refer to generalized recursive utility simply as "recursive utility", except for emphasis.

Within the class of recursive utilities, we characterize the homothetic case, and derive the significantly simpler first-order conditions in this case, leading to several interesting applications. For example, we specify the class of homothetic recursive utilities for which the optimal consumption strategy is a prescribed deterministic function, for any underlying price dynamics, generalizing a familiar logarithmic example. For the case of incomplete markets, we identify a class of homothetic utilities that result in instantaneously mean–variance efficient trading strategies, generalizing related examples by Giovannini and Weil (1989) and Schroder and Skiadas (1999).

Other applications provide links between market incompleteness and source-dependent risk aversion. For example, any incomplete markets solution under homothetic Duffie–Epstein (1992) utility (e.g., time-additive HARA) with a coefficient of relative risk aversion $\gamma < 2$ is mapped to a complete markets solution obtained by pricing risk-neutrally all non-marketed risk, and using a homothetic recursive utility with coefficient of relative risk aversion toward marketed risk equal to γ, and coefficient of relative risk aversion toward non-marketed risk equal to $1/(2-\gamma)$. Another application provides general conditions under which first-order risk aversion results in market non-participation in the context of a stochastic investment opportunity set and trading constraints. Moreover, under a deterministic investment opportunity set, an example is given of an agent who will always go long an asset with positive expected instantaneous excess return, but will not short the asset unless the expected instantaneous excess return is sufficiently negative.

This paper's solutions include the classic Merton (1971) optimal strategies;[1] the complete market solutions of Svensson (1989), Obstfeld (1994), and Schroder and Skiadas (1999); the incomplete market examples of Kim and Omberg (1996), Chacko and Viceira (1999), Liu (2001), and Zariphopoulou and Tiu (2002); and the multiple priors examples of Chen and Epstein (2002). Moreover, the utility specification includes the robust-control type criterion used by Anderson et al. (2000), Hansen et al. (2001),

[1] The extension of the exponential utility case is treated in Schroder and Skiadas (2003a).

and Maenhout (1999), via a utility equivalence result by Skiadas (2003). The recent solution of Uppal and Wang (2002) (developed concurrently to this paper) is also a special case. Approximate solutions, not discussed in this paper, are reviewed in the recent book by Campbell and Viceira (2002) and the references therein.

Methodologically, we extend to the constrained case the utility gradient approach originating in Cox and Huang (1989) and Karatzas et al. (1987) for the additive-utility complete-markets case, proposed for non-additive utilities with a utility gradient by Skiadas (1992) and Duffie and Skiadas (1994), and implemented by Schroder and Skiadas (1999) for the case of recursive utility in complete markets.[2] El Karoui et al. (2001) extend the first-order conditions in complete markets to include non-linear wealth dynamics under generalized recursive utility. For the additive utility case, related convex duality characterizations have been developed by He and Pearson (1991) and Karatzas et al. (1991) (incomplete markets); Xu (1990) and Shreve and Xu (1992) (short-sale constraints); and Cvitanić and Karatzas (1992) (convex constraints). This literature has dealt mainly with applications to existence proofs. Our focus in this paper is not duality or existence, but rather necessary and sufficient first-order conditions of optimality that one would need to solve to compute a solution to either a primal or a dual formulation. Duality in a setting general enough to include this paper's formulation is developed in Schroder and Skiadas (2003b).

The first-order conditions under recursive preferences take the form of a constrained forward–backward stochastic differential equation (FBSDE). Wealth is computed in a recursion starting with a time-zero value forward in time, while utility and the shadow price of wealth are computed in a recursion starting with a terminal date value backward in time. The forward and backward components are coupled. Under the additional assumption of homotheticity, we show that the FBSDE uncouples, resulting in a single constrained backward stochastic differential equation (BSDE). While the conditions we derive are sufficient and necessary for optimality (under regularity assumptions), general appropriate existence results are lacking in the literature (referenced later in the paper), which imposes technical assumptions that are typically violated in the type of applications we discuss. The study of constrained BSDE systems is in its infancy (see, for example, Cvitanić et al., 2002), and we expect it will receive a lot more attention in mathematical and applied research in the future. For a broader historical perspective of the relationship between control problems and BSDEs we refer to the book by Yong and Zhou (1999).

This paper's theory can be extended in various directions. In Schroder and Skiadas (2003a) we consider the case of a nontradeable endowed income stream and constraints on the vector of portfolio market values, and we show simplifications of the solution for recursive utility specifications implying no wealth effects. In Schroder and Skiadas (2003b) we extend this paper's main abstract argument by relaxing the assumption of Brownian information and continuous price processes, by introducing joint constraints in wealth, portfolio positions, and consumption level, and by allowing wealth dynamics

[2] Heuristic but insightful computations in the spirit of Schroder and Skiadas (1999) were also carried out independently in a working paper by Fisher and Gilles (1998).

to be non-linear, a generality that can be used to model differential borrowing or lending rates, market impact, or certain types of taxation.

The remainder of the paper is organized in four sections and an appendix with proofs. The problem is formulated in its general form in Section 2, where sufficiency and necessity of the first-order conditions are shown. Consecutive sections then specialize the theory to recursive utility (Section 3), homothetic recursive utility (Section 4), and "quasi-quadratic" recursive utility (Section 5), which includes all homothetic Duffie–Epstein specifications.

2. Optimality and utility gradient

This section defines the securities market and optimality, and presents the first-order conditions of optimality given any utility function over consumption plans with a well-defined (super)gradient density. The study of recursive utility begins in the following section.

2.1. Stochastic setting

We consider a probability space $(\Omega, \mathscr{F}, P)$ supporting a d-dimensional standard Brownian motion, B, over the finite time horizon[3] $[0, T]$. All processes appearing in this paper are assumed to be progressively measurable with respect to the augmented filtration $\{\mathscr{F}_t : t \in [0, T]\}$ generated by B. We also assume that $\mathscr{F}_T = \mathscr{F}$. The conditional expectation operator $E[\cdot \,|\mathscr{F}_t]$ will be abbreviated to E_t throughout.

Given any subset, S, of a Euclidean space, we define $\mathscr{L}(S)$ to be the set of all S-valued progressively measurable processes, and, with $\|\cdot\|$ denoting the usual Euclidean norm, $\mathscr{L}_p(S) = \{x \in \mathscr{L}(S) : \int_0^T \|x_t\|^p \, \mathrm{d}t < \infty \text{ a.s.}\}$, $p = 1, 2$. Of frequent use will be the spaces of real-valued processes:

$$\mathscr{H} = \left\{ x \in \mathscr{L}(\mathbb{R}) : E\left[\int_0^T x_t^2 \, \mathrm{d}t + x_T^2 \right] < \infty \right\},$$

$$\mathscr{S} = \left\{ x \in \mathscr{L}(\mathbb{R}) : E\left[\left(\operatorname*{ess\,sup}_{t \in [0,T]} |x_t| \right) \right] < \infty \right\}.$$

The space $\mathscr{H}$ will be regarded as a Hilbert space with inner product

$$(x|y) = E\left[\int_0^T x_t y_t \, \mathrm{d}t + x_T y_T \right].$$

[3] We do not discuss the infinite-horizon case in this paper. From an application point of view, the life-time effects of consumption/portfolio choice are very important, and motivate our modeling choice. From a mathematical point of view, the first-order conditions can be extended essentially unchanged with $T = \infty$, although some of the regularity assumptions may appear overly restrictive in this case, and one may wish to relax them to more general transversality conditions. In addition, one must be careful about properly formulating BSDEs with an infinite horizon, as discussed in the context of recursive utility in the Appendix with Skiadas of Duffie and Epstein (1992b).

An element $x \in \mathscr{H}$ is to be thought of as a *cash flow*, where x_t, for $t < T$, represents a time-t payment rate, and x_T represents a lump-sum terminal payment. Given the reflexivity of $\mathscr{H}$, we will see that its elements can also be interpreted as state price densities.

As usual, we identify any elements x and $\tilde{x}$ in $\mathscr{H}$ such that $(x-\tilde{x}|x-\tilde{x})=0$. Given any process x valued in $\mathbb{R}^m$ for some m, and any subset S of $\mathbb{R}^m$, statements of the form "$x_t \in S$ for all t", or "x is valued in S", should be interpreted to mean that the indicator function of the condition $x_t \notin S$ is zero as an element of $\mathscr{H}$. The qualification "almost surely" (or a.s.) will be omitted where it is obviously implied by the context. The subset of strictly positive elements of $\mathscr{H}$ is denoted $\mathscr{H}_{++} = \mathscr{H} \cap \mathscr{L}(\mathbb{R}_{++})$.

On occasion, given any $b \in \mathscr{L}_2(\mathbb{R}^d)$, we utilize the notation B^b and ξ^b, to denote the processes satisfying

$$\mathrm{d}B_t^b = \mathrm{d}B_t + b_t\,\mathrm{d}t, \quad B_0^b = 0, \quad \text{and} \quad \frac{\mathrm{d}\xi_t^b}{\xi_t^b} = -b_t'\,\mathrm{d}B_t, \; \xi_0^b = 1. \tag{1}$$

If ξ^b is a martingale (for example, if b satisfies the Novikov condition), then B^b is Brownian motion under the probability P^b with density $\mathrm{d}P^b/\mathrm{d}P = \xi_T^b$.

The mathematical background is given in detail in Karatzas and Shreve (1998), and is summarized in the appendices of Duffie (2001).

2.2. Market and optimality

We consider a securities market allowing short-term default-free borrowing and lending at a rate given by the stochastic process r, which, for simplicity of exposition, is assumed bounded. We refer to trading at this rate as the "money market". The securities market also allows trading in $n \leqslant d$ risky assets, whose instantaneous excess returns (relative to r) are represented by the n-dimensional Ito process R, with dynamics

$$\mathrm{d}R_t = \mu_t^R\,\mathrm{d}t + \sigma_t^{R\prime}\,\mathrm{d}B_t,$$

where $\mu^R \in \mathscr{L}_1(\mathbb{R}^n)$ and $\sigma^R \in \mathscr{L}_2(\mathbb{R}^{d\times n})$. We assume throughout that σ_t^R is everywhere full-rank for all t (and therefore is invertible if $n=d$).

A *trading strategy* is any process $\psi \in \mathscr{L}(\mathbb{R}^n)$ such that

$$\int_0^t |\psi_s'\mu_s^R| + \psi_s'\sigma_s^{R\prime}\sigma_s^R\psi_s\,\mathrm{d}s < \infty, \quad \text{a.s. for all } t < T.$$

We interpret the ith component of ψ_t as the time-t proportion of wealth invested in security $i \in \{1,\ldots,n\}$, the remaining wealth being invested in the money market. A *consumption strategy* is any process $\rho \in \mathscr{L}_1(\mathbb{R}_+)$. For every time t, ρ_t represents a time-t consumption rate as a proportion of wealth. A *strategy*, (ψ,ρ), is a trading and consumption strategy pair.

We consider an agent characterized by the primitives $(\mathscr{C}, U_0, w_0, K)$, where $\mathscr{C} \subseteq \mathscr{H}_{++}$ is a set of *consumption plans*, $U_0 : \mathscr{C} \to \mathbb{R}$ is a utility function, $w_0 > 0$ is the agent's initial wealth, and $K \subseteq \mathbb{R}^n$ is a nonempty convex closed set used to define trading constraints. Given any consumption plan $c \in \mathscr{C}$, we think of c_t, $t < T$, as the time-t consumption rate, while c_T represents the terminal consumption or bequest.

While only needed for the necessity of the first-order conditions, for simplicity, we assume the following condition throughout the paper: $\mathscr{C}$ is a convex cone that contains all strictly positive constant (deterministic) processes, and $c' \leqslant c$ implies $c' \in \mathscr{C}$, for all $c \in \mathscr{C}$ and $c' \in \mathscr{H}_{++}$.

Associated with the strategy (ψ, ρ) is a strictly positive *wealth process* $W^{\psi,\rho}$, defined by

$$W_0^{\psi,\rho} = w_0, \qquad \frac{\mathrm{d}W_t^{\psi,\rho}}{W_t^{\psi,\rho}} = (r_t - \rho_t)\,\mathrm{d}t + \psi_t'\,\mathrm{d}R_t. \tag{2}$$

The strategy (ψ, ρ) *finances* the consumption plan c if

$$c_t = \rho_t W_t^{\psi,\rho} \text{ for } t < T, \qquad \text{and} \qquad c_T = W_T^{\psi,\rho}.$$

The strategy (ψ, ρ) is *feasible* if $\psi_t \in K$ for all $t < T$, and the strategy finances some consumption plan $c \in \mathscr{C}$. A consumption plan is *feasible* if it is financed by some feasible strategy. A consumption plan c is *optimal* if it is feasible and $U_0(c) \geqslant U_0(c')$ for any other feasible consumption plan c', while the strategy (ψ, ρ) is *optimal* if it finances an optimal consumption plan. A trading strategy ψ is *feasible* (resp. *optimal*) if (ψ, ρ) is feasible (resp. optimal) for some ρ. Similarly, a consumption strategy ρ is *feasible* (resp. *optimal*) if (ψ, ρ) is feasible (resp. optimal) for some ψ.

Remark 1. One can always assume without loss of generality that $n = d$. To see this, suppose $n = m < d$ and the constraint set is $K \subseteq \mathbb{R}^m$. An equivalent formulation results if we attach $d - m$ fictitious securities, and we let the new constraint set be $\{k \in \mathbb{R}^n : (k_1, \ldots, k_m) \in K,\ k_{m+1} = \cdots = k_d = 0\}$. Despite the redundancy, we will see that allowing the possibility $n < d$ results in more parsimonious modeling of incomplete markets.

Remark 2. For simplicity of exposition, we have taken K to be a subset of $\mathbb{R}^n$. The entire analysis goes through, however, if we allow K to be possibly time-dependent and stochastic. That is, K can be taken to be a function from $\Omega \times [0, T]$ to convex subsets of $\mathbb{R}^n$ (satisfying suitable technical restrictions).

2.3. Geometry of first-order conditions

Geometrically, the first-order conditions of optimality amount to the separation of the set of feasible incremental cash flows and the set of utility improving incremental cash flows, as we now explain.

We fix a reference feasible plan (ψ, ρ), with corresponding wealth process $W = W^{\psi,\rho}$, that finances the consumption plan c. The set of all *feasible incremental cash flows* relative to c is defined by

$$\mathscr{X}(c) = \{x \in \mathscr{H} : c + x \text{ is a feasible consumption plan}\}. \tag{3}$$

The set of *utility improving incremental cash flows* relative to c is defined by

$$\mathscr{Y}(c) = \{x \in \mathscr{H} : U_0(c + x) > U_0(c),\ c + x \in \mathscr{C}\},$$

a convex set if U_0 is quasi-concave. An optimum is obtained if and only if any incremental cash flow that improves utility is infeasible. Optimality of (ψ, ρ) can therefore be stated as

$$\mathscr{X}(c) \cap \mathscr{Y}(c) = \emptyset.$$

The first-order conditions of optimality amount to the strict separation of the sets $\mathscr{X}(c)$ and $\mathscr{Y}(c)$, that is, the existence of a process $\pi \in \mathscr{H}$ such that $(\pi|x) \leqslant 0$ for all $x \in \mathscr{X}(c)$, and $(\pi|x) > 0$ for all $x \in \mathscr{Y}(c)$. Clearly, such a condition is sufficient for optimality. In a finite-dimensional version of this model, necessity would follow from the separating hyperplane theorem (assuming convexity of $\mathscr{Y}(c)$), which does not apply in our setting, however, since neither set being separated need have a non-empty interior.

We follow the *utility gradient approach* of Skiadas (1992) and Duffie and Skiadas (1994), which utilizes the utility gradient at the optimum to support the set $\mathscr{Y}(c)$, and computes the density of the utility gradient explicitly based on the utility specification. Besides overcoming the technical issues of non-empty interiors, this approach has the main benefit that it results in explicit expressions for the state price density at the optimum.

With the above separation argument in mind, we define the set of *state price densities at* c by

$$\Pi(c) = \{\pi \in \mathscr{H} : (\pi|x) \leqslant 0 \text{ for all } x \in \mathscr{X}(c)\}. \tag{4}$$

We can think of $(\pi|x)$ as defining a present value of x, which in a perfectly competitive equilibrium must be non-positive. This position-dependent notion of a state price density extends the familiar position independent one in complete or incomplete markets (see, for example, Duffie, 2001).

The process $\pi \in \mathscr{H}$ is a *supergradient density* of U_0 at c if

$$U_0(c + h) \leqslant U_0(c) + (\pi|h) \quad \text{for all } h \text{ such that } c + h \in \mathscr{C}.$$

The process $\pi \in \mathscr{H}$ is a *utility gradient density* of U_0 at c if

$$(\pi|h) = \lim_{\alpha \downarrow 0} \frac{U_0(c + \alpha h) - U_0(c)}{\alpha} \quad \text{for all } h \text{ such that } c + \alpha h \in \mathscr{C} \text{ for some } \alpha > 0.$$

If π is a supergradient density of U_0 at c and the utility gradient of U_0 at c exists, then the utility gradient density is π.

Proposition 3 (First-order conditions). (a) (*Sufficient conditions*) *Suppose that* $\pi \in \mathscr{H}$ *is a supergradient density of* U_0 *at* c *that is also a state price density at* c. *Then the strategy* (ψ, ρ) *is optimal.*

(b) (*Necessary conditions*) *Suppose that* (ψ, ρ) *is optimal and* $\pi \in \mathscr{H}$ *is a utility gradient density of* U_0 *at* c. *Then* π *is a state price density at* c.

Proof. (a) Since π is a supergradient density at c, $(\pi|x) > 0$ for all $x \in \mathscr{Y}(c)$. Since $\pi \in \Pi(c)$, $(\pi|x) \leqslant 0$ for all $x \in \mathscr{X}(c)$. Therefore $\mathscr{X}(c) \cap \mathscr{Y}(c) = \emptyset$, proving optimality. (b) Consider any $x \in \mathscr{X}(c)$, and define the function $u(\alpha) = U_0(c + \alpha x)$ for all $\alpha \in [0, 1]$.

Since c is optimal, u is maximized at zero, and therefore $u'(0)=(\pi|x) \leqslant 0$. This proves $\pi \in \Pi(c)$. $\square$

2.4. Characterization of state price densities

Having characterized optimality in terms of the state price density property of the utility (super)gradient, we now turn to the characterization of state price densities, which when coupled with utility (super)gradient computations leads to more explicit first-order conditions of optimality.

Defining the support function, $\delta_K : \mathbb{R}^n \to (-\infty, \infty]$ of K by

$$\delta_K(\varepsilon) = \sup\{k'\varepsilon : k \in K\}, \tag{5}$$

we show that the state price density property of a strictly positive Ito process, π, is characterized by the following condition.

Condition 4. The process $\pi \in \mathscr{H}_{++}$ follows the dynamics

$$\frac{\mathrm{d}\pi_t}{\pi_t} = -(r_t + \delta_K(\varepsilon_t))\,\mathrm{d}t - \eta_t'\,\mathrm{d}B_t,$$

where $\varepsilon_t = \mu_t^R - \sigma_t^{R\prime}\eta_t$, and $\eta \in \mathscr{L}_2(\mathbb{R}^d)$ is such that $\psi_t'\varepsilon_t = \delta_K(\varepsilon_t)$, $t < T$.

To interpret the condition, we can think of

$$\hat{r} = r + \delta_K(\varepsilon) \quad \text{and} \quad \hat{\mu}^R = \sigma^{R\prime}\eta = \mu^R - \varepsilon \tag{6}$$

as instantaneous expected returns implied by the agent's "fundamental" marginal pricing at c. At the optimum the agent is constrained from further exploiting fundamental mispricings in this sense. On the other hand, the optimal portfolio cannot be mispriced, since it is always feasible to increase or decrease exposure in the market, without affecting the asset allocation in terms of wealth proportions, by simply adjusting the amount consumed. This reasoning suggests the following conditions, which are easily seen to be equivalent to Condition 4:

$$r_t + k'\mu_t^R \leqslant \hat{r}_t + k'\hat{\mu}_t^R \quad \text{for all } k \in K, \quad \text{and} \quad r_t + \psi_t'\mu_t^R = \hat{r}_t + \psi_t'\hat{\mu}_t^R. \tag{7}$$

(The above intuition is further extended in Schroder and Skiadas (2003b), based on the notion of quasiarbitrage, a notion not discussed here as it is peripheral to this paper's objectives.)

The following examples include as special cases incomplete markets and short-sale constraints.

Example 5 (Conical constraints). Suppose K contains zero, then $\delta_K \geqslant 0$, and therefore $\hat{r} \geqslant r$. Suppose further that K is a cone. Then δ_K vanishes on

$$\hat{K} = \{\varepsilon \in \mathbb{R}^n : \delta_K(\varepsilon) < \infty\} = \{\varepsilon \in \mathbb{R}^n : k'\varepsilon \leqslant 0 \text{ for all } k \in K\}.$$

The restriction $\psi_t'\varepsilon_t = \delta_K(\varepsilon_t)$ of Condition 4 becomes $\varepsilon_t \in \hat{K}$ and $\psi_t'\varepsilon_t = 0$, and implies that $\hat{r} = r$.

