The Annals of Applied Probability
2015, Vol. 25, No. 4, 2066–2095
DOI: 10.1214/14-AAP1043

HEDGING, ARBITRAGE AND OPTIMALITY WITH SUPERLINEAR FRICTIONS

BY PAOLO GUASONI[1] AND MIKLÓS RÁSONYI

Boston University and Dublin City University, and MTA Alfréd Rényi Institute of Mathematics, Budapest and University of Edinburgh

In a continuous-time model with multiple assets described by càdlàg processes, this paper characterizes superhedging prices, absence of arbitrage, and utility maximizing strategies, under general frictions that make execution prices arbitrarily unfavorable for high trading intensity. Such frictions induce a duality between feasible trading strategies and shadow execution prices with a martingale measure. Utility maximizing strategies exist even if arbitrage is present, because it is not scalable at will.

1. Introduction. In financial markets, trading moves prices against the trader: buying faster increases execution prices, and selling faster decreases them. This aspect of liquidity, known as market depth [Black (1986)] or price-impact, is widely documented empirically [Cont, Kukanov and Stoikov (2014), Dufour and Engle (2000)], and has received increasing attention in models of asymmetric information [Kyle (1985)], illiquid portfolio choice [Garleanu and Pedersen (2013), Rogers and Singh (2010)] and optimal liquidation [Almgren and Chriss (2001), Bertsimas and Lo (1998), Schied and Schöneborn (2009)]. These models depart from the literature on frictionless markets, where prices are the same for any amount traded. They also depart from proportional transaction costs models, in which prices differ for buying and selling, but are insensitive to quantities.[2]

The growing interest in price-impact has also highlighted a shortage of effective theoretical tools. In these models, what is the analogue of a martingale measure? Which contingent claims are hedgeable, and at what price? How do the familiar optimality conditions for utility maximization look in this context? In discrete time, several researchers have studied these fundamental questions [Astic and Touzi (2007), Dolinsky and Soner (2013), Pennanen (2011a), Pennanen and

Received August 2013; revised March 2014.

[1]Supported in part by the ERC (278295), NSF (DMS-11-09047), SFI (07/SK/M1189, 08/SRC/FMC1389), and FP7 (RG-248896).

MSC2010 subject classifications. 91G10, 91G80.

Key words and phrases. Hedging, arbitrage, price-impact, frictions, utility maximization.

[2]A separate class of models [e.g., Bank and Kramkov (2011a, 2011b)] investigates the conditions under which current prices depend on past trading activity, a distinct effect also referred to as (permanent) price-impact. This paper focuses on temporary price-impact, or market depth, which some authors call nonlinear transaction costs [cf. Garleanu and Pedersen (2013)].

Penner (2010)], but extensions to continuous time have proved challenging. This paper aims at filling the gap.

Tackling price-impact in continuous-time requires to clarify two basic concepts that remain concealed in discrete models: the relevant classes of trading strategies and of dual variables. First, to retain price-impact effects in continuous time, execution prices must depend on the traded quantities per unit of time, that is, on trading intensity, rather than on the traded quantities themselves, otherwise price-impact can be avoided with judicious policies [Çetin, Jarrow and Protter (2004), Çetin and Rogers (2007), Çetin, Soner and Touzi (2010)]. Various classes of trading strategies have appeared in different models [Çetin, Soner and Touzi (2010), Schied and Schöneborn (2009)], but a generally agreed definition of what kind of strategies should be allowed has not yet emerged. The second key concept is the relevant notion of dual variables—the analogue of a martingale measure. The proportional transaction costs literature identifies the corresponding dual variable as a consistent prices system, a pair $(\tilde{S}, Q)$ of a price $\tilde{S}$ evolving within the bid-ask spread, and a probability Q under which $\tilde{S}$ is a martingale.[3] Such a definition suggests that with frictions, passing to the risk-neutral setting requires both a change in the probability measure and a change in the price process.

Superlinear frictions in the sense of the present paper, such as price-impact models, entail that execution prices become arbitrarily unfavorable as traded quantities per unit of time grow: buying or selling too fast becomes impossible. As a result, trading is feasible only at finite rates—the number of shares is absolutely continuous. This feature sets apart superlinear frictions from frictionless markets, in which the number of shares is merely predictable, and from models with proportional transaction costs, in which they have finite variation.