Example 6 (Rectangular constraints). Suppose

$$K = \{k \in \mathbb{R}^n : k_i \in [\alpha_i, \beta_i],\ i = 1, \dots, n\},$$

where $-\infty \leqslant \alpha_i \leqslant 0 \leqslant \beta_i \leqslant \infty$ for each i. Since K contains zero, $\delta_K \geqslant 0$. The restriction $\psi_t' \varepsilon_t = \delta_K(\varepsilon_t)$ becomes $(k_i - \psi_i)\varepsilon_i \leqslant 0$ for all $k_i \in [\alpha_i, \beta_i]$ and $i = 1, \dots, n$. If trading on asset i is impossible (incomplete markets) then $\alpha_i = \beta_i = 0$, and ε_i is unrestricted. If $\alpha_i < \beta_i$, the restriction $\psi_t' \varepsilon_t = \delta_K(\varepsilon_t)$ can equivalently be written as

$$\psi_i = \alpha_i \Rightarrow \varepsilon_i \leqslant 0; \quad \psi_i \in (\alpha_i, \beta_i) \Rightarrow \varepsilon_i = 0; \quad \psi_i = \beta_i \Rightarrow \varepsilon_i \geqslant 0.$$

No short selling of asset i corresponds to $\alpha_i = 0$ and $\beta_i = \infty$, and therefore $\varepsilon_i \leqslant 0$ and $\psi_i \varepsilon_i = 0$.

The section's main conclusion is given in the following key theorem, proved in the appendix.

Theorem 7. *Suppose that* (ψ, ρ) *is a feasible strategy with wealth process* $W = W^{\psi,\rho}$, $\pi \in \mathscr{H}_{++}$ *is an Ito process, and* $\pi W \in \mathscr{S}$. *Let* c *be the consumption plan financed by* (ψ, ρ), *and let* $\Pi(c)$ *be the corresponding set of state price densities, defined in Eq.* (4).

(a) (*Sufficiency*) *Condition* 4 *implies* $\pi \in \Pi(c)$.
(b) (*Necessity*) *If* ρ *is continuous, then* $\pi \in \Pi(c)$ *implies Condition* 4.

Remark 8. The necessity part excludes the formulation in which utility is defined over terminal consumption only. While sufficiency is all we need to embed available solutions, such as those of Liu (2001), to our setting, a necessity result is obtained as a corollary to Theorem 7, under the assumption that trading in the money market is unrestricted, in which case utility for terminal wealth can be extended by treating intermediate consumption as a money-market cash flow.

Complete markets are obtained if $n = d$ and $K = \mathbb{R}^d$, in which case the first-order conditions of optimality reduce to the ones presented in Schroder and Skiadas (1999). In complete markets, Condition 4 can be restated as: $\mathrm{d}\pi/\pi = -r\,\mathrm{d}t - \eta'\,\mathrm{d}B$, where $\eta = \sigma^{R\prime-1}\mu^R$ is the unique (*market*) *price of risk process.* In the language of equivalent martingale measures (EMM), assuming $E\xi_T^\eta = 1$, $\mathrm{d}P^\eta/\mathrm{d}P = \xi_T^\eta$ defines the unique EMM P^η.

Suppose now that $n{=}d$ (see Remark 1), that a utility gradient density at the optimum exists, and markets are constrained: $K \subset \mathbb{R}^d$. In this case, the first-order conditions of optimality select one of possibly many state price densities at the optimum, namely, the utility gradient density π at the optimum. The subspace orthogonal to π can be thought of as the set of marketed incremental cash flows in a fictitious complete market, in which the agent selects the same optimal strategy and consumption plan. This fictitious complete market is described in terms of the quantities in Condition 4 in the following result.

Corollary 9. *Suppose that* $n=d$, (ψ,ρ) *is an optimal strategy financing the consumption plan* c, $\pi\in\mathscr{H}_{++}$ *is the utility gradient density of* U_0 *at* c, $\pi W^{\psi,\rho}\in\mathscr{S}$, *and* ρ *is continuous. Then Condition 4 holds, and* (ψ,ρ) *is optimal in a fictitious market obtained from the original market by relaxing the trading constraints (that is, letting* $K=\mathbb{R}^d$*), and assuming that the short rate process is* $\hat{r}=r+\delta_K(\varepsilon)$*, and the instantaneous expected excess returns are* $\hat{\mu}^R=\mu^R-\varepsilon$.

Proof. In this fictitious market, the price of risk process is $\eta=\sigma^{R\prime-1}\hat{\mu}^R$. Moreover, since $r+\psi'\mu^R=\hat{r}+\psi'\hat{\mu}^R$, feasibility of (ψ,ρ) in the original market implies feasibility of (ψ,ρ) in the fictitious complete market. The result follows from Theorem 7. □

The above characterization leads immediately to a duality formulation, as explained in Schroder and Skiadas (2003b).

2.5. Incomplete markets

Suppose there is an $m\leqslant n$ such that trading is possible only in the first m risky assets and the money market, possibly subject to some constraints. As pointed out in Remark 1, we can model such market incompleteness either by letting $n=m$, or through K with $n=d$. In this section we relate the first-order conditions in the two approaches, used on several occasions in examples in subsequent sections.

Given any vector $x\in\mathbb{R}^n$ and matrix $y\in\mathbb{R}^{n\times n}$, for any $m\leqslant n\leqslant d$, we write

$$x=\begin{bmatrix}x_M\\ x_N\end{bmatrix}\quad\text{and}\quad y=\begin{bmatrix}y_{MM} & y_{MN}\\ y_{NM} & y_{NN}\end{bmatrix}=[y_{*M}\;\; y_{*N}],$$

where $x_M\in\mathbb{R}^m$ and $y_{MM}\in\mathbb{R}^{m\times m}$. To simplify notation, we often suppress time indices, writing, for example, $\mathrm{d}R_M=\mu_M^R\,\mathrm{d}t+\sigma_{*M}^{R\prime}\,\mathrm{d}B$, for the dynamics of the marketed assets. The d-dimensional identity matrix is denoted I_d.

The following lemma will be useful in simplifying the return dynamics:

Lemma 10. *There exists a progressively measurable process* Φ *that is valued in the set of orthogonal* $d\times d$ *matrices* $(\Phi_t'\Phi_t=I_d)$*, in terms of which the excess return dynamics are of the form* $\mathrm{d}R_M=\mu_M^R\,\mathrm{d}t+\bar{\sigma}_{MM}^{R\prime}\,\mathrm{d}\bar{B}_M$*, where* $\bar{B}=[\bar{B}_M',\bar{B}_N']'$ *is the Brownian motion defined by* $\mathrm{d}\bar{B}_t=\Phi_t'\,\mathrm{d}B_t$, $\bar{B}_0=0$.

We can therefore assume without loss of generality, passing to a new Brownian motion if necessary, that the marketed assets have the normalized dynamics

$$\mathrm{d}R_M=\mu_M^R\,\mathrm{d}t+\sigma_{MM}^{R\prime}\,\mathrm{d}B_M. \tag{8}$$

It should be emphasized, however, that while only the first m components of the Brownian motion B appear in the above equation, the processes μ_M^R and σ_{MM}^R need not be adapted to the filtration generated by B_M.

Adopting this normalization, we now compare the form of the utility gradient at the optimum in the two modeling approaches of incomplete markets. In both approaches, we assume the trading restriction $\psi_M \in K_M$, for some convex set $K_M \subseteq \mathbb{R}^m$, and we define the corresponding support function

$$\delta_{K_M}(\varepsilon_M) = \sup\{k'\varepsilon_M : k \in K_M\}, \quad \varepsilon_M \in \mathbb{R}^m.$$

Modeling Approach A. $n = m \leqslant d$ and $K = K_M \subseteq \mathbb{R}^m$. In this case, $\delta_K = \delta_{K_M}$, $R = R_M$, $\mu^R = \mu^R_M$, and $\sigma^{R\prime} = [\sigma^{R\prime}_{MM}, 0]$. The restriction $\mu^R - \varepsilon = \sigma^{R\prime}\eta$ of Condition 4 can be equivalently stated as

$$\eta_M = (\sigma^{R\prime}_{MM})^{-1}\hat{\mu}^R_M, \quad \hat{\mu}^R_M = \mu^R_M - \varepsilon_M. \tag{9}$$

This expression specifies the price of marketed risk, modified to reflect the shadow price of constrains on the marketed assets. If $K_M = \mathbb{R}^m$, then $\varepsilon_M = 0$. The price of non-marketed risk, η_N, is unrestricted, and parameterizes the set, $\Pi(c)$, of state price densities at c:

$$\pi = \pi^M \xi^{\eta_N}, \quad \text{where}$$
$$\frac{d\pi^M}{\pi^M} = -(r + \delta_{K_M}(\varepsilon_M))\,dt - \eta'_M\,dB_M \quad \text{and} \quad \frac{d\xi^{\eta_N}}{\xi^{\eta_N}} = -\eta'_N\,dB_N.$$

Modeling Approach B. $n = d$ and $K = \{k \in \mathbb{R}^d : k_M \in K_M \text{ and } k_N = 0\}$. In this case, $\delta_K(\varepsilon) = \delta_{K_M}(\varepsilon_M)$, while the return normalization (8) implies

$$\sigma^R = \begin{bmatrix} \sigma^R_{MM} & \sigma^R_{MN} \\ 0 & \sigma^R_{NN} \end{bmatrix}. \tag{10}$$

The first m components of the restriction $\mu^R - \varepsilon = \sigma^{R\prime}\eta$ of Condition 4 give the price of marketed risk expression (9) once again. The remaining $d - m$ components can be used to solve for the price of non-marketed risk:

$$\eta_N = (\sigma^{R\prime}_{NN})^{-1}(\hat{\mu}^R_N - \sigma^{R\prime}_{MN}\eta_M), \quad \hat{\mu}_N = \mu_N - \varepsilon_N.$$

Since ε_N is unrestricted, so is η_N, recovering the above parametrization of all state price densities. In the fictitious complete market of Corollary 9, $\hat{\mu}_N$ is a non-marketed asset expected instantaneous excess return that induces zero optimal demand for the non-marketed assets.

3. Generalized recursive utility

Having established the first-order conditions of optimality in terms of the utility (super)gradient density, in this section we specialize the results to (generalized) recursive utility. We begin with the utility definition and some examples, followed by the first-order conditions, and an outline of the corresponding PDE system in a Markovian setting.

3.1. Utility specification

Taking as primitive a set $\mathscr{U}$ of progressively measurable processes, for every consumption plan c, $U_0(c)$ is assumed to be the initial value of the unique process $U = U(c)$

in $\mathscr{U}$ that solves the BSDE

$$dU_t = -F(t, c_t, U_t, \Sigma_t)\,dt + \Sigma_t'\,dB_t, \qquad U_T = F(T, c_T). \tag{11}$$

The function $F : \Omega \times [0,T] \times (0,\infty) \times \mathbb{R}^{1+d} \to \mathbb{R}$ is called the *aggregator*, and is always defined so that $F(T, c_T, U, \Sigma)$ does not depend on the arguments (U, Σ), which are therefore notationally suppressed. $F(t, c_t, U_t, \Sigma_t)$ denotes the random variable that maps ω to $F(\omega, t, c(\omega,t), U(\omega,t), \Sigma(\omega,t))$. The symbols (c, U, Σ) are also used to denote dummy variables in $(0,\infty) \times \mathbb{R} \times \mathbb{R}^d$, with the meaning being clear from the context.

Condition 11 below will be assumed throughout the rest of this paper (although concavity is not needed for the necessity of the first-order conditions, and the Inada condition is not needed for sufficiency).

Condition 11 (Standing assumption). For any $c \in \mathscr{C}$, $U_0(c) = U_0$, where $(U, \Sigma) \in \mathscr{U} \times \mathscr{L}_2(\mathbb{R}^d)$ uniquely solves *BSDE* (11). The aggregator $F : \Omega \times [0,T] \times (0,\infty) \times \mathbb{R}^{1+d} \to \mathbb{R}$ is *regular*, meaning that the following conditions hold for all $(\omega, t, c, U, \Sigma) \in \Omega \times [0,T] \times (0,\infty) \times \mathbb{R}^{1+d}$:

1. $F(\cdot, c, U, \Sigma)$ is progressively measurable.
2. $F(\omega, t, \cdot)$ is a concave function.
3. (*Inada condition*) The derivative, $F_c(\omega, t, \cdot, U, \Sigma)$, of $F(\omega, t, \cdot, U, \Sigma)$ exists and maps $(0,\infty)$ onto $(0,\infty)$.

Concrete examples of utilities satisfying the above conditions are given by Schroder and Skiadas (1999), along with existence results. General BSDE existence results for regular aggregators are not available in the literature. The original results of Pardoux and Peng (1990) and Duffie and Epstein (1992b) assume Lipschitz-growth conditions that are violated in this context. Further existence results have been developed by Lepeltier and San Martín (1997), Kobylanski (2000), and Lepeltier and San Martín (2002). Kobylanski's results apply quite generally to Duffie–Epstein utilities, defined below, under the assumption that a and b in Eq. (12) are bounded, which precludes commonly used regular aggregators. Given Kobylanski's stability results, however, the extension to unbounded a and b should not be difficult.

The Inada condition guarantees that the optimal consumption plan is strictly positive, and therefore the non-negativity constraint in consumption is non-binding. While the Inada condition excludes the case of no intermediate consumption, the extension of the first-order conditions to this case amounts to simply omitting the consumption argument c_t for $t < T$.

Remark 12. The results can be extended to include a habit formation term among the aggregator's arguments. If markets are complete, or under conical constraints with deterministic short rate, the isomorphism of Schroder and Skiadas (2002) can be used to mechanically transform this paper's solutions to solutions incorporating linear habit formation. In more general formulations, a utility gradient computation in Duffie and Skiadas (1994) can be used.

3.2. Examples of generalized recursive utility

The utility process V is *ordinally equivalent* to U if there exists a progressively measurable function $\chi : \Omega \times [0,T] \times \mathbb{R} \to \mathbb{R}$ such that, for every $(\omega,t) \in \Omega \times [0,T]$, $\chi(\omega,t,\cdot)$ is strictly increasing, and $V_t = \chi(t,U_t)$. In this case, V and U represent the same preference ordering over consumption plans.

F is a *Duffie–Epstein aggregator* if

$$F(t,c,U,\Sigma) = b(t,c,U) - \frac{a(t,U)}{2}\Sigma'\Sigma, \tag{12}$$

for some functions $b : \Omega \times [0,T] \times (0,\infty) \times \mathbb{R} \to \mathbb{R}$ and $a : \Omega \times [0,T] \times \mathbb{R} \to \mathbb{R}$. Duffie and Epstein (1992b) show that a measures comparative risk aversion, and that,[4] subject to some technical regularity restrictions, if F is a Duffie–Epstein aggregator, then there is an ordinally equivalent version, V, of U that uniquely solves a BSDE of the form

$$\mathrm{d}V_t = -f(t,c_t,V_t)\,\mathrm{d}t + \Sigma_t'\,\mathrm{d}B_t, \qquad V_T = f(T,c_T), \tag{13}$$

for some function $f : \Omega \times [0,T] \times (0,\infty) \times \mathbb{R} \to \mathbb{R}$ (which is related to preferences for information in Skiadas (1998)). While this form is simpler in that the aggregator does not depend on Σ, it can be less convenient in other ways. For example, as the parametric examples in Schroder and Skiadas (1999) show, f need not be jointly concave in consumption and utility.

The *time-additive* case corresponds to f taking the form

$$f_t(c,V) = u_t(c) - \beta_t V, \tag{14}$$

for some process β and a function $u : \Omega \times [0,T] \times (0,\infty) \to \mathbb{R}$, in which case (under minor integrability restrictions),

$$V_t(c) = E_t\left[\int_t^T \mathrm{e}^{-\int_t^s \beta_u\,\mathrm{d}u} u_s(c_s)\,\mathrm{d}s + \mathrm{e}^{-\int_t^T \beta_u\,\mathrm{d}u} u_T(c_T)\right].$$

In the analysis of homothetic utility in the following section, even if the utility specification takes the above simple additive form, or the Duffie–Epstein form (13), it will be more convenient to work with an ordinally equivalent version of the utility that measures utility in certainty equivalent terms. Such a transformation results in an aggregator of the more general form (12).

A multiple-priors formulation that is also generalized recursive utility is presented by Chen and Epstein (2002), who derive the representation

$$\mathrm{d}V_t = -\left(f(t,c_t,V_t) - \max_{\theta \in \Theta_t} \theta'\Sigma_t\right)\mathrm{d}t + \Sigma_t'\,\mathrm{d}B_t, \tag{15}$$

for some function Θ from $\Omega \times [0,T]$ to the set of convex compact subsets of $\mathbb{R}^d$. We return to a concrete example of this type of preferences in the last section.

Anderson et al. (2000) and Hansen et al. (2001) consider a different multiple-prior formulation. As shown in Skiadas (2003), their formulation is equivalent to a type

[4] More precisely, Duffie and Epstein consider the case in which the aggregator is state and time independent. The same arguments, however, extend readily to the more general specification considered here.

of Duffie–Epstein utility, and is therefore also embedded in the current framework. Maenhout (1999) reinterprets the Schroder and Skiadas (1999) results in terms of a variant of the Hansen-Sargent et al. criterion.

3.3. Supergradient density

The formulation of the first-order conditions requires the computation of the utility supergradient density, which is the topic of this subsection.

While we have assumed F to be differentiable with respect to consumption, with corresponding partial F_c, we have not assumed smoothness of F in (U,Σ), a generality that we will see is useful in applications. The *superdifferential* of F with respect to the variables (U,Σ) at (ω,t,c,U,Σ) is defined as the set $(\partial_{U,\Sigma}F)(\omega,t,c,U,\Sigma)$ of all pairs $(a,b)\in\mathbb{R}\times\mathbb{R}^d$ such that

$$F(\omega,t,c,U+y,\Sigma+z)\leqslant F(\omega,t,c,U,\Sigma)+ay+b'z.$$

In particular, if $F(\omega,t,\cdot)$ is differentiable, $(\partial_{U,\Sigma}F)(\omega,t,c,U,\Sigma)$ consists of the single element, $(F_U(\omega,t,c,U,\Sigma),F_\Sigma(\omega,t,c,U,\Sigma))$, which is the pair of the partial derivatives of F with respect to U and Σ.

Given any pair of processes $(a,b)\in\mathscr{L}_1(\mathbb{R})\times\mathscr{L}_2(\mathbb{R}^d)$, we define the stochastic exponential process $\mathscr{E}(a,b)$ by the SDE

$$\frac{\mathrm{d}\mathscr{E}_t(a,b)}{\mathscr{E}_t(a,b)}=a_t\,\mathrm{d}t+b_t'\,\mathrm{d}B_t,\qquad \mathscr{E}_0(a,b)=1. \tag{16}$$

Extending the superdifferential notation, given any triple of processes $(c,U,\Sigma)\in\mathscr{C}\times\mathscr{U}\times\mathscr{L}_2(\mathbb{R}^d)$, the set $(\partial_{U,\Sigma}F)(c,U,\Sigma)$ consists of all processes $(a,b)\in\mathscr{L}_1(\mathbb{R})\times\mathscr{L}_2(\mathbb{R}^d)$ such that $(a_t,b_t)\in(\partial_{U,\Sigma}F)(t,c_t,U_t,\Sigma_t)$ and the process $\mathscr{E}_t=\mathscr{E}_t(a,b)$ satisfies the regularity condition

$$\mathscr{E}U\in\mathscr{S}\quad\text{ for all } U\in\mathscr{U}. \tag{17}$$

The last condition is tailored to the following result, where we use the symbols (F_U,F_Σ) to denote both the partials of F when they exist, and a typical element of $(\partial_{U,\Sigma}F)(c,U,\Sigma)$.

Proposition 13. *Suppose that $c\in\mathscr{C}$, $(U,\Sigma)\in\mathscr{U}\times\mathscr{L}_2(\mathbb{R}^d)$ solves BSDE (11), $(F_U,F_\Sigma)\in(\partial_{U,\Sigma}F)(c,U,\Sigma)$, and*

$$\pi_t=\mathscr{E}_t(F_U,F_\Sigma)F_c(t,c_t,U_t,\Sigma_t) \tag{18}$$

defines an element of $\mathscr{H}$. Then π is a utility supergradient density of U_0 at c. If $F(\omega,t,\cdot)$ is differentiable, then π is the utility gradient density of U_0 at c.

The supergradient density expression (18) is consistent with the calculations of Skiadas (1992), Duffie and Skiadas (1994), Chen and Epstein (2002), and El Karoui et al. (2001). All of these papers assume Lipschitz-growth conditions that are inconsistent with a regular aggregator. Parametric examples under a regular aggregator can be found in Schroder and Skiadas (1999).

3.4. First-order conditions

Having derived an expression for the utility (super)gradient density, in this section we specialize the general first-order condition of optimality to recursive utility, showing that they take the form of a constrained FBSDE.

An important role is played by the strictly positive process

$$\lambda_t = F_c(t, c_t, U_t, \Sigma_t), \tag{19}$$

computed at the optimum. In a time-t formulation of the agent's problem, λ_t is the Lagrange multiplier for the time-t wealth constraint, since it provides the first-order utility increment (per unit of wealth) as a result of slightly increasing time-t wealth. Although we have no need to formalize this interpretation of λ, it is nevertheless important in understanding the first-order conditions. We use the following notation for the dynamics of λ throughout:

$$\frac{\mathrm{d}\lambda_t}{\lambda_t} = \mu_t^\lambda \,\mathrm{d}t + \sigma_t^{\lambda\prime} \,\mathrm{d}B_t. \tag{20}$$

The optimal consumption can be expressed as $c_t = \mathscr{I}(t, \lambda_t, U_t, \Sigma_t)$, where the function $\mathscr{I} : \Omega \times [0, T] \times (0, \infty) \times \mathbb{R}^{1+d} \to (0, \infty)$ is well-defined (by the regularity of F) implicitly through the equation

$$F_c(t, \mathscr{I}(t, \lambda, U, \Sigma), U, \Sigma) = \lambda, \qquad \lambda \in (0, \infty).$$

A utility (super)gradient density π was computed in Proposition 13 as $\pi_t = \mathscr{E}_t \lambda_t$, resulting in the dynamics

$$\frac{\mathrm{d}\pi_t}{\pi_t} = (F_U(t) + \mu_t^\lambda + \sigma_t^{\lambda\prime} F_\Sigma(t))\,\mathrm{d}t + (F_\Sigma(t) + \sigma_t^\lambda)' \,\mathrm{d}B_t.$$

Combining this expression with Condition 4, we can now formulate the first-order conditions for recursive utility as a constrained FBSDE system:

Condition 14 (First-order conditions). The processes $(U, \Sigma, \lambda, \sigma^\lambda, W) \in \mathscr{U} \times \mathscr{L}_2(\mathbb{R}^d) \times \mathscr{L}(\mathbb{R}_{++}) \times \mathscr{L}_2(\mathbb{R}^d) \times \mathscr{L}(\mathbb{R}_{++})$, and the trading strategy ψ, solve the constrained FBSDE:

$$\mathrm{d}U_t = -F(t, \mathscr{I}(t, \lambda_t, U_t, \Sigma_t), U_t, \Sigma_t)\,\mathrm{d}t + \Sigma_t' \,\mathrm{d}B_t, \qquad U_T = F(T, W_T),$$

$$\frac{\mathrm{d}\lambda_t}{\lambda_t} = -(r_t + \delta_K(\varepsilon_t) + F_U(t) + \sigma_t^{\lambda\prime} F_\Sigma(t))\,\mathrm{d}t + \sigma_t^{\lambda\prime} \,\mathrm{d}B_t, \qquad \lambda_T = F_c(T, W_T),$$

$$\mathrm{d}W_t = (W_t(r_t + \psi_t' \mu_t^R) - \mathscr{I}(t, \lambda_t, U_t, \Sigma_t))\,\mathrm{d}t + W_t \psi_t' \sigma_t^{R\prime} \,\mathrm{d}B_t, \qquad W_0 = w_0,$$

$$\varepsilon_t = \mu_t^R + \sigma_t^{R\prime}(F_\Sigma(t) + \sigma_t^\lambda), \;\; \psi_t \in K, \;\; \psi_t' \varepsilon_t = \delta_K(\varepsilon_t), \qquad t < T,$$

$$(F_U, F_\Sigma) \in (\partial_{U,\Sigma} F)(\mathscr{I}(t, \lambda, U, \Sigma), U, \Sigma).$$

The characterization of the optimum is summarized in the following result:

Theorem 15. *Suppose the utility function is specialized by Condition* 11.

(a) (*Sufficiency*) *Suppose Condition* 14 *holds, and let* $c_t = \mathscr{I}(t, \lambda_t, U_t, \Sigma_t)$ *and* $\rho_t = c_t/W_t$. *If* $c \in \mathscr{C}$, $\pi = \mathscr{E}(F_U, F_\Sigma)\lambda \in \mathscr{H}$, *and* $\pi W \in \mathscr{S}$, *then the strategy* (ψ, ρ) *is optimal,* $W^{\psi,\rho} = W$, *and* $U(c) = U$.

(b) (*Necessity*) *Suppose that* (ψ, ρ) *is an optimal strategy financing the continuous consumption plan* c, *and let* $\lambda_t = F_c(t, c_t, U_t, \Sigma_t)$, *where* $(U, \Sigma) \in \mathscr{U} \times \mathscr{L}_2(\mathbb{R}^d)$ *solves BSDE* (11). *Suppose further that* $F(\omega, t, \cdot)$ *is differentiable for all* (ω, t), $(F_U, F_\Sigma) \in (\partial_{U,\Sigma} F)(c, U, \Sigma)$, $\mathscr{E}(F_U, F_\Sigma)\lambda \in \mathscr{H}$, $W = W^{\psi,\rho}$, *and* $\pi W \in \mathscr{S}$. *Then Condition* 14 *is satisfied.*

Proof. (a) By Proposition 13, π is a utility supergradient density at c. By Theorem 7, π is also a state price density, and the result follows by Proposition 3.

(b) By Proposition 13, $\pi = \mathscr{E}(F_U, F_\Sigma)\lambda$ is the utility gradient density at c, which, by Proposition 3, must be a state price density. Theorem 7 completes the proof. □

Remark 16. The first-order conditions for the case of no intermediate consumption are obtained by omitting the consumption argument in the utility and supergradient dynamics. Sufficiency in this case follows by the same arguments, while necessity is qualified by Remark 8.

Given the assumptions of Corollary 9 (which include $n = d$), the constrained solution is obtained as the solution to a fictitious complete market with short rate process $\hat{r} = r + \delta_K(\varepsilon)$ and instantaneous expected excess returns of $\hat{\mu}^R = \mu^R - \varepsilon$. The first-order conditions show that the price of risk process in the fictitious complete market is given by $\eta_t = -(F_\Sigma(t) + \sigma_t^\lambda)$. Recalling the notation in (1), the dynamics of λ can be expressed as

$$\frac{\mathrm{d}\lambda_t}{\lambda_t} = -(\hat{r}_t + F_U(t) - \sigma_t^{\lambda\prime}\sigma_t^\lambda)\,\mathrm{d}t + \sigma_t^{\lambda\prime}\,\mathrm{d}B_t^\eta.$$

If the original market is complete, then $\hat{r} = r$ and $\eta = \sigma^{R\prime -1}\mu^R$.

3.5. The Markovian case

Assuming a Markovian underlying structure, the constrained FBSDE system of the first-order conditions naturally corresponds to a PDE system outlined below (for a smooth aggregator). Essentially, we adapt the Ma et al. (1994) approach to the current setting.

Throughout the subsection, we assume that uncertainty in the model is driven by a Markov process Z valued in $\mathbb{R}^k$, with given initial value $Z_0 \in \mathbb{R}^k$, uniquely solving the SDE

$$\mathrm{d}Z_t = \mu^Z(t, Z_t)\,\mathrm{d}t + \sigma^Z(t, Z_t)'\,\mathrm{d}B_t, \tag{21}$$

for functions $\mu^Z : [0,T] \times \mathbb{R}^k \to \mathbb{R}^k$ and $\sigma^Z : [0,T] \times \mathbb{R}^k \to \mathbb{R}^{d\times k}$. The aggregator and the price dynamics parameters are assumed to be possibly time-dependent but deterministic functions of the state Z. With some convenient duplication of notation, we write these functional relationships as

$$r_t = r(t, Z_t), \quad \mu_t^R = \mu^R(t, Z_t), \quad \sigma_t^R = \sigma^R(t, Z_t), \tag{22}$$

while the aggregator F is now regarded as a smooth function of (t, Z, c, U, Σ).

The central idea is to express the backward variables, (U, λ), as possibly time-dependent but deterministic functions of the forward variables, (W, Z). With the same sort of duplicate notation as in Eqs. (22), we write $U_t = U(t, W_t, Z_t)$ and $\lambda_t = \lambda(t, W_t, Z_t)$. Moreover, recalling the interpretation of λ as the Lagrange multiplier for wealth, we conjecture that $\lambda = U_W$.

To state the PDE system, we introduce two differential operators applied on any (sufficiently smooth) function $f(t, W, Z)$. Both sides of the following definitions are functions of the arguments (t, W, Z), which are omitted in the notation. The parameter functions $c(t, W, Z)$ and $\psi(t, W, Z)$ are valued in $\mathbb{R}_+$ and $\mathbb{R}^n$, respectively.

$$\mathscr{D}_1^{\psi} f = \sigma^R \psi w f_W + \sigma^Z f_Z,$$

$$\mathscr{D}_2^{\psi,c} f = f_t + f_Z' \mu^Z + f_W (wr + w\psi' \mu^R - c) + \tfrac{1}{2} w^2 f_{WW} \psi' \sigma^{R\prime} \sigma^R \psi + w f_{WZ}' \sigma^{Z\prime} \sigma^R \psi + \tfrac{1}{2} \operatorname{trace}(f_{ZZ} \sigma^{Z\prime} \sigma^Z).$$

The PDE system is obtained by matching the Ito expansions of U and λ in the first-order conditions to those obtained by applying Ito's lemma to the functions $U(t, W_t, Z_t)$ and $\lambda(t, W_t, Z_t)$. We state the resulting system, omitting the arguments (t, W, Z). The set Ψ_K consists of all (product-measurable) functions of the form $\psi(t, W, Z)$ that are valued in the set K.

$$0 = F(c, U, \mathscr{D}_1^{\psi} U) + \mathscr{D}_2^{\psi,c} U, \quad c = \mathscr{I}(U_W, U, \mathscr{D}_1^{\psi} U),$$

$$0 = (r + \delta_K(\varepsilon) + F_U(c, U, \mathscr{D}_1^{\psi} U)) U_W + F_{\Sigma}(c, U, \mathscr{D}_1^{\psi} U)' \mathscr{D}_1^{\psi} U_W + \mathscr{D}_2^{\psi,c} U_W,$$

$$\varepsilon = \mu^R + \sigma^{R\prime}(\mathscr{D}_1^{\psi} U_W + F_{\Sigma}(c, U, \mathscr{D}_1^{\psi} U)), \quad \psi \in \Psi_K, \ \ \psi' \varepsilon = \delta_K(\varepsilon),$$

$$U(T, \cdot) = F(T, \cdot) \quad U_W(T, \cdot) = F_c(T, \cdot).$$

Given a sufficiently smooth solution (U, λ, ψ), where $\lambda = U_W$, to the above system, and the forward SDEs describing Z and W, one can use Ito's lemma to recover the first-order conditions of optimality (provided the integrals are well-defined).

4. Homothetic recursive utility

In the remainder of this paper we explore simplifications as a result of preference homotheticity. We begin with a characterization of homothetic recursive utilities. As always, Condition 11 is assumed. For each $c \in \mathscr{C}$, we let $U(c) \in \mathscr{U}$ denote the corresponding utility process, and we use the term *utility* to refer to the function, $c \mapsto U(c)$, mapping consumption processes to utility processes.

4.1. Homotheticity and proportional aggregators

The utility U is *homothetic* if for any time t, and any $c^1, c^2 \in \mathscr{C}$, and $k \in \mathbb{R}_+$, $U_t(c^1) = U_t(c^2)$ implies $U_t(kc^1) = U_t(kc^2)$. Defining the sets $\mathscr{C}_t = \{c \in \mathscr{C} : c_s = c_t, \text{for all } s \geqslant t\}$, the utility U is in *certainty equivalent form* if it is valued in $(0, \infty)$, and for every time t and $c \in \mathscr{C}_t$, $U_t(c) = c_t$. Given this normalization, $U_t(c)$, given any $c \in \mathscr{C}$, is the consumption level that, if frozen in time from time t to T, has the same time-t utility as c.

Remark 17. Any strictly monotone (generalized) recursive utility has an ordinally equivalent version in certainty equivalent form, given some regularity assumptions. We outline an argument for a recursive utility $\tilde{U}$ with state-independent aggregator $\tilde{F}$. We assume that, for every $x > 0$, there exists a deterministic process $\chi_t(x)$ that uniquely solves $\mathrm{d}\chi_t(x) = -\tilde{F}(t, x, \chi_t(x), 0)\,\mathrm{d}t$, $\chi_T(x) = \tilde{F}(T, x)$, and $\chi_t(\cdot)$ is strictly increasing and maps $(0, \infty)$ onto $(0, \infty)$. Noting that $c \in \mathscr{C}_t$ implies $\tilde{U}_t(c) = \chi_t(c_t)$, it follows that $U_t = \chi_t^{-1}(\tilde{U}_t)$ is in certainty equivalent form. If χ^{-1} is twice continuously differentiable, then, by Ito's lemma, U is also recursive utility.

Suppose, for now, that for any $U \in \mathscr{U}$ and time t, there exists some $c \in \mathscr{C}_t$ such that $c_t = U_t$. Then, if the utility U is homothetic and in certainty equivalent form, it must be homogeneous of degree one;[5] that is, $U(kc) = kU(c)$ for all $k > 0$. In the converse direction, a utility that is homogeneous of degree one is necessarily homothetic. With this motivation, we henceforth focus on positive-valued utilities that are homogeneous of degree one. Comparing the BSDEs defining $U(c)$ and $U(kc)$ shows that homogeneity of degree one is essentially equivalent to the following aggregator restriction, which is directly assumed throughout the remainder of this paper:

Condition 18 (Standing assumption). Utility processes are strictly positive (that is, $\mathscr{U} \subseteq \mathscr{L}(\mathbb{R}_{++})$), and the aggregator F takes the *homothetic form*

$$F(\omega, t, c, U, \Sigma) = UG(\omega, t, c/U, \Sigma/U), \qquad F(T, c) = c, \tag{23}$$

for some function $G : \Omega \times [0, T] \times (0, \infty) \times \mathbb{R}^d \to \mathbb{R}$, that we will refer to as a *proportional aggregator*.

We adopt the following convenient change of variables throughout

$$c_t^U = \frac{c_t}{U_t} \quad \text{and} \quad \sigma_t^U = \frac{\Sigma_t}{U_t}, \tag{24}$$

in terms of which the dynamics of $U = U(c)$ can be written as

$$\frac{\mathrm{d}U_t}{U_t} = -G(t, c_t^U, \sigma_t^U)\,\mathrm{d}t + \sigma_t^{U\prime}\,\mathrm{d}B_t. \tag{25}$$

[5] To see this, fix any $c \in \mathscr{C}$, $t \in [0, T]$, and $k \in \mathbb{R}_{++}$. Select $\bar{c} \in \mathscr{C}_t$ such that $\bar{c}_t = U_t(c)$. Since U in certainty equivalent form, $U_t(\bar{c}) = \bar{c}_t = U_t(c)$. Using homotheticity and the fact that $k\bar{c} \in \mathscr{C}_t$, we conclude $U_t(kc) = U_t(k\bar{c}) = k\bar{c}_t = kU_t(c)$.

The partial derivative of G with respect to consumption is $G_c(\omega, t, c^U, \sigma^U) = F_c(\omega, t, c, U, \Sigma)$. The superdifferential of G at $(\omega, t, c^U, \sigma^U) \in \Omega \times [0, T] \times \mathbb{R}_+ \times \mathbb{R}^d$ with respect to volatility is the set, $(\partial_\sigma G)(\omega, t, c^U, \sigma^U)$, consisting of all $a \in \mathbb{R}^d$ such that, for any $z \in \mathbb{R}^d$, $G(\omega, t, c^U, \sigma^U + z) \leqslant G(\omega, t, c^U, \sigma^U) + a'z$. If the partial derivative, G_σ, of G with respect to volatility exists at $(\omega, t, c^U, \sigma^U)$, then $(\partial_\sigma G)(\omega, t, c^U, \sigma^U) = \{G_\sigma(\omega, t, c^U, \sigma^U)\}$.

Lemma 19. *Given any $(\omega, t, c, U, \Sigma) \in \Omega \times [0, T] \times \mathbb{R}_+^2 \times \mathbb{R}^d$, let $c^U = c/U$ and $\sigma^U = \Sigma/U$. Then $(a, b) \in (\partial_{U,\Sigma} F)(\omega, t, c, U, \Sigma)$ if and only if*

$$a = G(\omega, t, c^U, \sigma^U) - G_c(\omega, t, c^U, \sigma^U)c^U - b'\sigma^U \quad \text{and} \quad b \in (\partial_\sigma G)(\omega, t, c^U, \sigma^U).$$

A corollary of the above lemma is that concavity of $F(\omega, t, \cdot)$ is equivalent to concavity of $G(\omega, t, \cdot)$. Our standing assumption that F is regular, therefore, implies that G is *regular*, in the sense that, for all $(\omega, t, c^U, \sigma^U) \in \Omega \times [0, T] \times (0, \infty) \times \mathbb{R}^d$, $G(\cdot, c^U, \sigma^U)$ is progressively measurable, $G(\omega, t, \cdot)$ is concave, and $G_c(\omega, t, \cdot, \sigma^U)$ exists and maps $(0, \infty)$ onto $(0, \infty)$.

Finally, given any processes $(c, U, \Sigma) \in \mathscr{C} \times \mathscr{U} \times \mathscr{L}_2(\mathbb{R}^d)$, we define the set $(\partial_\sigma G)(c^U, \sigma^U)$ to consist of all processes $G_\sigma \in \mathscr{L}_2(\mathbb{R}^d)$ such that $(F_U, F_\Sigma) \in (\partial_{U,\Sigma} F)(t, c_t, U_t, \Sigma_t)$, where

$$F_U(t) = G(t, c_t^U, \sigma_t^U) - G_c(t, c_t^U, \sigma_t^U)c_t^U - G'_\sigma(t)\sigma_t^U, \quad F_\Sigma(t) = G_\sigma(t). \tag{26}$$

In particular, the restriction $G_\sigma \in (\partial_\sigma G)(c^U, \sigma^U)$ implies that $\mathscr{E} = \mathscr{E}(F_U, F_\Sigma)$ satisfies the regularity condition (17).

4.2. Examples of homothetic recursive utility

Homothetic Duffie–Epstein utilities are characterized as follows:

Proposition 20. *The homothetic form (23) defines a Duffie–Epstein aggregator F if and only if*

$$G(\omega, t, c, \sigma) = g(\omega, t, c) - \frac{\gamma(\omega, t)}{2}\sigma'\sigma, \tag{27}$$

for some functions $g : \Omega \times [0, T] \times (0, \infty) \to \mathbb{R}$ and $\gamma : \Omega \times [0, T] \to \mathbb{R}$.

Given representation (27), the standing assumption of a regular G is equivalent to the following restrictions: $g(\cdot, c)$ and γ are progressively measurable; $g(\omega, t, \cdot)$ is concave and $\gamma(\omega, \cdot)$ is non-negative; and $g_c(\omega, t, \cdot)$ maps $(0, \infty)$ onto $(0, \infty)$.

Given the Duffie–Epstein proportional aggregator (27), with γ differentiable, the ordinally equivalent utility version with dynamics (13) is given by

$$V_t = \frac{U_t^{1-\gamma_t} - 1}{1 - \gamma_t}, \quad t \in [0, T],$$

where, for $\gamma_t = 1$, the above expression is to be interpreted as the natural logarithm of U_t. In particular, Ito's lemma implies that (13) is satisfied with

$$f(t,c,V) = h_t(V)^{1-\gamma_t} g\left(t, \frac{c}{h_t(V)}\right) + \frac{\dot{\gamma}_t}{(1-\gamma_t)}(h_t(V)^{1-\gamma_t} \log(h_t(V)) - V), \quad (28)$$

where $\dot{\gamma}_t$ stands for the derivative of γ, and

$$h_t(V) = (1 + (1-\gamma_t)V)^{1/(1-\gamma_t)}.$$

For $\gamma_t = 1$, we interpret the last expression as $h_t(V) = \exp(V)$. (It is worth noting that the renormalized aggregator f need not be concave, even though G is assumed concave.)

Imposing homotheticity to an additive utility results in the familiar HARA type utilities:

Proposition 21. *The Duffie–Epstein aggregator f combines the additive representation (14) with the homothetic representation (28) if and only if $\gamma_t = \gamma$ is a deterministic constant, and there exist processes a and b, such that*

$$g(t,c) = a_t + b_t \frac{c^{1-\gamma} - 1}{1-\gamma}, \quad c > 0, \ \ t \in [0,T],$$

in which case, $u = g$, $\beta_t = b_t - (1-\gamma)a_t$, and b and γ are strictly positive.

Finally, the argument used in the proof of Proposition 20 shows that the homothetic case of the Chen and Epstein (2002) specification of Eq. (15) corresponds to a proportional aggregator of the form

$$G(\omega,t,c,\sigma) = g(\omega,t,c) - \frac{\gamma(\omega,t)}{2}\sigma'\sigma - \max_{\theta \in \Theta(\omega,t)} \theta'\sigma.$$

4.3. First-order conditions under homotheticity

The key to simplifying the first-order conditions under homotheticity is the homogeneity of the utility function. Suppose that (ψ,ρ) is an optimal strategy with corresponding wealth process W and utility process U. Recalling the interpretation of the process λ as the sensitivity of the optimal time-t utility value on time-t wealth, we conjecture that

$$U_t = \lambda_t W_t, \quad (29)$$

and therefore, using integration by parts, $\sigma_t^U = \sigma_t^\lambda + \sigma_t^R \psi_t$. These relationships uncouple the forward and backward components of the first-order conditions.