Finite trading rates have two central implications: first, portfolio values are well defined for asset prices that follow general càdlàg processes, not only for semimartingales. Second, immediate portfolio liquidation is impossible and, therefore, the usual notion of admissibility, based on a lower bound for liquidation values, is inappropriate. We define below a *feasible* strategy as any trading policy with finite trading rate and trading volume, without any lower bounds on portfolio values. In particular, this definition does not involve the asset price. In frictionless markets, or under proportional transaction costs, this approach would fail for two reasons: first, the set of claims attainable by feasible strategies would not be closed in any reasonable sense, as a block trade is approximated by intense trading over small time intervals. Second, portfolios unbounded from below allow doubling strategies, which lead to arbitrage even with martingale prices.

Neither issue arises in our models with superlinear frictions. Block trades are infeasible, even in the limit, as intense trading incurs exorbitant costs: put differently,

[3] These dual objects first appeared in Jouini and Kallal (1995). They were baptized "consistent price systems" in Schachermayer (2004). See Kabanov and Safarian (2009) for further developments.

bounded losses imply bounded *trading volume* (Lemma 3.4). The bound on trading volume in turn yields the closedness of the payoffs of feasible strategies (Proposition 3.5), and the martingale property of portfolio values under shadow execution prices, which excludes arbitrage through doubling strategies (Lemma 5.6).

Arbitrage also occurs differently in the present setting. Unlike models without friction or with proportional transaction costs, where an arbitrage opportunity scales freely, superlinear frictions imply that scaling trading rates results in a less than proportional scaling of payoffs [see Pennanen (2011b) for more about scalable arbitrage]. In fact, in our setting (Assumption 2.3) we prove a stronger result, whereby *all* payoffs are dominated by a single random variable, the *market bound*, which depends on the friction and on the asset price only (Lemma 3.5). This bound implies that price-impact defeats arbitrage, if pursued on a large scale.

All these definitions and properties come together in the main superhedging result, Theorem 3.7, which characterizes the initial asset positions that can dominate a given claim through trading, in terms of shadow execution prices. The main message of this theorem is that the superhedging price of a claim is the supremum of its expected value under a martingale measure for a shadow execution price, *minus* a penalty, which reflects how far the shadow price is from the base price. The penalty depends on the *dual friction*, introduced by Dolinsky and Soner (2013) in discrete time, and is zero for any equivalent martingale measure of the asset price. Importantly, the theorem is valid even if there are no martingale measures, or if the price is not a semi-martingale.

The superhedging theorem, which does not assume absence of arbitrage, characterizes a large class of models that do not admit arbitrage of the second kind (strategies that lead to a sure minimal gain) even in limited amounts. As for proportional transaction costs, this class contains any price process that satisfies the conditional full support property Guasoni, Rásonyi and Schachermayer (2008), including fractional Brownian motion.

We conclude the paper by addressing utility maximization. First, a general theorem guarantees that optimal solutions exist. This holds true even in the eventual presence of arbitrage opportunities, which must be chosen optimally, lest price-impact offset gains. Second, optimal strategies are identified by a version of the familiar first-order condition that the marginal utility of the optimal payoff be proportional to a stochastic discount factor. Technicalities aside, price-impact leads to a novel condition, which prescribes that a stochastic discount factor makes the shadow execution price, not the base price, a martingale. In models with proportional transaction costs this criterion formally reduces to the usual shadow price approach for optimality [Kallsen and Muhle-Karbe (2010)].

The rest of the paper proceeds with Section 2, which describes the model in detail. The main theoretical tools are developed in Section 3, which proves the market bound, the trading volume bound, the closedness of the payoff space, and the main superhedging result. Section 4 discusses the implications for arbitrage of the second kind, and its absence with prices with conditional full support. Section 5 concludes with the results on utility maximization.

2. The model. For a finite time horizon $T > 0$, consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ with $\mathcal{F}_0$ trivial, satisfying the usual hypotheses as well as $\mathcal{F} = \mathcal{F}_T$. $\mathcal{O}$ denotes the optional sigma-field on $\Omega \times [0, T]$. The market includes a riskless and perfectly liquid asset S^0, used as numeraire, hence $S^0_t \equiv 1$, $t \in [0, T]$, and d risky assets, described by càdlàg, adapted processes $(S^i_t)^{1\le i\le d}_{t\in[0,T]}$. Henceforth, S denotes the d-dimensional process with components S^i, $1 \le i \le d$, the concatenation xy of two vectors x, y of equal dimensions denotes their scalar product, and $|x|$ denotes the Euclidean norm of x. The components of a $(d + 1)$-dimensional vector x are denoted by $x^0, \dots, x^d$.