The first-order conditions will be stated in terms of the functions $\mathscr{I}^G, G^* : \Omega \times [0,T] \times (0,\infty) \times \mathbb{R}^d \to \mathbb{R}$, defined by (omitting state variables)

$$G_c(t, \mathscr{I}^G(t,\lambda,\sigma), \sigma) = \lambda,$$

$$G^*(t,\lambda,\sigma) = \max_{c>0}(G(t,c,\sigma) - \lambda c) = G(t, \mathscr{I}^G(t,\lambda,\sigma),\sigma) - \mathscr{I}^G(t,\lambda,\sigma)\lambda.$$

Note that, at the optimum, $\lambda_t = G_c(t, c_t^U, \sigma_t^U)$ and $\rho_t = c_t^U \lambda_t$, implying

$$G^*(t, \lambda_t, \sigma_t^U) = G(t, \rho_t/\lambda_t, \sigma_t^U) - \rho_t.$$

Direct computation, using Lemma 19 and the above discussion, shows that the equation for λ at the optimum becomes a constrained BSDE:

Condition 22 (First-order conditions: homothetic case). The processes $(\lambda, \sigma^\lambda) \in \mathscr{L}(\mathbb{R}_{++}) \times \mathscr{L}_2(\mathbb{R}^d)$ and the trading strategy ψ solve

$$\frac{\mathrm{d}\lambda_t}{\lambda_t} = -(r_t + G^*(t, \lambda_t, \sigma_t^\lambda + \sigma_t^R \psi_t) + \psi_t'(\mu_t^R + \sigma_t^{R\prime} \sigma_t^\lambda))\,\mathrm{d}t + \sigma_t^{\lambda\prime}\,\mathrm{d}B_t, \quad \lambda_T = 1,$$

$$\varepsilon_t = \mu_t^R + \sigma_t^{R\prime}(G_\sigma(t) + \sigma_t^\lambda), \quad \psi_t \in K, \quad \psi_t' \varepsilon_t = \delta_K(\varepsilon_t), \quad t < T,$$

$$G_\sigma \in (\partial_\sigma G)(\mathscr{I}^G(\lambda, \sigma^\lambda + \sigma^R \psi), \sigma^\lambda + \sigma^R \psi).$$

The characterization of optimality in the homothetic case is summarized in the following result, whose proof (in the appendix) includes a rigorous justification of the conjectured relationship (29). (The counterpart to Remark 16 applies for the case of no intermediate consumption.)

Theorem 23. *Suppose the utility is specified in Conditions* 11 and 18. *In each part below*, (F_U, F_Σ) *is specified in Eqs.* (26).

(a) (*Sufficiency*) *Suppose Condition* 22 *holds, and* $\rho_t = \lambda_t \mathscr{I}^G(t, \lambda_t, \sigma_t^\lambda + \sigma_t^R \psi_t)$. *If* (ψ, ρ) *finances* $c \in \mathscr{C}$, $\lambda W^{\psi,\rho} \in \mathscr{U}$, *and* $\mathscr{E}(F_U, F_\Sigma)\lambda \in \mathscr{H}$, *then the strategy* (ψ, ρ) *is optimal and* $U(c) = \lambda W^{\psi,\rho}$.

(b) (*Necessity*) *Suppose* (ψ, ρ) *is an optimal strategy financing* $c \in \mathscr{C}$, *let* (U, Σ) *and* (c^U, σ^U) *be defined by* (11) *and* (24), *respectively, and let* $\lambda = G_c(c^U, \sigma^U)$. *Suppose further that* $G(\omega, t, \cdot)$ *is differentiable for all* (ω, t), ρ *is continuous,* $G_\sigma \in (\partial_\sigma G)(c^U, \sigma^U)$, *and* $\mathscr{E}(F_U, F_\Sigma)\lambda \in \mathscr{H}$. *Then Condition* 22 *is satisfied with* $\sigma^\lambda = \sigma^U - \sigma_t^R \psi_t$, *and* $U = \lambda W^{\psi,\rho}$.

Remark 24. The homotheticity assumption causes the Markovian PDE characterization of the first-order conditions to simplify significantly. In the Markovian setting of Section 3.5, we express λ and ψ as functions of the state variable: $\lambda_t = \lambda(t, Z_t)$ and $\psi_t = \psi(t, Z_t)$. Using subscripts to denote partial derivatives, Ito's lemma implies the following PDE version of the first-order conditions:

$$0 = r + G^* + \psi'(\mu^R + \sigma^{R\prime}\sigma^\lambda) + \frac{\lambda_t}{\lambda} + \frac{\lambda_Z'}{\lambda}\mu^Z + \frac{1}{2} tr\left(\frac{\lambda_{ZZ}}{\lambda}\sigma^{Z\prime}\sigma^Z\right),$$

$$\lambda(T, \cdot, \cdot) = 1, \quad \sigma^\lambda(t, z) = \frac{\lambda_Z(t, z)}{\lambda(t, z)}\sigma^Z(t, z),$$

$$\varepsilon = \mu^R + \sigma^{R\prime}(G_\sigma + \sigma^\lambda), \quad \psi \in K, \quad \psi'\varepsilon = \delta_K(\varepsilon),$$

where $\mathscr{I}^G$ and G^* are computed at $(t, \lambda, \sigma^\lambda + \sigma^R \psi)$, and G_σ is computed at $(t, \mathscr{I}^G, \sigma^\lambda + \sigma^R \psi)$.

4.4. Some special optimal strategies

In this subsection we outline formulations that result in particularly simple consumption or trading strategies. First we consider a preference specification that results in a prescribed consumption strategy independently of the investment opportunity set, extending the familiar logarithmic utility analysis. We then provide conditions resulting in instantaneously mean–variance efficient trading strategies in complete or incomplete markets, again for any specification of price dynamics. These results generalize Theorem 2 and Section 7 of Schroder and Skiadas (1999), the latter being a continuous-time extension of the discrete-time example of Giovannini and Weil (1989).

4.4.1. A robustly optimal consumption strategy

Let β be a strictly positive and (for simplicity) bounded (progressively measurable) process. We are interested in determining the class of proportional aggregators G for which the following condition holds (given G and β):

Condition 25. For any price dynamics specification (r, μ^R, σ^R), and any trading constraint convex set K, the optimal consumption strategy is $\rho = \beta$.

Recalling that at the optimum $\lambda_t = G_c(t, c_t^U, \sigma_t^U)$ and $\rho_t = c_t^U \lambda_t$, the functional restriction that corresponds to the above condition is

$$c^U G_c(t, c^U, \sigma^U) = \beta(t), \quad t < T, \tag{30}$$

for any optimum (c^U, σ^U), given any price dynamics and constraint set.

For this, it is sufficient that Eq. (30) holds for all $(c^U, \sigma^U) \in (0, \infty) \times \mathbb{R}^d$, in which case G assumes the functional form:

$$G(t, c^U, \sigma^U) = \beta(t) \log(c^U) + H(t, \sigma^U), \quad t < T, \tag{31}$$

for some function H. Given this specification of G, the conclusion of Theorem 23(a) can be strengthened to include the validity of Condition 25. (We leave the statement of the simplified first-order conditions to the reader.)

If we further assume Duffie–Epstein preferences, then H must take the form $H(t, \sigma^U) = -\gamma_t \sigma^{U\prime} \sigma^U / 2$ for some positive process γ, providing a measure of comparative risk aversion that does not affect preferences over deterministic plans. If one further imposes additivity, one must set $\gamma_t = 1$, in which case $\log U$ is the usual time-additive expected discounted logarithmic utility with discount process β. In the additive case, risk aversion is fully determined by preferences over deterministic plans.

Finally, we outline an argument showing that representation (31) is essentially necessary for Condition 25 (eschewing uninteresting technicalities). Consider any positive process c^U and $\mathbb{R}^d$-valued process σ^U such that U is well-defined by Eq. (25) with $U_T = w_0$, and $c_t = c_t^U U_t$, $t < T$, $c_T = w_0$, defines a consumption plan c. Consider the market with $r = c/w_0$ and $K = \{0\}$ (no risky asset trading). In such a market the consumption plan c is optimal, as can be confirmed from the first-order conditions with $\lambda = U/w_0$, $\psi = 0$, $\rho = r$, and $W = w_0$. If Condition 25 holds, it must therefore be the case that Eq. (30) is satisfied for the essentially arbitrarily chosen (c^U, σ^U).

4.4.2. Mean–variance efficient trading strategies

The trading strategy, ψ, is instantaneously *mean-variance efficient* (MVE) if there exists $k \in \mathscr{L}(\mathbb{R})$ such that

$$\psi = k(\sigma^{R\prime}\sigma^R)^{-1}\mu^R.$$

Unit relative risk aversion toward marketed risk. We consider the incomplete markets formulation of Section 2.5 (modeling approach A), including the return dynamics normalization (8). The proportional aggregator is assumed to be of the form

$$G(t,c,\sigma) = g(t,c,\sigma_N) - \tfrac{1}{2}\,\sigma_M'\sigma_M,$$

and therefore $G_\sigma' = [-\sigma_M', g_{\sigma_N}']$. The first-order conditions imply the optimal strategy

$$\psi_M = (\sigma_{MM}^{R\prime}\sigma_{MM}^R)^{-1}\hat{\mu}_M^R = (\sigma_{MM}^R)^{-1}\eta_M,$$

which is MVE in a market with modified expected instantaneous excess returns $\hat{\mu}_M^R$, reflecting the shadow price of the constraint $\psi_M \in K_M$. If $K_M = \mathbb{R}^m$, then $\hat{\mu}^R = \mu^R$, and the trading strategy ψ_M is MVE relative to the actual expected returns. The complete markets case is obtained by letting $m = d$ and $N = \emptyset$.

Deterministic investment opportunity set. Suppose that the processes (r, μ^R, σ^R) are deterministic, and G takes the state-independent homothetic Duffie–Epstein form $G(\omega,t,c,\sigma) = g(t,c) - \gamma(t)\sigma'\sigma/2$. Letting $\sigma^\lambda = 0$, the BSDE characterizing λ reduces to an ordinary differential equation. Assuming the latter has a (necessarily deterministic) solution λ, then the optimal consumption strategy, $\rho_t = \mathscr{I}^g(t,\lambda_t)$, is also deterministic. Moreover, the expression for $\hat{\mu}^R = \mu^R - \varepsilon$ becomes

$$\hat{\mu}_t^R = -\sigma_t^{R\prime} G_\sigma(t, c_t^U, \sigma_t^R\psi_t) = \gamma_t\sigma_t^{R\prime}\sigma_t^R\psi_t.$$

Rearranging we obtain the optimal trading strategy

$$\psi_t = \frac{1}{\gamma_t}(\sigma_t^{R\prime}\sigma_t^R)^{-1}\hat{\mu}_t^R,$$

which is MVE in the fictitious complete market with expected excess returns $\hat{\mu}^R$. If $K = \mathbb{R}^n$, corresponding to complete markets if $n = d$, and incomplete markets otherwise, then $\hat{\mu}^R = \mu^R$, and the resulting optimal strategy is MVE.

4.5. Optimal consumption dynamics

In this subsection we assume the proportional aggregator form

$$G(\omega,t,c^U,\sigma^U) = g(t,c^U) + H(\omega,t,\sigma^U),$$

for some state-independent function $g : [0,T] \times (0,\infty) \to \mathbb{R}$, and we derive the optimal consumption dynamics. The aggregator form includes all state-independent Duffie–Epstein homothetic aggregators, as well as "quasi-quadratic" extensions discussed in the following section.

We consider an optimal trading strategy (ψ,ρ), financing consumption c, and with associated utility and wealth process U and W, respectively. The Ito decomposition of

λ is denoted as in (20). Using the shorthand notation $G(t) = G(t, c_t^U, \sigma_t^U)$, and defining the processes:

$$\mathscr{S}_t = -\frac{g_{ct}(t, c_t^U)}{g_c(t, c_t^U)}, \quad \mathscr{R}_t = -\frac{c_t^U g_{cc}(t, c_t^U)}{g_c(t, c_t^U)}, \quad \text{and} \quad \mathscr{P}_t = -\frac{c_t^U g_{ccc}(t, c_t^U)}{g_{cc}(t, c_t^U)},$$

we show below that the dynamics of optimal consumption are

$$\frac{dc_t}{c_t} = \mu_t^c \, dt + \sigma_t^{c\prime} \, dB_t, \quad \text{where}$$

$$\mu_t^c = -G(t) - \frac{1}{\mathscr{R}_t}\left(\mathscr{S}_t + \mu_t^\lambda + \sigma_t^{\lambda\prime}\sigma_t^c + \frac{1}{\mathscr{R}_t}\left(1 - \frac{1}{2}\mathscr{P}_t\right)\sigma_t^{\lambda\prime}\sigma_t^\lambda\right),$$

$$\sigma_t^c = \left(1 - \frac{1}{\mathscr{R}_t}\right)\sigma_t^\lambda + \sigma_t^W. \tag{32}$$

Alternatively, the expected instantaneous consumption growth rate can be expressed in terms of the expected instantaneous wealth growth rate $\mu^W = r + \psi'\mu^R - \rho$. Since $U = \lambda W$, we have $\mu^\lambda + \sigma^{\lambda\prime}\sigma^W = -(G + \mu^W)$, which in conjunction with the above equations results in the expression:

$$\mu_t^c = -\left(1 - \frac{1}{\mathscr{R}_t}\right)G(t) - \frac{1}{\mathscr{R}_t}\left(\mathscr{S}_t - \mu_t^W + \left(1 - \frac{1}{2}\frac{\mathscr{P}_t}{\mathscr{R}_t}\right)\sigma_t^{\lambda\prime}\sigma_t^\lambda\right).$$

Finally, to confirm Eqs. (32), suppose that $dc_t^U / c_t^U = a_t \, dt + b_t' \, dB_t$, and combine the Ito expansion of $\lambda_t = g_c(t, c_t^U)$ with Eq. (20), to obtain

$$a_t = -\frac{1}{\mathscr{R}_t}\left(\mu_t^\lambda + \mathscr{S}_t - \frac{1}{2}\frac{\mathscr{P}_t}{\mathscr{R}_t}\sigma_t^{\lambda\prime}\sigma_t^\lambda\right) \quad \text{and} \quad b_t = -\frac{1}{\mathscr{R}_t}\sigma_t^\lambda.$$

On the other hand, since $c = c^U U$ and $U = \lambda W$, we have

$$\mu^c = a - G + b'(\sigma^\lambda + \sigma^W) \quad \text{and} \quad \sigma^c = b + \sigma^\lambda + \sigma^W.$$

Substituting the above expressions for a and b, results in Eqs. (32).

4.6. The dynamic programming approach

The homothetic specification lends itself to the application of the dynamic programming principle, without the usual assumption of an underlying Markov structure. The dynamic programming verification argument is presented in this subsection, and is compared to the utility gradient approach.

To motivate the Bellman equation, we informally consider the time-t problem of the agent with wealth level w. The homogeneity of the utility function implies that the agent's time-t value function is of the form $J_t(w) = \lambda_t w$, for a strictly positive process λ (with $\lambda_T = 1$). Consider now a candidate optimal strategy $(\hat{\psi}, \hat{\rho})$, with wealth process $\hat{W} = W^{\hat{\psi}, \hat{\rho}}$, and financing consumption plan $\hat{c} = \hat{\rho}\hat{W}$, whose utility process is $\hat{U} = U(\hat{c})$. Optimality implies that $\hat{U}_t = J_t(\hat{W}_t) = \lambda_t \hat{W}_t$, and therefore $\hat{c}_t / \hat{U}_t = \hat{\rho}_t / \lambda_t$, for all t. Denoting the Ito decomposition of λ as in Eq. (20), and applying Ito's lemma to $\hat{U} = \lambda\hat{W}$, we obtain

$$\frac{d\hat{U}_t}{\hat{U}_t} = (\mu_t^\lambda + r_t - \hat{\rho}_t + \hat{\psi}_t'\mu_t^R + \hat{\psi}_t'\sigma_t^{R\prime}\sigma_t^\lambda)\,dt + (\sigma_t^\lambda + \sigma_t^R\hat{\psi}_t)'\,dB_t.$$

Comparing to the utility dynamics (25), we conclude that, at the optimum,

$$\mu_t^\lambda + r_t - \hat{\rho}_t + \hat{\psi}_t'(\mu_t^R + \sigma_t^{R\prime}\sigma_t^\lambda) + G\left(t, \frac{\hat{\rho}_t}{\lambda_t}, \sigma_t^\lambda + \sigma_t^R\hat{\psi}_t\right) = 0. \tag{33}$$

Given any other feasible strategy (ψ, ρ), the usual dynamic programming argument leads to the conjecture

$$\mu_t^\lambda + r_t - \rho_t + \psi_t'(\mu_t^R + \sigma_t^{R\prime}\sigma_t^\lambda) + G\left(t, \frac{\rho_t}{\lambda_t}, \sigma_t^\lambda + \sigma_t^R\psi\right) \leqslant 0. \tag{34}$$

Recalling the definition of $(\partial_{U,\Sigma}F)(c, U, \Sigma)$ (and associated regularity restriction (17)) in Section 3.3, we have the following result, which is proved in the appendix using a traditional (but non-Markovian) dynamic programming verification argument.

Theorem 26. *Suppose the utility is specified in Conditions* 11 *and* 18, *and* λ *is a strictly positive Ito process with decomposition* (20) *and terminal value* $\lambda_T = 1$. *Suppose also that for every feasible strategy* (ψ, ρ), $\lambda W^{\psi,\rho} \in \mathscr{U}$ *and* $(\partial_{U,\Sigma}F)$ $(\rho W^{\psi,\rho}, \lambda W^{\psi,\rho}, \lambda W^{\psi,\rho}(\sigma^\lambda + \sigma^R\psi))$ *is non-empty. If the feasible strategy* $(\hat{\psi}, \hat{\rho})$ *satisfies Eq.* 33, *while any other feasible strategy* (ψ, ρ) *satisfies inequality* (34), *then* $\left(\hat{\psi}, \hat{\rho}\right)$ *is optimal.*

Conditions (33) and (34) are implied by the Bellman equation

$$\max_{x>0,\, y\in K} \left\{\mu_t^\lambda + r_t - x + y'(\mu_t^R + \sigma_t^{R\prime}\sigma_t^\lambda) + G\left(t, \frac{x}{\lambda_t}, \sigma_t^\lambda + \sigma_t^R y\right)\right\} = 0, \tag{35}$$

with the optimal $\hat{\rho}_t$ and $\hat{\psi}_t$ providing (almost everywhere) the maximizing arguments. One can easily check that the first-order conditions characterizing the Bellman equation correspond to Condition 22. We confirm this claim assuming, for the sake of brevity, smoothness of $G(\omega, t, \cdot)$. (The analogous argument applies using the superdifferential of G.) Maximization with respect to x at $\hat{\rho}_t > 0$ gives $\hat{\rho}_t = \mathscr{I}^G(t, \lambda_t, \sigma_t^\lambda + \sigma_t^R\hat{\psi}_t)\lambda_t$, which when substituted into Eq. (33), gives the drift of $\mathrm{d}\lambda/\lambda$ as

$$\mu_t^\lambda = -(r_t + G^*(t, \hat{\rho}_t/\lambda_t, \sigma_t^\lambda + \sigma_t^R\hat{\psi}_t) + \hat{\psi}_t'(\mu_t^R + \sigma_t^{R\prime}\sigma_t^\lambda)).$$

The gradient of the expression in the Bellman equation with respect to y is

$$\varepsilon_t = \mu_t^R + \sigma_t^{R\prime}(\sigma_t^\lambda + G_\sigma(t, \hat{\rho}_t/\lambda_t, \sigma_t^\lambda + \sigma_t^R\hat{\psi}_t)).$$

Maximization with respect to $y \in K$ at ψ_t implies that, for every $y \in K$, $(y - \hat{\psi}_t)'\varepsilon_t \leqslant 0$, or, equivalently, $\hat{\psi}_t'\varepsilon_t = \delta_K(\varepsilon_t)$. We have therefore recovered the constrained FBSDE for λ of Condition 22.

The complexity of the dynamic programming verification argument is comparable to that of the proof of sufficiency using the utility gradient approach. The utility gradient approach begins by applying Ito's lemma to $\pi_t W_t$, while the dynamic programming verification argument applies Ito's lemma to $\lambda_t W_t$. At the optimum the two quantities are related by $\pi_t W_t = \mathscr{E}_t \lambda_t W_t$. Finally, the same type of comparison argument completes the proof in both cases, delivering the gradient inequality in the utility gradient approach, and the comparison of the utility value and the value function in the dynamic programming approach.