The next definition identifies those strategies for which the number of shares changes over time at some finite rate ϕ, hence it is absolutely continuous.

DEFINITION 2.1. A *feasible strategy* is a process ϕ in the class

$$\mathcal{A} := \left\{\phi : \phi \text{ is a } \mathbb{R}^d\text{-valued, optional process, } \int_0^T |\phi_u|\,du < \infty \text{ a.s.}\right\}. \tag{1}$$

In this definition, the process ϕ represents the *trading rate*, that is, the speed at which the number of shares in each asset changes over time, and the condition $\int_0^T |\phi_u|\,du < \infty$ means that *absolute turnover* (the cumulative number of shares bought or sold) remains finite in finite time.

The above definition compares to the one of admissible strategies in frictionless markets as follows. On one hand, it relaxes the solvency constraint typical of admissibility, since a feasible strategy can lead to negative wealth. On the other hand, this definition restricts the number of shares to be differentiable in time, while usual admissible strategies have an arbitrarily irregular number of shares.[4] Note also that the definition of feasibility does not involve the asset price at all.

With this notation, in the absence of frictions the self-financing condition would imply a position at time T in the safe asset (henceforth, cash) equal to:[5]

$$z^0 - \int_0^T S_t \phi_t\,dt, \tag{2}$$

where z^0 represents the initial capital, and the integral reflects the cost of purchases and the proceeds of sales. For a given trading strategy ϕ, frictions reduce the cash position, by making purchases more expensive, and sales less profitable. With a similar notation to Dolinsky and Soner (2013), we model this effect by introducing a function G, which summarizes the impact of frictions on the execution price at different trading rates:

---

[4]In the definition of feasible strategy an optional trading rate leads to a continuous, hence predictable, number of shares, as for usual admissible strategies.

[5]By the càdlàg property of S_t, the function $S_t(\omega)$, $t \in [0, T]$ is bounded for almost every $\omega \in \Omega$, hence the integral in (2) is finite a.s. for each ϕ satisfying $\int_0^T |\phi_t|\,dt < \infty$ a.s.

ASSUMPTION 2.2 (Friction). Let $G:\Omega\times[0,T]\times\mathbb{R}^d\to\mathbb{R}_+$ be a $\mathcal{O}\otimes\mathcal{B}(\mathbb{R}^d)$-measurable function, such that $G(\omega,t,\cdot)$ is convex with $G(\omega,t,x)\geq G(\omega,t,0)$ for all ω,t,x. Henceforth, set $G_t(x):=G(\omega,t,x)$, that is, the dependence on ω is omitted, and t is used as a subscript.

With this definition, for a given strategy $\phi\in\mathcal{A}$ and an initial asset position $z\in\mathbb{R}^{d+1}$, the resulting positions at time $t\in[0,T]$ in the risky and safe assets are defined as

$$V_t^i(z,\phi):=z^i+\int_0^t\phi_u^i\,du,\qquad 1\leq i\leq d, \tag{3}$$

$$V_t^0(z,\phi):=z^0-\int_0^t\phi_u S_u\,du-\int_0^t G_u(\phi_u)\,du. \tag{4}$$

The first equation merely says that the cumulative number of shares V_t^i in the ith asset is given by the initial number of shares, plus subsequent flows. The second equation contains the new term involving the friction G, which summarizes the impact of trading on execution prices. The condition $G(\omega,t,x)\geq G(\omega,t,0)$ means that inactivity is always cheaper than any trading activity. Most models in the literature assume $G(\omega,t,0)=0$, but the above definition allows for $G(\omega,t,0)>0$, which is interpreted as a cost of participation in the market, such as the fees charged by exchanges to trading firms, or as a monitoring cost. The convexity of $x\mapsto G_t(x)$ implies that, excluding monitoring costs, trading twice as fast for half the time locally increases execution costs—speed is expensive.[6] Finally, note that in general V_t^0 may take the value $-\infty$ for some (unwise) strategies.