5. Quasi-quadratic homothetic utility

The analysis of homothetic recursive utility continues in this section, under the special assumption of a quasi-quadratic proportional aggregator, which includes all homothetic Duffie–Epstein aggregators, the Chen–Epstein "κ-ignorance" specification, as well as the criteria in Anderson et al. (2000), Maenhout (1999), and Uppal and Wang (2002).

5.1. Utility specification

In the remainder of this paper we assume the following condition, using the notation

$$|x|' = (|x_1|, \ldots, |x_d|), \qquad x \in \mathbb{R}^d.$$

Condition 27 (Standing assumption). The proportional aggregator G is *quasi-quadratic*, meaning that it takes the form

$$G(\omega,t,c,\sigma) = g(\omega,t,c) - q(\omega,t)'\sigma - \kappa(\omega,t)'|\sigma| - \tfrac{1}{2}\,\sigma' Q(\omega,t)\sigma, \tag{36}$$

for some (progressively measurable) functions $g : \Omega \times [0,T] \times (0,\infty) \to \mathbb{R}$, $q : \Omega \times [0,T] \to \mathbb{R}^d$, $\kappa : \Omega \times [0,T] \to \mathbb{R}^d_+$, and $Q : \Omega \times [0,T] \to \mathbb{R}^{d\times d}$, such that $Q(\omega,t)$ is symmetric positive definite for all (ω,t). Finally, the processes κ, q, and Q are assumed bounded (for simplicity).

Regularity of G implies that $g(\omega,t,\cdot)$ is differentiable and concave, and $g_c(\omega,t,\cdot)$ maps $(0,\infty)$ onto $(0,\infty)$.

If the first term, g, of the quasi-quadratic representation is state-independent, it is completely determined by the agent's preferences over deterministic consumption plans and the assumption that the utility is in certainty-equivalent form. The remaining coefficients of G can then be used to adjust the agent's attitude towards risk, without modifying preferences over deterministic plans.

The linear term, $q'\sigma$, of G can be thought of as reflecting the agent's beliefs. To see that, given any process b, consider a Girsanov change of measure, summarized in Eqs. (1). The utility and excess return dynamics can be written in terms of the Brownian motion B^b under the probability P^b, as

$$\frac{\mathrm{d}U_t}{U_t} = -\left(g_t(c_t^U) - (q_t - b_t)'\sigma_t^U - \kappa(t)'|\sigma_t^U| - \frac{1}{2}\,\sigma_t^{U\prime} Q_t \sigma_t^U \right) \mathrm{d}t + \sigma_t^{U\prime}\,\mathrm{d}B_t^b,$$

$$\mathrm{d}R_t = (\mu_t^R - \sigma_t^{R\prime} b_t)\,\mathrm{d}t + \sigma_t^{R\prime}\,\mathrm{d}B_t^b.$$

Changing beliefs from P to P^b has the effect of modifying the linear term of G from q to $q-b$, and the instantaneous expected excess returns from μ^R to $\mu^R - \sigma^{R\prime} b$. Setting $b=q$, reduces the problem to the case with $q=0$.

The remaining coefficients, Q and κ, of G determine the concavity of the proportional aggregator with respect to utility volatility. The Q term provides a measure of second-order relative risk aversion, while the κ term provides a measure of first-order relative risk aversion. (A static version of these notions is discussed by Segal and

Spivak (1990).) The Q and κ terms allow risk aversion to be dependent on the source of risk. Varying risk attitudes toward different sources of risk can be thought of as reflecting the ambiguity of the risk source in the sense of the well-known experiments by Ellsberg (1961). (Related ideas appear in independent work by Klibanoff et al. (2002) and Uppal and Wang (2002).) The coefficients κ and Q can also arise from multiple prior formulations, as shown in Chen and Epstein (2002) and Skiadas (2003).

Given the source-dependence of risk aversion, the quasi-quadratic representation depends on the choice of a Brownian motion rotation. More specifically, suppose $\Phi \in \mathscr{L}(\mathbb{R}^{d\times d})$ satisfies $\Phi_t'\Phi_t = I_d$, and define $\bar{B}_t = \int_0^t \Phi_s' \, \mathrm{d}B_s$, which is a standard Brownian motion (under the original measure P). Defining $\bar{\sigma}_t^R = \Phi_t'\sigma_t^R$, $\bar{\sigma}_t^U = \Phi_t'\sigma_t^U$, $\bar{q}_t = \Phi_t' q_t$, and $\bar{Q}_t = \Phi_t' Q_t \Phi_t$, the utility and excess return dynamics become

$$\frac{\mathrm{d}U_t}{U_t} = -\left(g_t(c_t^U) - \bar{q}_t'\bar{\sigma}^U - \kappa_t'|\Phi_t\bar{\sigma}_t^U| - \frac{1}{2}\,\bar{\sigma}_t^{U\prime}\bar{Q}_t\bar{\sigma}_t^U \right)\mathrm{d}t + \bar{\sigma}_t^{U\prime}\,\mathrm{d}\bar{B}_t,$$

$$\mathrm{d}R_t = \mu_t^R\,\mathrm{d}t + \bar{\sigma}_t^{R\prime}\,\mathrm{d}\bar{B}_t.$$

The rotation Φ_t can always be chosen so that $\bar{Q}_t$ is diagonal, with the (positive) eigenvalues of Q_t on the diagonal indicating the coefficients of second-order relative risk aversion to sources of risk as measured by the elements of $\bar{\sigma}^U$. On the other hand, the (also positive) elements of κ measure coefficients of first-order risk aversion to sources of risk as measured by σ^U. A Duffie–Epstein proportional aggregator, by Proposition 20, corresponds to $q = \kappa = 0$ and $Q = \gamma I$ for some process γ, in which case $\bar{q} = q$ and $\bar{Q} = Q$. The Duffie–Epstein specification is, therefore, invariant to Brownian motion rotations.

5.2. Utility examples

The following two examples include the parametric homothetic utility forms analyzed by Schroder and Skiadas (1999), and used earlier in asset pricing applications by Duffie and Epstein (1992a) and Duffie et al. (1997). Through the type of argument given in Skiadas (2003), the examples also include the robust-control criteria used in Anderson et al. (2000), Hansen et al. (2001), Maenhout (1999), and Uppal and Wang (2002).

Example 28 (Log-quasi-quadratic aggregator). Suppose the recursive utility V is well-defined by

$$V_t = E_t\left[\int_t^T \mathrm{e}^{-\int_t^s \beta_u\,\mathrm{d}u}\left(a_s\log(c_s) - \kappa_s'|\Sigma_s| + \frac{1}{2}\,\Sigma_s' A_s \Sigma_s\right)\mathrm{d}s + \mathrm{e}^{-\int_t^T \beta_u\,\mathrm{d}u} V_T\right],$$

with terminal value $V_T = a_T\log(c_T)$, where Σ is the diffusion coefficient of V, the processes $a > 0$ and β are deterministic, while the processes $A \in \mathscr{L}(\mathbb{R}^{d\times d})$ and $\kappa \in \mathscr{L}(\mathbb{R}_+^d)$ can be stochastic. Defining

$$D_t = \int_t^T \mathrm{e}^{-\int_t^s \beta_u\,\mathrm{d}u} a_s\,\mathrm{d}s + \mathrm{e}^{-\int_t^T \beta_u\,\mathrm{d}u} a_T,$$

the ordinally equivalent utility $U_t = \exp(V_t/D_t)$ has a quasi-quadratic proportional aggregator:

$$\frac{\mathrm{d}U_t}{U_t} = -\left[\frac{a_t}{D_t}\log\left(\frac{c_t}{U_t}\right) - \kappa_t'|\sigma_t^U| - \frac{1}{2}\,\sigma_t^{U\prime}(I - D_tA_t)\sigma_t^U\right]\mathrm{d}t + \sigma_t^{U\prime}\,\mathrm{d}B_t,$$

with terminal value $U_T = c_T$. Regularity of the proportional aggregator requires that $I - D_tA_t$ be positive definite for all t. (In the parametric case considered in Schroder and Skiadas (1999) this corresponds to the assumption $\alpha \leqslant \beta$.) As shown in Section 4.4.1, the optimal consumption strategy under the above specification is $\rho_t = a_t/D_t$, for any specification of (r, μ^R, σ^R) and K. It is worth noting that if all the parameters in the original specification of V are time-independent, the proportional aggregator of U and the optimal consumption strategy ρ are both time-dependent, since D is time-dependent.

Example 29 (Power-quasi-quadratic aggregator). This example includes the continuous-time version of Epstein and Zin (1989) recursive utility, and time-additive HARA utility. Suppose the utility V is well-defined by

$$V_t = E_t\left[\int_t^T \mathrm{e}^{-\int_t^s \beta_u\,\mathrm{d}u}\left(a_s\,\frac{c_s^{1-\delta}}{1-\delta} - \kappa_s'|\Sigma_s| + \frac{1}{2}\,\frac{\Sigma_s'A_s\Sigma_s}{V_s}\right)\mathrm{d}s + \mathrm{e}^{-\int_t^T \beta_u\,\mathrm{d}u}\,\frac{c_T^{1-\delta}}{1-\delta}\right],$$

where Σ is the diffusion coefficient of V, δ is a scalar such that $0 < \delta \neq 1$, and $a \in \mathscr{L}(\mathbb{R}_{++})$, $\beta \in \mathscr{L}(\mathbb{R})$, $A \in \mathscr{L}(\mathbb{R}^{d\times d})$, and $\kappa \in \mathscr{L}(\mathbb{R}_+^d)$ can be stochastic. The ordinally equivalent utility process $U_t = ((1-\delta)V_t)^{1/(1-\delta)}$ is in the quasi-quadratic class:

$$\frac{\mathrm{d}U_t}{U_t} = -\left(\frac{a_t}{1-\delta}\left(\frac{c_t}{U_t}\right)^{1-\delta} - \frac{\beta_t}{1-\delta} - \kappa_t'|\sigma_t^U| - \frac{1}{2}\sigma_t^{U\prime}\left[\delta I - (1-\delta)A_t\right]\sigma_t^U\right)\mathrm{d}t$$
$$+\,\sigma_t^{U\prime}\,\mathrm{d}B_t, \qquad U_T = c_T.$$

Concavity of the aggregator requires that $\delta I - (1-\delta)A_t$ be positive definite for all t. Schroder and Skiadas (1999) compute the optimal strategy in complete markets for the Duffie–Epstein case of this utility specification. Assuming further time-additivity ($A = 0$), they derive a closed-form expression for λ at the optimum, as an expectation of a forward-looking integral involving only the problem primitives, a result that can be easily confirmed from the dynamics of λ in this section.

5.3. First-order conditions

With quasi-quadratic utility, we obtain the simplifications

$$\mathscr{I}^G(\omega, t, \lambda, \sigma) = \mathscr{I}^g(\omega, t, \lambda), \quad \text{and}$$

$$G^*(\omega, t, \lambda, \sigma) = g^*(\omega, t, \lambda) - q(\omega, t)'\sigma - \kappa(\omega, t)'|\sigma| - \tfrac{1}{2}\,\sigma' Q(\omega, t)\sigma,$$

where the functions $\mathscr{I}^g, g^* : \Omega \times [0, T] \times (0, \infty) \to (0, \infty)$ are defined analogously to $\mathscr{I}^G$ and G^* by

$$g_c(\omega, t, \mathscr{I}^g(\omega, t, \lambda)) = \lambda, \quad \text{and}$$

$$g^*(\omega, t, \lambda) = \max_{c>0}\,(g(t, c) - \lambda c) = g(t, \mathscr{I}^g(t, \lambda)) - \mathscr{I}^g(t, \lambda)\lambda.$$

To characterize the superdifferential $\partial_\sigma G$, we define, for any $\sigma \in \mathscr{L}_2(\mathbb{R}^d)$, the set $\partial|\sigma|$ of all processes Λ that are valued in $[-1,1]^d$, and whose ith coordinate, Λ_i, satisfies

$$\Lambda_i(t) = 1 \text{ on } \{\sigma_i(t) > 0\}, \quad \Lambda_i(t) = -1 \text{ on } \{\sigma_i(t) < 0\},$$
$$\text{and } \Lambda_i(t) \in [-1,1] \text{ on } \{\sigma_i(t) = 0\}.$$

Moreover, for every $x, y \in \mathbb{R}^d$, $x \otimes y \in \mathbb{R}^d$ denotes element-by-element multiplication, that is, $(x \otimes y)_i = x_i y_i$ for all i. The calculation of the superdifferential of G below follows easily from the definitions.

Lemma 30. *$G_\sigma \in (\partial_\sigma G)(c^U, \sigma^U)$ if and only if $G_\sigma(t) = -(q_t + \kappa_t \otimes \Lambda_t + Q_t \sigma_t^U)$ for some $\Lambda \in \partial|\sigma^U|$.*

Direct computation using the above lemma shows:

Proposition 31. *For the quasi-quadratic proportional aggregator G, the constrained BSDE of Condition 22 is equivalent to*:

$$\frac{\mathrm{d}\lambda_t}{\lambda_t} = -\left(r_t + \delta_K(\varepsilon_t) + g^*(t,\lambda_t) - \frac{1}{2}\sigma_t^{\lambda\prime} Q_t \sigma_t^\lambda + \frac{1}{2}\psi_t' \sigma_t^{R\prime} Q_t \sigma_t^R \psi_t\right)\mathrm{d}t$$
$$+ \sigma_t^{\lambda\prime}(\mathrm{d}B_t + (q_t + \kappa_t \otimes \Lambda_t)\,\mathrm{d}t), \quad \lambda_T = 1,$$

$$\psi_t = (\sigma_t^{R\prime} Q_t \sigma_t^R)^{-1}(\mu_t^R - \varepsilon_t - \sigma_t^{R\prime}[(q_t + \kappa_t \otimes \Lambda_t) + (Q_t - I)\sigma_t^\lambda])$$
$$\Lambda \in \partial|\sigma^\lambda + \sigma^R \psi|, \quad \psi_t \in K, \quad \psi_t' \varepsilon_t = \delta_K(\varepsilon_t), \quad t < T. \tag{37}$$

Remark 32. If $K = \mathbb{R}^n$, $n \leqslant d$, corresponding to incomplete ($n < d$) or complete markets ($n = d$), the first-order conditions simplify by letting $\varepsilon = 0$. Another simplification arises if the processes $(r, \mu^R, \sigma^R, q, \kappa, Q)$ are deterministic, and g is state-independent, in which case $\sigma^\lambda = 0$.

Theorem 23 provides conditions for the sufficiency and necessity of the above conditions for optimality. Moreover, at the optimum,

$$\rho_t = \mathscr{I}^g(t,\lambda_t)\lambda_t \quad \text{and} \quad g^*(t,\lambda_t) = g\left(t, \frac{\rho_t}{\lambda_t}\right) - \rho_t.$$

As with any recursive utility that is homogeneous of degree one, the optimal utility and wealth processes are related by Eq. (29).

We close this section with two examples with $\kappa = 0$, while first-order risk aversion is discussed in the following subsection. The first example unifies and extends the parametric solutions of Schroder and Skiadas (1999), who assume Duffie–Epstein homothetic utility of either the logarithmic or power form defined in Section 5.2. The second example considers linear-constraints that include the case of a borrowing constraint as a proportion of wealth.

Example 33 (Complete markets). Suppose $n=d$ and the quasi-quadratic proportional aggregator is smooth ($\kappa=0$), and let $\hat{r}=r+\delta_K(\varepsilon)$ and $\eta=(\sigma^{R\prime})^{-1}(\mu^R-\varepsilon)$. Then the dynamics of λ and optimal trading strategy can be written as

$$\frac{\mathrm{d}\lambda_t}{\lambda_t}=-\left(\hat{r}_t+g^*(t,\lambda_t)+\frac{1}{2}(\eta_t-q_t+\sigma_t^\lambda)'Q_t^{-1}(\eta_t-q_t+\sigma_t^\lambda)-\sigma_t^{\lambda\prime}\sigma_t^\lambda\right)\mathrm{d}t$$

$$+\sigma_t^{\lambda\prime}\mathrm{d}B_t^\eta,\qquad \lambda_T=1;\ \ \psi_t=(Q_t\sigma_t^R)^{-1}(\eta_t-q_t+(I-Q_t)\sigma_t^\lambda).$$

If the market is complete ($K=\mathbb{R}^d$), then $\varepsilon=0$ and $r=\hat{r}$. For general K, $(\hat{r},\eta)$ represent the short rate and price of risk in the fictitious complete market of Corollary 9.

Example 34 (Linear constraints). Given a smooth ($\kappa=0$) quasi-quadratic proportional aggregator, a particularly simple expression for the optimal trading strategy is obtained if $K=\{k\in\mathbb{R}^n: \alpha\leqslant l'k\leqslant\beta\}$ where $l\in R^n$ and α and β are valued in $[-\infty,+\infty]$. The case of no short-selling of asset i corresponds to $\alpha=0$, $\beta=\infty$, and l a vector of zeros, except for a one in the ith position. The case of a cap on the proportion of wealth borrowed, possibly combined with a limit on short sales as a fraction of wealth, corresponds to letting l be a vector of ones. We assume that K is non-empty, and define

$$\psi_t^*=A_t(\mu_t^R-\sigma_t^{R\prime}q_t+\sigma_t^{R\prime}(I-Q_t)\sigma_t^\lambda),\qquad A_t=(\sigma_t^{R\prime}Q_t\sigma_t^R)^{-1}.$$

The above expression gives the optimal trading strategy as a function of σ^λ in the unconstrained case ($\alpha=-\infty$, $\beta=\infty$). The (constrained) optimal trading strategy ψ and process ε in the dynamics of λ are given by

$$\psi_t=\psi_t^*-A_t\varepsilon_t,\qquad \varepsilon_t=-(l'A_tl)^{-1}l(\min\{\max\{l'\psi_t^*,\alpha\},\beta\}-l'\psi_t^*). \tag{38}$$

This follows from the more general problem of p linear constraints, $K=\{k\in\mathbb{R}^n: L'k\leqslant b\}$, where $L\in R^{n\times p}$ and $b\in R^p$. The set K is assumed non-empty. Eq. (37) implies $\psi_t=\psi_t^*-A_t\varepsilon_t$. The condition $\psi_t'\varepsilon_t=\delta_K(\varepsilon_t)$ is a linear program, whose solution is characterized by the first-order conditions of optimality

$$\varepsilon_t=Lv_t,\qquad \delta_t=b-L'\psi_t\geqslant 0,\qquad v_t\geqslant 0,\qquad v_t'\delta_t=0,$$

for a process v, valued in $\mathbb{R}^m$. Substituting the expressions for ψ and ε, we obtain the linear complementarity problem

$$\delta_t=b-L'\psi_t^*+L'A_tLv_t\geqslant 0,\qquad v_t\geqslant 0,\ \ v_t'\delta_t=0.$$

The special solution (38) follows by letting $L=[-l,l]$ and $b'=[-\alpha,\beta]$.

5.4. Optimal trading and first-order risk aversion

This subsection concerns the effects of first-order risk aversion ($\kappa\neq 0$) on optimal trading, and can be skipped by the reader interested in the smooth case only. For tractability, we assume throughout the subsection that Q is diagonal. The setting and notation is that of Section 2.5, Modeling Approach A, unless otherwise indicated. The main result follows.

Proposition 35. *Suppose that Q is diagonal, there are $m = n \leqslant d$ risky assets traded, and the excess return dynamics take the normalized form* (8). *Then the optimal trading strategy ψ in the first-order conditions is given by*

$$\psi = (\sigma^R_{MM})^{-1}(Q^{-1}_{MM}\min\{\max\{0, \alpha_M - \kappa_M\}, \alpha_M + \kappa_M\} - \sigma^\lambda_M),$$

where

$$\alpha_M = \eta_M - q_M + \sigma^\lambda_M, \quad \eta_M = (\sigma^{R\prime}_{MM})^{-1}(\mu^R_M - \varepsilon_M). \tag{39}$$

The term $\kappa \otimes \Lambda$ in the BSDE for λ can be set to

$$\kappa \otimes \Lambda = \begin{pmatrix} \max\{\min\{\alpha_M, \kappa_M\}, -\kappa_M\} \\ \kappa_N \otimes \operatorname{sign}(\sigma^\lambda_N) \end{pmatrix}.$$

Proof. Given the normalized form (8), Proposition 31 implies

$$\psi = (Q_{MM}\sigma^R_{MM})^{-1}(\alpha_M - \kappa_M \otimes \Lambda_M - Q_{MM}\sigma^\lambda_M).$$

Since $\sigma^U = \sigma^\lambda + \sigma^R\psi$, the above expression for ψ is equivalent to $Q_{ii}\sigma^U_i = \alpha_i - \kappa_i\Lambda_i$, $i \in M$. Since $\Lambda_i \in \partial|\sigma^U_i|$, there are three possibilities: (1) $\sigma^U_i > 0$, $\Lambda_i = 1$, $\alpha_i > \kappa_i$; (2) $\sigma^U_i = 0$, $\alpha_i = \kappa_i\Lambda_i$; and (3) $\sigma^U_i < 0$, $\Lambda_i = -1$, $\alpha_i < -\kappa_i$. In all three cases, $\kappa_M \otimes \Lambda_M = \max\{\min\{\alpha_M, \kappa_M\}, -\kappa_M\}$. Substituting back in the above expression for ψ proves the proposition's first claim. Given the normalization (8), $\sigma^U_N = \sigma^\lambda_N$, and therefore $\sigma^{\lambda\prime}_N(\kappa_N \otimes \Lambda_N) = \sigma^{\lambda\prime}_N(\kappa_N \otimes \operatorname{sign}(\sigma^\lambda_N))$, confirming the proposition's last claim. □

Remark 36. For any $i \in M$, when $\alpha_i \in [-\kappa_i, \kappa_i]$, $\sigma^U_i = 0$; that is, the agent completely hedges utility risk in the ith direction. Such perfect hedging is not encountered with second-order risk aversion alone.

Remark 37. Suppose $Q = I$, and $K = \mathbb{R}^n$, $n \leqslant d$ (so markets can be incomplete). Applying the first-order conditions with $\varepsilon = 0$ shows that the optimal portfolio, ψ_t, is instantaneously mean–variance efficient when $\kappa = q = 0$; or $\kappa = -q$ and $\eta_M \geqslant -\sigma^\lambda_M$; or $\kappa = q$ and $\eta_M \leqslant -\sigma^\lambda_M$.

The following example generalizes Section 5.3 of Chen and Epstein (2002) (who assume $q = 0$, $Q = (1-\alpha)I$, $\alpha \leqslant 1$, and $0 \leqslant \kappa < |\eta_M|$, $\varepsilon = 0$).

Example 38 (Deterministic investment opportunity set). Suppose that r, μ^R, σ^R, κ, q, Q, are all deterministic, and the function g is state independent. Suppose also that $K = \mathbb{R}^n$, $n \leqslant d$, meaning markets can be incomplete, but there are no further restrictions. Then the BSDE for λ reduces to an ODE by setting $\sigma^\lambda = 0$. Assuming the ODE has a (necessarily deterministic) solution λ, the optimal trading strategy is given by Proposition 35 with $\varepsilon = 0$ and $\sigma^\lambda = 0$. Let us assume, for simplicity, that σ^R_{MM} is diagonal with positive diagonal, and let $i \in M$. Then $\psi_i = 0$ when $\eta_i - q_i \in [-\kappa_i, +\kappa_i]$; the agent will not assume a position (long or short) unless the subjective price of marketed risk is sufficiently far from zero. Suppose further that $q_i = -\kappa_i$, then $\psi_i = Q^{-1}_{ii}\mu^R_i/(\sigma^R_{ii})^2$

when $\mu_i^R > 0$ (just as with $\kappa = q = 0$), but the agent will only short asset i when $\mu_i^R < -2\kappa_i \sigma_{ii}^R$.

We conclude with another example of market non-participation as a result of first-order risk aversion. Thinking of B_M as generating "domestic" uncertainty, and B_N as generating "foreign" uncertainty, the following example shows that first-order risk aversion toward foreign uncertainty can have the same effect as if foreign uncertainty were not traded altogether. A closely related two-person equilibrium parametric example (with no domestic constraints) is given by Epstein and Miao (2000).

Example 39 (Risk aversion induced non-participation). The setting is that of Section 2.5, with $m < n = d$ and σ^R taking the canonical form (10). We further assume that Q is diagonal, the processes r, q_M, η_M, Q_{MM}, σ_{MM}^R, and the function g are adapted to the filtration generated by B_M, and that κ satisfies

$$\kappa_M = 0 \quad \text{and} \quad -\kappa_N \leqslant (\sigma_{NN}^{R\prime})^{-1}\mu_N - q_N \leqslant \kappa_N.$$

Consider the market in the first m securities with the filtration generated by B_M alone; in other words, the market that ignores the existence of information source B_N and securities $m+1, \ldots, d$. In this market, the first-order conditions of optimality are (omitting time indices):

$$\frac{\mathrm{d}\lambda}{\lambda} = -\left(\hat{r} + g^*(t,\lambda) - \frac{1}{2}\,\sigma_M^{\lambda\prime} Q_{MM} \sigma_M^{\lambda} + \frac{1}{2}\,\psi_M' \sigma_{MM}^{R\prime} Q_{MM} \sigma_{MM}^{R} \psi_M\right) \mathrm{d}t$$

$$+\,\sigma_M^{\lambda\prime}(\mathrm{d}B_M + q_M dt), \qquad \lambda_T = 1,$$

$$\psi_M = (\sigma_{MM}^R)^{-1}(Q_{MM}^{-1}\alpha_M - \sigma_M^{\lambda}) \in K_M, \qquad \psi_M' \varepsilon_M = \delta_{K_M}(\varepsilon_M),$$

with α_M given in Eq. (39) and $\hat{r} = r + \delta_{K_M}(\varepsilon_M)$.

Consider now the larger market obtained from the first one by revealing information source B_N and allowing unrestricted trading in all asset in N; that is, the filtration is generated by B and the constraint set is $K = \{k \in \mathbb{R}^d : k_M \in K_M\}$. Using Proposition 35, one can easily confirm that a solution $(\psi_M, \lambda, \sigma_M^{\lambda})$ to the above conditions, together with $\psi_N = 0$ and $\sigma_N^{\lambda} = 0$, is also a solution to the first-order conditions of optimality in this larger market. To the extent that the first-order conditions are necessary for optimality in the first market and sufficient for optimality in the second market, it follows that the optimal trading strategy in the first market is also optimal in the second market.

5.5. Links between complete and incomplete markets

In this subsection we present two connections between incomplete and complete markets solutions, given a smooth ($\kappa = 0$) quasi-quadratic proportional aggregator.

The setting and notation are those of Section 2.5, with trading in the money market and the first m risky assets, under the constraint $\psi_M \in K_M$. Excess returns of the traded assets are assumed to follow the normalized dynamics (8). For simplicity of exposition, we also assume that Q takes the block-diagonal form

$$Q = \begin{bmatrix} Q_{MM} & 0 \\ 0 & Q_{NN} \end{bmatrix}.$$

Following Modeling Approach A of Section 2.5, and applying Proposition 31, the first-order conditions of optimality in this context are as follows.

Proposition 40. *Suppose that $n = m$, $K = K_M$, $\sigma^{R\prime} = [\sigma^{R\prime}_{MM}, 0]$, $Q_{NM} = 0$, and $\kappa = 0$. Then the constrained BSDE of Condition 22 is equivalent to*

$$\begin{aligned}
\frac{\mathrm{d}\lambda}{\lambda} = &-\left(r + \delta_{K_M}(\varepsilon_M) + g^*(t,\lambda) - \frac{1}{2}\,\sigma^{\lambda\prime}_N Q_{NN}\sigma^{\lambda}_N\right)\mathrm{d}t + \sigma^{\lambda\prime}_N(\mathrm{d}B_N + q_N\,\mathrm{d}t)\\
&-\left(\frac{1}{2}\,(\eta_M - q_M + \sigma^{\lambda}_M)' Q^{-1}_{MM}(\eta_M - q_M + \sigma^{\lambda}_M) - \sigma^{\lambda\prime}_M\sigma^{\lambda}_M\right)\mathrm{d}t\\
&+\sigma^{\lambda\prime}_M(\mathrm{d}B_M + \eta_M\,\mathrm{d}t), \qquad \lambda_T = 1,
\end{aligned}$$

$$\begin{aligned}
&\psi_M = (Q_{MM}\sigma^R_{MM})^{-1}(\eta_M - q_M + (I - Q_{MM})\sigma^{\lambda}_M),\\
&\psi_M \in K_M, \quad \psi'_M\varepsilon_M = \delta_{K_M}(\varepsilon_M), \quad \eta_M = (\sigma^{R\prime}_{MM})^{-1}(\mu^R_M - \varepsilon_M).
\end{aligned} \tag{40}$$

The following example illustrates the alternative Modeling Approach B of Section 2.5 and Corollary 9, by showing how to derive an incomplete markets solution from the complete markets solution of Example 33.

Example 41 (Fictitious market completion). Using Modeling Approach B of Section 2.5, let $n = d$ and $K = \{k \in \mathbb{R}^n : k_M \in K_M,\ k_N = 0\}$, and therefore $\delta_K(\varepsilon) = \delta_{K_M}(\varepsilon_M)$. Given that the last $d - m$ assets are not traded, we assume, without loss in generality, that

$$\sigma^R = \begin{bmatrix} \sigma^R_{MM} & 0 \\ 0 & I_{d-m} \end{bmatrix}.$$

Let $\hat{r} = r + \delta_K(\varepsilon)$ and $\eta = (\sigma^{R\prime})^{-1}(\mu^R - \varepsilon)$ be the short rate and market price of risk, respectively, of the fictitious complete market of Corollary 9. Given the assumption $Q_{MN} = 0$, the unconstrained optimal trading strategy, given $\hat{r}$ and η, can be computed from Example 33 to be given by Eq. (40) and

$$\psi_N = Q^{-1}_{NN}(\eta_N - q_N + (I - Q_{NN})\sigma^{\lambda}_N).$$

Setting $\eta_N = q_N - (I - Q_{NN})\sigma^{\lambda}_N$ implies the unconstrained optimal demand $\psi_N = 0$, and therefore the non-tradeability of the last $d - m$ assets becomes a non-binding constraint. Substituting this expression for η_N into the BSDE for λ of Example 33, we recover the BSDE for λ of the last Proposition.

A different type of connection between incomplete and complete markets solutions is obtained by assigning an arbitrary price to non-marketed risk, and suitably modifying beliefs and second-order risk aversion with respect to non-marketed risk. This result requires the additional assumption that $2I - Q_{NN}$ is positive definite. For example, in the Duffie–Epstein case, $Q = \gamma I$, we assume γ is valued in $(0, 2)$ (an example being time-additive HARA utility with coefficient of relative risk aversion $\gamma \in (0, 2)$). Given this condition, the following proposition shows that the optimal strategy given trading only in the first m assets (possibly under constraints) can be characterized in terms of the solution obtained by introducing unrestricted trading of assets in $N = \{m+1, \ldots, d\}$, assigning any value to η_N, and, instead of G, using the quasi-quadratic proportional aggregator:

$$\bar{G} = g - \frac{1}{2}(\eta_N - q_N)'(2I - Q_{NN})^{-1}(\eta_N - q_N) - \bar{q}'\sigma - \frac{1}{2}\sigma'\bar{Q}\sigma,$$

$$\bar{Q} = \begin{bmatrix} Q_{MM} & 0 \\ 0 & (2I - Q_{NN})^{-1} \end{bmatrix}, \quad \bar{q} = \begin{bmatrix} q_M \\ \eta_N - (2I - Q_{NN})^{-1}(\eta_N - q_N) \end{bmatrix}. \tag{41}$$

Note that the specification of $\bar{G}$ is not dependent on the original underlying price dynamics. For simplicity of exposition, we assume the sufficiency and necessity of the first-order conditions for optimality, referring to Theorem 23 for the relevant qualifications.

Proposition 42. *Suppose $m < d$, $Q_{MN} = 0$, $2I - Q_{NN}$ is positive definite, $\kappa = 0$, σ^R takes the canonical form* (10), *$\bar{G}$ is defined in Eq.* (41) *given any* (*say bounded*) *$\eta_N \in \mathscr{L}(\mathbb{R}^{d-m})$, and*

$$\frac{\bar{W}_t}{W_t} = \frac{\bar{U}_t}{U_t} = \exp\left(\int_0^t \psi_N'(s)\sigma_{*N}^{R\prime}(s)\left(\begin{bmatrix} -\sigma_{MM}^R(s)\psi_M(s) \\ \eta_N(s) \end{bmatrix} \mathrm{d}s + \mathrm{d}B_s\right)\right).$$

Then the following two statements are equivalent (*assuming the sufficiency and necessity of the first-order conditions*):

1. *The strategy (ψ_M, ρ) is optimal with $n = m$, proportional aggregator G, and constraint set $K = K_M \subseteq \mathbb{R}^m$. The corresponding optimal wealth and utility processes are W and U, respectively.*
2. *The strategy (ψ, ρ) is optimal with $n = d$, proportional aggregator $\bar{G}$, and constraint set $K = \{k \in \mathbb{R}^d : k_M \in K_M\}$. The corresponding optimal wealth and utility processes are $\bar{W}$ and $\bar{U}$, respectively.*

Proof. For each part, we can apply the first-order conditions of Proposition 40, to verify by direct calculation that the dynamics of λ are identical for both problems, as is the optimal trading strategy, ψ_M. The equality of the ratios $\bar{W}/W$ and $\bar{U}/U$ follows from the homogeneity of both problems, and finally the exponential expression for $\bar{W}/W$ follows from the budget equations in the two problems with the common value of ψ_M. $\square$

5.6. Incomplete markets and quadratic BSDEs

We conclude the main part of this paper with an incomplete markets application in which the BSDE characterizing $\log(\lambda)$ is quadratic. Under a suitable set of assumptions, we show that the quadratic BSDE reduces to an ODE system of the Riccati type. The technique can be applied either in terms of BSDEs or in terms of PDEs, as in the affine term-structure literature (see Duffie et al., 2003; Piazzesi, 2002). This section's results extend the complete markets calculations with log-Duffie–Epstein utility of the last section of Schroder and Skiadas (1999), and the stochastic-volatility incomplete-markets model of Chacko and Viceira (1999).[6] In the case of time-additive utility for terminal wealth only, the results extend those of Liu (2001), who in turn generalizes Kim and Omberg (1996). In complete markets, a similar reduction of the solution to a set of ODEs can be obtained under time-additive power utility, as shown by Schroder and Skiadas (1999) and Liu (2001).[7]

The following key condition is assumed throughout this section. The utility specification includes that of Example 28 with $\kappa = 0$.

Condition 43. Markets can be incomplete ($n \leqslant d$), but there are no further constraints ($K = \mathbb{R}^n$), the quasi-quadratic proportional aggregator is smooth ($\kappa = 0$), with $q = 0$, and

$$g(\omega, t, c) = \beta(\omega, t) + y(t) + y(t)\log\left(\frac{c}{y(t)}\right),$$

for some process $\beta \in \mathscr{L}(\mathbb{R})$ and some non-negative *deterministic* process y.

The assumption $q = 0$ is without loss of generality (given the change of measure argument of Section 5.1). The function $y\log(c/y)$ is defined to take the value zero for $y = 0$. Time-additive utility for terminal wealth only is obtained by letting $y = 0$ and $Q_t = \gamma I$, for some positive constant γ. In this case, the ordinally equivalent utility process $V_t = (1-\gamma)^{-1}U_t^{1-\gamma}$ is given by

$$V_t = E_t\left(\mathrm{e}^{\int_t^T (1-\gamma)\beta_s\,\mathrm{d}s}\,\frac{c_T^{1-\gamma}}{1-\gamma}\right).$$

This is the case, assuming constant β, studied by Liu (2001).

Making the convenient change of variables

$$\ell_t = \log(\lambda_t),$$

we note the above condition implies that $g^*(t, \lambda) = \beta_t - y_t\ell_t$. Direct computation then shows that, in the first-order conditions, ℓ solves a quadratic BSDE.

---

[6] Chacko and Viceira (1999) also provides an approximate solution for the case of power Duffie–Epstein utility.

[7] The solution in Wachter (2002) is a special case of Schroder and Skiadas (1999) in the case of intermediate consumption, and of Liu (2001) in the case of terminal consumption.

Proposition 44. *Given Condition* 43, *the BSDE of the first-order conditions* (37) *is equivalent to*

$$\mathrm{d}\ell_t = -\left(p_t - y_t\ell_t + \sigma^{\ell\prime}h_t + \frac{1}{2}\sigma_t^{\ell\prime}H_t\sigma_t^{\ell}\right)\mathrm{d}t + \sigma_t^{\ell\prime}\mathrm{d}B_t, \quad \ell_T = 0, \tag{42}$$

where

$$p_t = r_t + \beta_t + \tfrac{1}{2}\mu_t^{R\prime}(\sigma_t^{R\prime}Q_t\sigma_t^R)^{-1}\mu_t^R,$$

$$h_t = (I - Q_t)\sigma_t^R(\sigma_t^{R\prime}Q_t\sigma_t^R)^{-1}\mu_t^R, \quad \textit{and}$$

$$H_t = (I - Q_t)[I + \sigma_t^R(\sigma_t^{R\prime}Q_t\sigma_t^R)^{-1}\sigma_t^{R\prime}(I - Q_t)].$$

Finally, we introduce a set of conditions under which the quadratic BSDE (42) reduces to an ODE system of the Riccati type. We present the results at a formal level without addressing issues of existence (see Duffie et al. (2003) and the references therein).

We introduce a *state process* $Z \in \mathscr{L}(\mathbb{R}^k)$ with dynamics

$$\mathrm{d}Z_t = \mu_t^Z\,\mathrm{d}t + \sigma_t^{Z\prime}\,\mathrm{d}B_t, \quad \text{where } \mu^Z \in \mathscr{L}_1(\mathbb{R}^k),\ \sigma^Z \in \mathscr{L}_2(\mathbb{R}^{d\times k}).$$

Moreover, we split these state variables into two blocks, of dimensionality a and b, respectively, where $a + b = k$. We treat a and b as integers denoting dimensionality, as well as indices of corresponding matrix blocks, writing

$$Z = \begin{bmatrix} Z^a \\ Z^b \end{bmatrix} \quad \text{and} \quad \sigma^Z = [\sigma^{Za}\ \ \sigma^{Zb}],$$

where $Z^a \in \mathscr{L}(\mathbb{R}^a)$, $Z^b \in \mathscr{L}(\mathbb{R}^b)$, $\sigma^{Za} \in \mathscr{L}(\mathbb{R}^{d\times a})$, and $\sigma^{Zb} \in \mathscr{L}(\mathbb{R}^{d\times b})$.

We seek a solution of the form

$$\ell_t = \ell_0(t) + \ell_1(t)'Z_t + \tfrac{1}{2}Z_t^{a\prime}\ell_2(t)Z_t^a, \tag{43}$$

for some *deterministic* processes $\ell_0 \in \mathscr{L}(\mathbb{R})$, $\ell_1 = [\ell_1^{a\prime}, \ell_1^{b\prime}]' \in \mathscr{L}(\mathbb{R}^k)$, and $\ell_2 \in \mathscr{L}(\mathbb{R}^{a\times a})$. A sufficient set of conditions for this type of solution is stated below (omitting time indices). We assume, without loss of generality, that all matrices appearing in quadratic forms, including $\ell_2(t)$ above, are symmetric.

Condition 45. (a) The process σ^{Za} is deterministic.

(b) For some *deterministic* processes $K_0 \in \mathscr{L}(\mathbb{R})$, $K_1 \in \mathscr{L}(\mathbb{R}^k)$, and $K_2 \in \mathscr{L}(\mathbb{R}^{a\times a})$,

$$p = K_0 + K_1'Z + \tfrac{1}{2}Z^{a\prime}K_2Z^a.$$

(c) For some *deterministic* processes $L_0 = [L_0^{a\prime}, L_0^{b\prime}]' \in \mathscr{L}(\mathbb{R}^k)$, $L_1^a \in \mathscr{L}(\mathbb{R}^{a\times a})$, $L_1^b \in \mathscr{L}(\mathbb{R}^{b\times k})$, and $L_2^b[i] \in \mathscr{L}(\mathbb{R}^{a\times a})$,

$$\mu^Z + \sigma^{Z\prime}h = L_0 + \begin{pmatrix} L_1^aZ^a \\ L_1^bZ_t + [Z_t^{a\prime}L_2^b[i]Z^a]_{i=1,\dots,b} \end{pmatrix}.$$

(d) For some *deterministic* processes $D_0^{aa} \in \mathscr{L}(\mathbb{R}^{a\times a})$, $D_0^{ab} \in \mathscr{L}(\mathbb{R}^{a\times b})$, $D_0^{bb} \in \mathscr{L}(\mathbb{R}^{b\times b})$, $D_1^{ab}[i,j] \in \mathscr{L}(\mathbb{R}^a)$, $D_1^{bb}[i,j] \in \mathbb{R}^k$ and $D_2^{bb}[i,j] \in \mathscr{L}(\mathbb{R}^{a\times a})$,

$$\sigma^{Z\prime} H \sigma^Z = \begin{pmatrix} D_0^{aa} & D_0^{ab} \\ D_0^{ab\prime} & D_0^{bb} \end{pmatrix} + \begin{pmatrix} 0 & [Z^{a\prime} D_1^{ab}[i,j]]_{i=1,\dots,a;j=1,\dots,b} \\ [Z^{a\prime} D_1^{ab}[i,j]]'_{i=1,\dots,a;j=1,\dots,b} & [Z' D_1^{bb}[i,j]]_{i,j=1,\dots,b} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & [Z^{a\prime} D_2^{bb}[i,j] Z^a]_{i,j=1,\dots,b} \end{pmatrix}.$$

We let $D_0 \in \mathscr{L}(\mathbb{R}^{k\times k})$ denote the first term of the right-hand side.

Suppose Conditions 43 and 45 hold, $\Xi \in \mathscr{L}(\mathbb{R}^{a\times a})$ is defined to have ith row

$$\Xi_{i*} = \sum_{j=1}^{b} \ell_{1j}^b D_1^{ab}[i,j]',$$

and the deterministic processes (ℓ_0, ℓ_1, ℓ_2) solve the following ODE system (where the left-hand sides denote time-derivatives):

$$\dot{\ell}_0 = y\ell_0 - K_0 - \ell_1' L_0 - \frac{1}{2}\ell_1' D_0 \ell_1 - \frac{1}{2}\operatorname{trace}(\ell_2 \sigma^{Za\prime}\sigma^{Za}),$$

$$\dot{\ell}_1 = y\ell_1 - K_1 - L_1^{b\prime}\ell_1^b - \frac{1}{2}\sum_{i=1}^{b}\sum_{j=1}^{b} \ell_{1i}^b \ell_{1j}^b D_1^{bb}[i,j] - \begin{pmatrix} L_1^{a\prime}\ell_1^a + \sum_{i=1}^{a}\sum_{j=1}^{b} \ell_{1i}^a \ell_{1j}^b D_1^{ab}[i,j] + \ell_2 \begin{pmatrix} D_0^{aa} & D_0^{ab} \end{pmatrix} \ell_1 + \ell_2 L_0^a \\ 0 \end{pmatrix},$$

$$\dot{\ell}_2 = y\ell_2 - K_2 - \ell_2 L_1^a - L_1^{a\prime}\ell_2 - \ell_2 D_0^{aa}\ell_2 - \ell_2 \Xi - \Xi'\ell_2 - \sum_{i=1}^{b}\sum_{j=1}^{b} \ell_{1i}^b \ell_{1j}^b D_2^{bb}[i,j] - 2\sum_{i=1}^{b} \ell_{1i}^b L_2^b[i],$$

with terminal conditions

$$\ell_0(T) = \ell_1(T) = \ell_2(T) = 0.$$

Direct computation using Ito's lemma shows that then Eq. (43) defines a solution to BSDE (42) (provided that the drift and diffusion terms of (42) are suitably integrable so that the respective integrals are well-defined).

Example 46. We assume that Condition 43 holds, and adopt Modeling Approach A of Section 2.5: $n = m \leqslant d$, $K = \mathbb{R}^m$, and $\mathrm{d}R = \mu^R\,\mathrm{d}t + \sigma_{MM}^{R\prime}\,\mathrm{d}B_M$. The price of marketed risk is $\eta_M = (\sigma_{MM}^{R\prime})^{-1}\mu^R$. The following two problem classes satisfy Condition 45 (and extend corresponding complete markets examples in Schroder and Skiadas (1999)).

Class 1. $Z = Z^a$ $(a = k, b = 0)$, Q is a deterministic constant, and

$$\mathrm{d}Z_t = (\mu - \theta Z_t)\,\mathrm{d}t + \Sigma' \mathrm{d}B_t, \qquad \eta_M = \upsilon + V Z_t,$$

$$r_t + \beta_t = y_0 + y_1' Z_t + \tfrac{1}{2} Z_t' y_2 Z_t,$$

where $\mu \in \mathbb{R}^a$, $\Sigma \in \mathbb{R}^{d\times a}$, $\theta \in \mathbb{R}^{a\times a}$, $\upsilon \in \mathbb{R}^m$, $V \in \mathbb{R}^{m\times a}$, $y_0 \in \mathbb{R}$, $y_1 \in \mathbb{R}^a$, $y_2 \in \mathbb{R}^{a\times a}$. An Ornstein–Uhlenbeck market price of risk process is used in Kim and Omberg (1996), and in the variation of the Stein–Stein model examined by Liu (2001), both in the context of an investor maximizing expected power utility of terminal wealth.

Class 2. $Z = Z^b$ $(a = 0, b = k)$, Q is a deterministic and constant diagonal matrix, and

$$\mathrm{d}Z_t = (\mu - \theta Z_t)\,\mathrm{d}t + \Sigma' \mathrm{diag}\left(\sqrt{\upsilon + V Z_t}\right)\,\mathrm{d}B_t,$$

$$\eta_M = \mathrm{diag}\left(\sqrt{\upsilon_M + V_{M*} Z_t}\right)\varphi, \qquad r_t + \beta_t = y_0 + y_1' Z_t,$$

where $\mathrm{diag}(x)$ denotes the diagonal matrix with x on the diagonal, $\sqrt{x}$ denotes the vector with ith element $\sqrt{x_i}$, and $\mu \in \mathbb{R}^b$, $\Sigma \in \mathbb{R}^{d\times b}$, $\theta \in \mathbb{R}^{b\times b}$, $\upsilon = [\upsilon_M', \upsilon_N']' \in \mathbb{R}^d$, $V = [V_{M*}', V_{N*}']' \in \mathbb{R}^{d\times b}$, $\varphi \in \mathbb{R}^m$, $y_0 \in \mathbb{R}$, $y_1 \in \mathbb{R}^b$. In this case, $\ell_2 = 0$, and only the first two Riccati equations are needed, the second unlinked to the first. This formulation includes the square-root volatility process of Heston (1993), and the square-root market-price-of-risk process in Chacko and Viceira (1999).

Acknowledgements

We are thankful for discussions with Domenico Cuoco, Jakša Cvitanić, Darrell Duffie, Julien Hugonnier, Peter Klibanoff, Ali Lazrak, Steve Shreve, George Skoulakis, Tan Wang, and Thaleia Zariphopoulou. We are responsible for any errors.

Appendix A. Proofs

A.1. Proof of Theorem 7

1. *Sufficiency of Condition 4*. Suppose Condition 4 holds, and consider any feasible strategy $(\tilde{\psi}, \tilde{\rho})$ financing consumption plan $\tilde{c}$, and with corresponding wealth process $\tilde{W}$. Integration by parts implies, for any stopping time τ,

$$\pi_\tau \tilde{W}_\tau - \pi_0 w_0 = \int_0^\tau \tilde{W}_t \pi_t \left(\frac{\mathrm{d}\tilde{W}_t}{\tilde{W}_t} + \frac{\mathrm{d}\pi_t}{\pi_t} - \tilde{\psi}_t' \sigma_t^{R\prime} \eta_t\,\mathrm{d}t \right)$$

$$= \int_0^{\tau} \tilde{W}_t \pi_t (\tilde{\psi}_t' \varepsilon_t - \delta_K(\varepsilon_t))\,dt - \int_0^{\tau} \pi_t \tilde{c}_t\,dt + M_\tau^1$$

$$\leqslant - \int_0^{\tau} \pi_t \tilde{c}_t\,dt + M_\tau^1,$$

for a local martingale M^1 (collecting all the diffusion terms in the above expansion). The same argument for $\tilde{\psi} = \psi$ gives (using $\psi'\varepsilon = \delta_K(\varepsilon)$):

$$\pi_\tau W_\tau - \pi_0 w_0 = - \int_0^{\tau} \pi_t c_t\,dt + M_\tau^2,$$

for another local martingale M^2. Letting $x = \tilde{c} - c$, and $M = M^1 - M^2$, it follows that

$$\pi_\tau \tilde{W}_\tau - \pi_\tau \tilde{W}_\tau \leqslant - \int_0^{\tau} \pi_t x_t\,dt + M_\tau.$$

Consider now an increasing sequence of stopping times $\{\tau_n : n = 1, 2, \ldots\}$ that converges to T and such that M stopped at τ_n is a martingale. Taking expectations in the last inequality, we find

$$E\left[\int_0^{\tau_n} \pi_t x_t\,dt + \pi_{\tau_n}(\tilde{W}_{\tau_n} - W_{\tau_n})\right] \leqslant 0. \tag{A.1}$$

Taking the limit inferior on both sides as $n \to \infty$, we conclude $(\pi|x) \leqslant 0$. The interchange of limit and expectation is justified by the assumptions $x, \pi \in \mathscr{H}$, $\pi W \in \mathscr{S}$, dominated convergence, and Fatou's lemma.

2. *Necessity of Condition 4.* By taking as the unit of account the value of one unit invested in the money market at time zero, we assume, without loss of generality, that $r = 0$. (This is shown by letting $\bar{W}_t = W_t \exp(-\int_0^t r_s\,ds)$ and $\bar{\pi}_t = \pi_t \exp(\int_0^t r_s\,ds)$. Then $d\bar{W}_t/\bar{W}_t = (\psi_t' \mu_t^R - \rho_t)\,dt + \psi_t' \sigma_t^{R\prime} dB_t$ and, for any $x \in \mathscr{X}(c)$, $(\pi|x) = (\bar{\pi}|\bar{x})$, where $\bar{x}_t = x_t \exp(-\int_0^t r_s\,ds)$. Passing to the barred quantities, the theorem to be proved becomes the original theorem with $r = 0$.)

Suppose that $(\pi|x) \leqslant 0$ for all $x \in \mathscr{X}(c)$, and that ρ is continuous. The first lemma applies the restriction $(\pi|x) \leqslant 0$ to a suitably chosen $x \in \mathscr{X}(c)$ in order to show that the state-price density must correctly price the optimal consumption plan c (where $c_t = \rho_t W_t$ for $t < T$, and $c_T = W_T$).

Lemma A.1. *For any* $t \in [0, T]$,

$$\pi_t W_t = E_t\left[\int_t^T \pi_s c_s\,ds + \pi_T c_T\right].$$

Proof. Choose any $t \in [0, T]$ and event $F \in \mathscr{F}_t$. We consider a new strategy $(\tilde{\psi}, \tilde{\rho})$, where $\tilde{\psi} = \psi$, but the consumption strategy deviates slightly from ρ for a short time after t on the set F:

$$\tilde{\rho}_s - \rho_s = \begin{cases} \varepsilon \rho_t 1_F, & s \in [t, t + \tau(h)], \\ 0, & \text{otherwise,} \end{cases}$$

where $\varepsilon \in (-1/2, 1/2)$ and the stopping time $\tau(h)$ is defined below. Let $\tilde{W}$ and $\tilde{c}$ be the wealth process and consumption plan, respectively, corresponding to $(\tilde{\psi}, \tilde{\rho})$. It is easy to confirm that

$$\tilde{W}_s - W_s = \begin{cases} 0 & \text{for } s \leqslant t, \\ (\mathrm{e}^{-\varepsilon\rho_t(s-t)} - 1)W_s 1_F & \text{for } t < s < t + \tau(h), \\ (\mathrm{e}^{-\varepsilon\rho_t\tau(h)} - 1)W_s 1_F & \text{for } t + \tau(h) \leqslant s \leqslant T \end{cases}$$

and $\tilde{c} - c = x 1_F$ where

$$x_s = \begin{cases} 0 & \text{for } s \leqslant t, \\ (\mathrm{e}^{-\varepsilon\rho_t(s-t)} - 1)c_s + \tilde{W}_s\varepsilon\rho_t & \text{for } t < s < t + \tau(h), \\ (\mathrm{e}^{-\varepsilon\rho_t\tau(h)} - 1)c_s & \text{for } t + \tau(h) \leqslant s \leqslant T. \end{cases}$$

Now define, for each $h \in (0, T - t)$ and $|\varepsilon| < 1/2$, the stopping time

$$\tau(h) = \min\{s \geqslant t : \tilde{c}_s/c_s \notin [1/2, 2],\ \pi_s \geqslant 2\pi_t,\ \text{or } s - t = h\} - t.$$

By the definition of $\tau(h)$, $\tilde{c}$ is strictly positive and bounded above by $2c$. Based on these observations, and the assumed properties of $\mathscr{C}$, it is not hard to confirm that $\tilde{c} \in \mathscr{C}$, and therefore $x 1_F \in \mathscr{X}(c)$. The condition $(\pi | x 1_F) \leqslant 0$ can be stated as

$$0 \geqslant E \left\{ \begin{array}{l} 1_F \left(\dfrac{1}{h} \displaystyle\int_t^{t+\tau(h)} (\mathrm{e}^{-\varepsilon\rho_t(s-t)} - 1)\pi_s c_s \, \mathrm{d}s + \dfrac{1}{h}\, \varepsilon\rho_t \int_t^{t+\tau(h)} \pi_s \tilde{W}_s \, \mathrm{d}s \right) \\ \quad + 1_F \dfrac{1}{h} (\mathrm{e}^{-\varepsilon\rho_t\tau(h)} - 1) \left(\displaystyle\int_t^T \pi_s c_s \, \mathrm{d}s + \pi_T c_T \right) \end{array} \right\}.$$

Taking the limit as $h \to 0$,

$$0 \geqslant E \left\{ 1_F \varepsilon\rho_t \left(\pi_t \tilde{W}_t - \int_t^T \pi_s c_s \, \mathrm{d}s - \pi_T c_T \right) \right\},$$

and since ε can be positive or negative, we have

$$0 = E \left[1_F \left(\pi_t W_t - \int_t^T \pi_s c_s \, \mathrm{d}s - \pi_T c_T \right) \right].$$

Applied over all $F \in \mathscr{F}_t$ and $t \in [0, T]$, this proves the lemma. $\square$

Suppose now that π has the Ito decomposition

$$\frac{\mathrm{d}\pi_t}{\pi_t} = -\delta_t \, \mathrm{d}t - \eta_t' \, \mathrm{d}B_t.$$

Lemma A.2. $\delta_t = \psi_t' \varepsilon_t$.

Proof. By the last lemma:

$$\pi_t W_t + \int_0^t \pi_s c_s \, \mathrm{d}s = M_t = E_t \left[\int_0^T \pi_s c_s \, \mathrm{d}s + \pi_T c_T \right].$$

Using integration by parts and the dynamics for π and W, we obtain

$$(\psi_t' \varepsilon_t - \delta_t) \, \mathrm{d}t = \frac{\mathrm{d}M_t}{\pi_t W_t} + (\eta_t - \sigma_t^R \psi_t)' \, \mathrm{d}B_t.$$

Since the right-hand-side is a local martingale, both sides must vanish, proving that $\delta_t = \psi_t' \varepsilon_t$. □

The next two lemmas conclude the proof by showing that $k' \varepsilon_t \leqslant \delta_t$ for all $k \in K$. So far, we have only considered perturbations of the optimal consumption strategy ρ. The following lemma considers perturbations in the trading strategy.

Lemma A.3. *Consider any feasible strategy* $(\tilde{\psi}, \tilde{\rho})$, *with corresponding wealth process* $\tilde{W}$ *such that* $\pi \tilde{W} \in \mathscr{S}$ *and* $\tilde{\psi}_t' \varepsilon_t - \delta_t$ *is bounded below. Then*

$$E \left[\int_0^T \pi_t \tilde{W}_t (\tilde{\psi}_t' \varepsilon_t - \delta_t) \, \mathrm{d}t \right] \leqslant 0. \tag{A.2}$$

Proof. Given any stopping time τ, integration by parts implies

$$\begin{aligned} \pi_\tau \tilde{W}_\tau - \pi_0 w_0 &= \int_0^\tau (\pi_t \, \mathrm{d}\tilde{W}_t + \tilde{W}_t \, \mathrm{d}\pi_t - \tilde{W}_t \tilde{\psi}_t' \sigma_t^{R\prime} \eta_t \, \mathrm{d}t) \\ &= \int_0^\tau \tilde{W}_t \pi_t (\tilde{\psi}_t' \varepsilon_t - \delta_t) \, \mathrm{d}t - \int_0^\tau \pi_t \tilde{c}_t \, \mathrm{d}t + \int_0^\tau \pi_t \tilde{W}_t (\sigma_t^R \tilde{\psi}_t - \eta_t)' \, \mathrm{d}B_t. \end{aligned}$$

Letting $x = \tilde{c} - c$, and since $\delta_t = \psi_t' \varepsilon_t$, we obtain

$$\pi_\tau (\tilde{W}_\tau - W_\tau) + \int_0^\tau \pi_t x_t \, \mathrm{d}t = \int_0^\tau \pi_t \tilde{W}_t (\tilde{\psi}_t \varepsilon_t - \delta_t)' \, \mathrm{d}t + M_\tau,$$

where M is the local martingale collecting all the diffusion terms. Consider an increasing sequence of stopping times $\{\tau_n : n = 1, 2, \ldots\}$ that converges to T and such that M stopped at τ_n is a martingale. It then follows that

$$E \left[\pi_{\tau_n} (\tilde{W}_{\tau_n} - W_{\tau_n}) + \int_0^{\tau_n} \pi_t x_t \, \mathrm{d}t \right] \geqslant E \left[\int_0^{\tau_n} \pi_t \tilde{W}_t (\tilde{\psi}_t \varepsilon_t - \delta_t)' \, \mathrm{d}t \right].$$

Taking the limit inferior as $n \to \infty$ on both sides the result follows from the inequality $(\pi | x) \leqslant 0$. The interchange of limits and expectation on the left-hand side is justified by the assumptions $\pi W, \pi \tilde{W} \in \mathscr{S}$ and $\pi, x \in \mathscr{H}$. For the right-hand side we apply Fatou's lemma. □

The last lemma intuitively suggests the following result, whose proof provides the required technicalities to complete the proof of the theorem.

Lemma A.4. *$k'\varepsilon_t \leqslant \delta_t$ for all $k \in K$.*

Proof. For each $N \in \{1, 2, \ldots\}$, define the correspondence

$$F^N(\omega, t) = \{y \in K : |y| \leqslant N, \quad y'\varepsilon(\omega, t) - \delta(\omega, t) \geqslant 1/N\}.$$

We show that the set $S^N = \{(\omega, t) : F^N(\omega, t) \neq \emptyset\}$ is null (meaning that its indicator function is zero as an element of $\mathscr{H}$). Using a measurable selections theorem (see, for example, Klein and Thompson, 1984, Theorem 14.2.1) we define the trading strategy $\tilde{\psi}$ by letting $\tilde{\psi} \in F^N$ on S^N and $\tilde{\psi} = \psi$ on the complement of S^N. Letting $\tilde{W}$ denote the wealth process generated by $(\tilde{\psi}, \rho)$, and given any scalar $b > 1$, we define τ^b to be the minimum of T and the first time that $\tilde{W}$ hits bW. We also define the new trading strategy $\tilde{\psi}^b$ to be equal to $\tilde{\psi}$ up to τ^b, and equal to ψ from τ^b to T. The wealth process generated by $(\tilde{\psi}^b, \rho)$ is denoted $\tilde{W}^b$. By construction, $\rho\tilde{W}^b \leqslant b\rho W$, and therefore $\rho\tilde{W}^b \in \mathscr{C}$ and $(\tilde{\psi}^b, \rho)$ is feasible. Similarly, since $\pi\tilde{W}^b \leqslant b\pi W$ and $\pi W \in \mathscr{S}$, it is also true that $\pi\tilde{W}^b \in \mathscr{S}$. Lemma A.3 therefore implies (using $\psi'\varepsilon = \delta$):

$$E\left[\int_0^{\tau^b} \pi_t \tilde{W}_t^b (\tilde{\psi}_t'\varepsilon_t - \delta_t)\,\mathrm{d}t\right] \leqslant 0.$$

This is consistent with the definition of $\tilde{\psi}$ only if $S^N \cap [0, \tau^b]$ is null. Taking the union over all integers b implies that S^N is null, and taking the union over all N, shows that $\{(\omega, t) : y'\varepsilon(\omega, t) > \delta(\omega, t), \text{ some } y \in K\}$ is null, completing the proof. □

A.2. Proof of Lemma 10

Let b_{*N} be a $d \times (d - m)$-dimensional process whose columns form a basis for the orthogonal linear subspace spanned by the columns of σ_{*M}^R at almost every (ω, t). In other words, b_{*N} is full rank and solves $\sigma_{*M}^{R\prime} b_{*N} = 0$. (We omit here a technical but routine measurable selection argument that ensures that b_{*N} is progressively measurable.) We define

$$\Phi_t = [\sigma_{*M}^R (\sigma_{*M}^{R\prime} \sigma_{*M}^R)^{-1/2} \quad b_{*N} (b_{*N}' b_{*N})^{-1/2}].$$

Direct calculation shows that $\Phi_t' \Phi_t = I_d$. Therefore $\mathrm{d}\bar{B}_t = \Phi_t'\,\mathrm{d}B_t$, $\bar{B}_0 = 0$, defines a standard Brownian motion, as shown by the quadratic variation calculation: $\mathrm{d}\bar{B}_t\,\mathrm{d}\bar{B}_t' = \Phi_t'\,\mathrm{d}B_t\,\mathrm{d}B_t'\,\Phi_t = I_d\,\mathrm{d}t$. Finally, we note that

$$\mathrm{d}R_t = \mu_t^R\,\mathrm{d}t + \bar{\sigma}_t^{R\prime}\,\mathrm{d}\bar{B}_t = \mu_t^R\,\mathrm{d}t + \sigma_t^{R\prime}\,\mathrm{d}B_t$$

if and only if

$$\bar{\sigma}^R = \Phi'\sigma^R = \begin{bmatrix} (\sigma_{*M}^{R\prime}\sigma_{*M}^R)^{1/2} & (\sigma_{*M}^{R\prime}\sigma_{*M}^R)^{-1/2}\sigma_{*M}^{R\prime}\sigma_{*N}^R \\ 0 & (b_{*N}'b_{*N})^{-1/2} b_{*N}'\sigma_{*N}^R \end{bmatrix},$$

which implies

$$\mathrm{d}R_M = \mu_M^R\,\mathrm{d}t + (\sigma_{*M}^{R\prime}\sigma_{*M}^R)^{1/2}\,\mathrm{d}\bar{B}_M.$$

A.3. Proof of Proposition 13

The result follows from the following lemma, by letting $t=0$. The lemma shows a more general conclusion that is used in the necessity part of the proof of Theorem 23. Given any $c\in\mathscr{C}$, $(U(c),\Sigma(c))$ denotes the corresponding solution to BSDE (11).

Lemma A.5. *Given any $t\in[0,T]$ and $c\in\mathscr{C}$, let $U=U(c)$ and $\Sigma=\Sigma(c)$. Suppose $(F_U,F_\Sigma)\in(\partial_{U,\Sigma}F)(c,U,\Sigma)$, and let $\mathscr{E}=\mathscr{E}(F_U,F_\Sigma)$. Then, for any process x such that $c+x\in\mathscr{C}$,*

$$U_t(c+x)\leqslant U_t(c)+E_t\left[\int_t^T\frac{\mathscr{E}_s}{\mathscr{E}_t}F_c(s,c_s,U_s,\Sigma_s)x_s\,\mathrm{d}s+\frac{\mathscr{E}_T}{\mathscr{E}_t}F_c(T,c_T)x_T\right].$$

Proof. Fixing any h such that $c+h\in\mathscr{C}$, we define the processes Δ and δ by

$$\Delta_t=U_t(c+h)-U_t(c),\qquad \delta_t=\Sigma_t(c+h)-\Sigma_t(c).$$

We use the simplified notation $U=U(c)$ and $\Sigma=\Sigma(c)$, $F(t)=F(t,c_t,U_t,\Sigma_t)$. By concavity of F,

$$F(t,c_t+h_t,U_t+\Delta_t,\Sigma_t+\delta_t)=F(t)+(F_c(t)h_t+F_U(t)\Delta_t+F_\Sigma'(t)\delta_t)-\varepsilon_t$$

for some non-negative process ε. Using the last expression and the BSDEs for $U(c+h)$ and $U(c)$, we obtain the BSDE

$$\mathrm{d}\Delta_t=-(F_c(t)h_t+F_U(t)\Delta_t+F_\Sigma'(t)\delta_t-\varepsilon_t)\,\mathrm{d}t+\delta_t'\,\mathrm{d}B_t,\qquad \Delta_T=F_c(T)h_T-\varepsilon_T.$$

Letting $\mathscr{E}=\mathscr{E}(F_U,F_\Sigma)$, we have $\mathrm{d}\mathscr{E}_t=\mathscr{E}_tF_U(t)\,\mathrm{d}t+\mathscr{E}_tF_\Sigma'(t)\,\mathrm{d}B_t$. Integration by parts gives

$$\mathrm{d}(\mathscr{E}_t\Delta_t)=\mathscr{E}_t\,\mathrm{d}\Delta_t+\Delta_t\,\mathrm{d}\mathscr{E}_t+\mathrm{d}\Delta_t\,\mathrm{d}\mathscr{E}_t=-\mathscr{E}_tF_c(t)h_t\,\mathrm{d}t+\mathscr{E}_t\varepsilon_t\,\mathrm{d}t+\mathrm{d}M_t,$$

for a local martingale M. We fix any time t, and let $\{\tau_N : N=1,2,\ldots\}$ be an increasing sequence of stopping times, valued in $[0,T]$, that converges to T almost surely, and such that $\{M_s : s\geqslant t\}$ stopped at τ_N is a martingale. Integrating the above dynamics from t to τ_N, give

$$\mathscr{E}_{\tau_N}\Delta_{\tau_N}-\mathscr{E}_t\Delta_t\geqslant-\int_t^{\tau_N}\mathscr{E}_sF_c(s)h_s\,\mathrm{d}s+M_{\tau_N}-M_t,$$

which implies

$$\mathscr{E}_t\Delta_t\leqslant E\left[\int_t^{\tau_N}\mathscr{E}_sF_c(s)h_s\,\mathrm{d}s+\mathscr{E}_{\tau_N}\Delta_{\tau_N}\right].$$

The proof of the lemma is completed by letting $N\to\infty$, and using dominated convergence and the condition $\mathscr{E}U\in\mathscr{S}$ for all $U\in\mathscr{U}$, which is implicit in the assumption $(F_U,F_\Sigma)\in(\partial_{U,\Sigma}F)(c,U,\Sigma)$. □

A.4. Proof of Lemma 19

Let $F_c = F_c(\omega, t, c, U, \Sigma) = G_c(\omega, t, c^U, \sigma^U) = G_c$, where $c^U = c/U$ and $\sigma^U = \Sigma/U$. Suppose $(F_U, F_\Sigma) \in (\partial_{U,\Sigma} F)(\omega, t, c, U, \Sigma)$, and therefore, given any $x \in (-c, \infty)$, $y \in (-U, \infty)$, and $z \in \mathbb{R}^d$,

$$F(\omega, t, c + x, U + y, \Sigma + z) \leqslant F(\omega, t, c, U, \Sigma) + F_c x + F_U y + F'_\Sigma z. \tag{A.3}$$

Consider any $(\alpha, \beta) \in \mathbb{R}_+ \times \mathbb{R}^d$, and define $G_\sigma = F_\Sigma$, and ε so that

$$F_U = G(\omega, t, c^U, \sigma^U) - G_c c^U - G_\sigma \sigma^U + \varepsilon.$$

Fixing any scalar y such that $U + y > 0$, we define x and z by the equations

$$\frac{c + x}{U + y} = c^U + \alpha \quad \text{and} \quad \frac{\Sigma + z}{U + z} = \sigma^U + \beta.$$

Substituting into Eq. (A.3), and simplifying, we obtain

$$G(\omega, t, c^U + \alpha, \sigma^U + \beta) \leqslant G(\omega, t, c^U, \sigma^U) + G_c \alpha + G'_\sigma \beta + \frac{y}{U + y} \varepsilon.$$

Applying this inequality for zero (α, β) and arbitrary $y > -U$, we conclude that $\varepsilon = 0$, proving both that $F_U = G - G_c \sigma^U - G_\sigma \sigma^U$, and that $G_\sigma \in \partial_\sigma G$. Conversely, if the latter conditions hold, we can reverse the above steps (with $\varepsilon = 0$), to conclude that $(F_U, F_\Sigma) \in \partial_{U,\Sigma} F$.

A.5. Proof of Proposition 20

Suppose F is a Duffie–Epstein aggregator taking the homothetic form (23). Letting $\Sigma = 0$ in Eq. (12), we obtain $b(\omega, t, c, U)/U = G(\omega, t, c/U, 0) \equiv g(\omega, t, c/U)$, which in turn leads to

$$G\left(\omega, t, \frac{c}{U}, \frac{\Sigma}{U}\right) = g\left(\omega, t, \frac{c}{U}\right) + \frac{Ua(\omega, t, U)}{2} \left(\frac{\Sigma}{U}\right)' \frac{\Sigma}{U}.$$

Since we can select the value of $U > 0$ arbitrarily, for any fixed value of c/U and Σ/U, it follows that $a(\omega, t, U)U = \gamma(\omega, t)$ for some function γ, and the result follows.

A.6. Proof of Proposition 21

Suppose the f combines the additive representation (14) with the homothetic representation (28). Then, using $U = h(V)$,

$$f_V = (1 - \gamma_t) g_t \left(\frac{c}{U}\right) - \frac{c}{U} g'_t \left(\frac{c}{U}\right) + \dot{\gamma}_t \log(U) = -\beta_t.$$

Setting $c = U$, it becomes clear that this equation can hold only if $\dot{\gamma} = 0$, and therefore $\beta_t = g'_t(1) - (1 - \gamma_t) g_t(1)$. Differentiating the above expression for f_V with respect to

c, we obtain the further restriction

$$-\frac{xg_t''(x)}{g_t'(x)} = \gamma_t = \gamma, \quad x > 0.$$

If g_t were expected utility, the above states that g_t has relative risk aversion γ, and therefore one obtains the familiar power representation. Substituting back into Eq. (28), we find that $u = g$.

A.7. Proof of Theorem 23

(a) (*Sufficiency*) Condition 14 can be verified by direct computation using Lemma 19. By Theorem 15, optimality follows provided we verify that $\pi W \in \mathscr{S}$. In this context, $\pi W = \mathscr{E}\lambda W = \mathscr{E} U \in \mathscr{S}$, where the last condition follows from (17), implicit in the condition $(F_U, F_\sigma) \in (\partial_{U,\Sigma} F)$, which in turn follows from the assumption $G_\sigma \in (\partial_\sigma G)$.

(b) (*Necessity*) Given the restriction $U = \lambda W$, Condition 14 can be verified directly using Lemma 19, and the result follows by Theorem 15. To confirm that $U = \lambda W$ at the optimum, we utilize the homogeneity of U. For any time t, we define the function $f_t(\alpha) = U_t(c + \alpha c) = (1+\alpha)U_t(c)$, $\alpha > 0$. Letting $\pi = \mathscr{E}\lambda$ be the gradient density at the optimum, Lemmas A.5 and A.1 imply:

$$\mathscr{E}_t f_t'(0) = E_t\left[\int_t^T \mathscr{E}_s \lambda_s c_s \, \mathrm{d}s + \mathscr{E}_T \lambda_T c_T\right] = E_t\left[\int_t^T \pi_s c_s \, \mathrm{d}s + \pi_T c_T\right] = \pi_t W_t.$$

On the other hand, homogeneity of U implies $f'(0) = U(c)$, and therefore $\mathscr{E}U = \pi W = (\mathscr{E}\lambda)W$, which simplifies to $U = \lambda W$.

A.8. Proof of Theorem 26

Consider any feasible strategy (ψ, ρ) financing $c \in \mathscr{C}$, and let $W = W^{\psi,\rho}$, $U = U(c)$, $\Sigma = \Sigma(c)$, $J = \lambda W$, and $\sigma^J = \sigma^\lambda + \sigma^R \psi$. Note that $\rho_t/\lambda_t = c_t/J_t$. We will show that $U_0 \leqslant J_0 = \lambda_0 w_0$, with equality holding if $(\psi, \rho) = (\hat{\psi}, \hat{\rho})$.

Applying integration by parts to $J = \lambda W$,

$$\frac{\mathrm{d}J_t}{J_t} = (\mu_t^\lambda + r_t - \rho_t + \psi_t'(\mu_t + \sigma_t^{R\prime}\sigma_t^\lambda))\,\mathrm{d}t + \sigma_t^{J\prime}\,\mathrm{d}B_t.$$

Let p^1 be the nonnegative (by inequality (34)) process such that

$$\mu_t^\lambda + r_t - \rho_t + \psi_t'(\mu_t^R + \sigma_t^{R\prime}\sigma_t^\lambda) + G\left(t, \frac{\rho_t}{\lambda_t}, \sigma_t^\lambda + \sigma_t^R \psi_t\right) + \frac{p_t^1}{J_t} = 0.$$

Then the above dynamics for J can be written as

$$\mathrm{d}J_t = -[F(t, c_t, J_t, \Sigma_t^J) + p_t^1]\,\mathrm{d}t + \Sigma^{J\prime}\,\mathrm{d}B_t.$$

On the other hand, $\mathrm{d}U_t = -F(t, c_t, U_t, \Sigma_t)\,\mathrm{d}t + \Sigma_t'\,\mathrm{d}B_t$. Letting $y = U - J$, and $z = \Sigma - \Sigma^J$, we therefore have

$$\mathrm{d}y_t = -[F(t, c_t, U_t, \Sigma_t) - F(t, c_t, J_t, \Sigma_t^J) - p_t^1]\,\mathrm{d}t + z_t'\,\mathrm{d}B_t.$$

Selecting $(F_U, F_\Sigma) \in (\partial_{U,\Sigma} F)(c, J, \Sigma^J)$,

$$F(t, c_t, U_t, \Sigma_t) - F(t, c_t, J, \Sigma^J) = F_U(t) y_t + F_\Sigma(t) z_t - p_t^2,$$

for some nonnegative process p^2. Therefore, with $p = p^1 + p^2$,

$$\mathrm{d} y_t = -[F_U(t) y_t + F_\Sigma(t)' z_t - p_t]\,\mathrm{d}t + z_t'\,\mathrm{d}B_t.$$

Letting $\mathscr{E} = \mathscr{E}(F_U, F_\Sigma)$, integration by parts gives

$$\mathrm{d}(\mathscr{E}_t y_t) = \mathscr{E}_t\,\mathrm{d} y_t + y_t\,\mathrm{d}\mathscr{E}_t + \mathrm{d} y_t\,\mathrm{d}\mathscr{E}_t = \mathscr{E}_t p_t\,\mathrm{d}t + \mathrm{d}M_t,$$

for a local martingale M. Let $\{\tau_N\}$ be a corresponding localizing stopping time sequence converging to T almost surely. Then

$$y_0 = E\left[-\int_0^{\tau_N} \mathscr{E}_s p_s\,\mathrm{d}s + \mathscr{E}_{\tau_N} y_{\tau_N}\right] \leqslant E[\mathscr{E}_{\tau_N} U_{\tau_N}] - E[\mathscr{E}_{\tau_N} J_{\tau_N}].$$

The first term converges by the assumption $\mathscr{E}U \in \mathscr{S}$, while we apply Fatou's lemma on the second term to conclude that $y_0 \leqslant E[y_T \mathscr{E}_T] = 0$ (since $y_T = 0$). We have proved that $U_0 \leqslant J_0 = \lambda_0 w_0$.

Repeating the first part of the above argument for $(\psi, \rho) = (\hat{\psi}, \hat{\rho})$ and $p^1 = 0$, gives $U_0 = J_0 = \lambda_0 w_0$, proving optimality of $(\hat{\psi}, \hat{\rho})$.

References

Anderson, E., Hansen, L., Sargent, T., 2000. Robustness, detection and the price of risk. Working Paper, Department of Economics, University of Chicago.

Campbell, J., Viceira, L., 2002. Strategic Asset Allocation. Oxford University Press, New York.

Chacko, G., Viceira, L., 1999. Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets. Working Paper, Harvard University.

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

Cox, J., Huang, C.-F., 1989. Optimal consumption and portfolio policies when asset prices follow a diffusion process. J. Econom. Theory 49, 33–83.

Cvitanić, J., Karatzas, I., 1992. Convex duality in constrained portfolio optimization. Ann. Appl. Probab. 2, 767–818.

Cvitanić, J., Karatzas, I., Soner, M., 2002. Backward stochastic differential equations with constraints on the gains process. Ann. Probab., forthcoming.

Duffie, D., 2001. Dynamic Asset Pricing Theory, 3rd Edition. Princeton University Press, Princeton, NJ.

Duffie, D., Epstein, L., 1992a. Asset pricing with stochastic differential utility. Rev. Financial Stud. 5, 411–436.

Duffie, D., Epstein, L., 1992b. Stochastic differential utility. Econometrica 60, 353–394.

Duffie, D., Skiadas, C., 1994. Continuous-time security pricing: a utility gradient approach. J. Math. Econom. 23, 107–131.

Duffie, D., Schroder, M., Skiadas, C., 1997. A term structure model with preferences for the timing of resolution of uncertainty. Econom. Theory 9, 3–23.

Duffie, D., Filipović, D., Schachermayer, W., 2003. Affine processes and applications in finance. Ann. Appl. Probab. 13, 984–1053.

El Karoui, N., Peng, S., Quenez, M.-C., 2001. A dynamic maximum principle for the optimization of recursive utilities under constraints. Ann. Appl. Probab. 11, 664–693.

Ellsberg, D., 1961. Risk, ambiguity, and the savage axioms. Quart. J. Econom. 75, 643–669.

Epstein, L., Miao, J., 2000. A two-person dynamic equilibrium under ambiguity. Working Paper, Department of Economics, University of Rochester, Rochester, NY.

Epstein, L., Zin, S., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica 57, 937–969.
Fisher, M., Gilles, C., 1998. Consumption and asset prices with recursive preferences. Working Paper, Federal Reserve Board, Atlanta, GA.
Giovannini, A., Weil, P., 1989. Risk aversion and intertemporal substitution in the capital asset pricing model. NBER Working Paper No. 2824, Cambridge, MA.
Hansen, L., Sargent, T., Turmuhambetova, G., Williams, N., 2001. Robustness and uncertainty aversion. Working Paper, Department of Economics, University of Chicago.
He, H., Pearson, N., 1991. Consumption and portfolio policies with incomplete markets and short-sale constrains: the infinite dimensional case. J. Econom. Theory 54, 259–304.
Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financial Stud. 6, 327–344.
Karatzas, I., Shreve, S., 1998. Brownian Motion and Stochastic Calculus. Springer, New York.
Karatzas, I., Lehoczky, J., Shreve, S., 1987. Optimal portfolio and consumption decisions for a 'small investor' on a finite horizon. SIAM J. Control Optim. 25, 1557–1586.
Karatzas, I., Lehoczky, J., Shreve, S., Xu, G., 1991. Martingale and duality methods for utility maximization in an incomplete market. SIAM J. Control Optim. 29, 702–730.
Kim, T., Omberg, E., 1996. Dynamic nonmyopic portfolio behavior. Rev. Financial Stud. 9, 141–162.
Klein, E., Thompson, A., 1984. Theory of Correspondences. Wiley, New York.
Klibanoff, P., Marinacci, M., Mukerji, S., 2002. A smooth model of decision making under ambiguity. Working Paper, MEDS, Kellogg School of Management, Northwestern University.
Kobylanski, M., 2000. Backward stochastic differential equations and partial differential equations with quadratic growth. Ann. Probab. 28, 558–602.
Lazrak, A., Quenez, M.C., 2003. A generalized stochastic differential utility. Math. Oper. Res. 28, 154–180.
Lepeltier, J.-P., San Martín, J., 1997. Backward stochastic differential equations with continuous coefficient. Statist. Probab. Lett. 32, 425–430.
Lepeltier, J.-P., San Martín, J., 2002. On the existence or non-existence of solutions for certain backward stochastic differential equations. Bernoulli 8, 123–137.
Liu, J., 2001. Portfolio selection in stochastic environments. Working Paper, UCLA.
Ma, J., Protter, P., Yong, J., 1994. Solving forward–backward stochastic differential equations explicitly—a four step scheme. Probab. Theory Related Fields 98, 339–359.
Maenhout, P., 1999. Robust portfolio rules and asset pricing. Working Paper, INSEAD.
Merton, R., 1971. Optimum consumption and portfolio rules in a continuous-time model. J. Econom. Theory 3, 373–413.
Obstfeld, M., 1994. Risk-taking, global diversification, and growth. Amer. Econom. Rev. 84, 1310–1329.
Pardoux, E., Peng, S., 1990. Adapted solution of a backward stochastic differential equation. Systems Control Lett. 14, 55–61.
Piazzesi, M., 2002. Affine term structure models. Working Paper, GSB, University of Chicago.
Schroder, M., Skiadas, C., 1999. Optimal consumption and portfolio selection with stochastic differential utility. J. Econom. Theory 89, 68–126.
Schroder, M., Skiadas, C., 2002. An isomorphism between asset pricing models with and without linear habit formation. Rev. Financial Stud. 15, 1189–1221.
Schroder, M., Skiadas, C., 2003a. Lifetime consumption-portfolio choice under trading constraints and nontradeable income. Working Paper 324, Department of Finance, Kellogg School of Management, Northwestern University.
Schroder, M., Skiadas, C., 2003b. Quasiarbitrage, optimality, state pricing and duality in constrained financial markets with continuous and discontinuous information. Working Paper 327, Department of Finance, Kellogg School of Management, Northwestern University.
Segal, U., Spivak, A., 1990. First-order versus second-order risk aversion. J. Econom. Theory 51, 111–125.
Shreve, S., Xu, G., 1992. A duality method for optimal consumption and investment under short-selling prohibition: I, general market coefficients; and II, constant market coefficients. Ann. Appl. Probab. 2, 87–112. and 314–328.
Skiadas, C., 1992. Advances in the Theory of Choice and Asset Pricing. Ph.D. Thesis, Stanford University.
Skiadas, C., 1998. Recursive utility and preferences for information. Econom. Theory 12, 293–312.

Skiadas, C., 2003. Robust control and recursive utility. Finance Stochastics 7, 475–489.
Svensson, L., 1989. Portfolio choice with non-expected utility in continuous time. Econom. Lett. 30, 313–317.
Uppal, R., Wang, T., 2002. Model misspecification and under-diversification. Working Paper, School of Commerce, University of British Columbia, Canada.
Wachter, J., 2002. Portfolio and consumption decisions under mean-reverting returns: an exact solution in complete markets. J. Financial Quantitative Anal. 37, 63–91.
Xu, G., 1990. A duality method for optimal consumption and investment under short-selling prohibition. Ph.D. Thesis, Department of Mathematics, Carnegie-Mellon University.
Yong, J., Zhou, X.Y., 1999. Stochastic Controls: Hamiltonian Systems and HJB Equations. Springer, New York.
Zariphopoulou, T., Tiu, C.-I., 2002. Optimization in incomplete markets with recursive utility. Working Paper, University of Texas, Austin.