With a single risky asset and for $G(\omega,t,0)=0$, the above specification is equivalent to assuming that a trading rate of $\phi_t\neq 0$ implies an execution price equal to

$$\tilde{S}_t=S_t+G_t(\phi_t)/\phi_t, \tag{5}$$

which is (since G is positive) higher than S_t when buying, and lower when selling. Thus, $G\equiv 0$ boils down to a frictionless market, while proportional transaction costs correspond to $G_t(x)=\varepsilon S_t|x|$ with some $\varepsilon>0$. Yet this paper focuses on neither of these settings, which entail either zero or linear costs, but rather on superlinear frictions, defined as those that satisfy the following conditions. Note that we require a strong form of superlinearity here (i.e., the cost functional grows at least as a superlinear power of the traded volume).

[6] Let $g(x)=G(\omega,t,x)$, that is, focus on a local effect. Then, by convexity, $g(x)\leq(1-1/k)g(0)+(1/k)g(kx)$ for $k>1$ and, therefore, $(g(kx)-g(0))T/k\geq(g(x)-g(0))T$, which means that increasing trading speed by a factor of k and reducing trading time by the same factor implies higher trading costs, excluding the monitoring cost captured by $g(0)$.

ASSUMPTION 2.3 (Superlinearity). There is $\alpha > 1$ and an optional process H such that[7]

$$\inf_{t\in[0,T]} H_t > 0 \qquad \text{a.s.,} \tag{6}$$

$$G_t(x) \geq H_t|x|^\alpha \qquad \text{for all } \omega, t, x, \tag{7}$$

$$\int_0^T \Big(\sup_{|x|\leq N} G_t(x)\Big)\, dt < \infty \qquad \text{a.s. for all } N > 0, \tag{8}$$

$$\sup_{t\in[0,T]} G_t(0) \leq K \qquad \text{a.s. for some constant } K. \tag{9}$$

Condition (7) is the central superlinearity assumption, and prescribes that trading twice as fast for half the time increases trading costs (in excess of monitoring) by a minimum positive proportion. Condition (6) requires that frictions never disappear, and (8) that they remain finite in finite time. By (9), the participation cost must be uniformly bounded in $\omega \in \Omega$. In summary, these conditions characterize nontrivial, finite, superlinear frictions. Note that (7) implies that $\tilde{S}_t$ in (5) becomes arbitrarily negative as ϕ_t becomes negative enough, that is, when selling too fast. This issue is addressed in more detail in Remarks 3.8 and 5.3 below.

The most common examples in the literature are, with one risky asset, the friction $G_t(x) := \Lambda|x|^\alpha$ for some $\Lambda > 0$, $\alpha > 1$ [see, e.g., Dolinsky and Soner (2013)] and, in multiasset models, the friction $G_t(x) := x'\Lambda x$ for some symmetric, positive-definite, $d \times d$ square matrix Λ (here x' stands for the transposition of the vector x); see Garleanu and Pedersen (2013).

REMARK 2.4. We conjecture that (7) could be weakened to the superlinearity condition

$$\lim_{x\to\infty} G_t(x)/|x| = \infty \qquad \text{a.s.,}$$

using Orlicz spaces instead of L^p-estimates (i.e., Hölder's inequality). This generalization is expected to involve substantial further technicalities for a limited increase in generality, hence it is not pursued here.

REMARK 2.5. Our results remain valid assuming that (7) holds for $|x| \geq M$ only, with some $M > 0$. Such an extension requires only minor modifications of the proofs, and may accommodate models for which a low trading rate incurs either zero or linear costs.

[7] We implicitly assume that $\inf_{t\in[0,T]} H_t$ is a random variable, which is always the case if, for example, H is càdlàg.

3. Superhedging and dual characterization of payoffs. Despite their similarity to models of frictionless markets and proportional transaction costs, superlinear frictions in the sense of Assumption 2.3 lead to a surprisingly different structure of attainable payoffs, as shown in this section. Indeed, the class of feasible strategies considered above, while still well defined even in a model without frictions or with proportional transaction costs, is virtually useless in such settings, as the set of terminal payoffs corresponding to feasible strategies is not closed in any reasonable sense.

As an example, a simple trading policy that buys one share of the risky asset at time t and sells it at time T is not a feasible strategy in the above sense, because it is not absolutely continuous, and in fact is discontinuous at t and T. Yet, in frictionless markets or with transaction costs, this policy is approximated arbitrarily well by another one that buys at rate n in the interval $[t, t + 1/n]$ and sells at rate n on $[T, T + 1/n]$. That is, the sequence of corresponding payoffs converges to a finite payoff, but this limit payoff does not belong to the payoff space of feasible strategies.

By contrast, with the superlinear frictions in Assumption 2.3, the set of terminal values corresponding to feasible strategies *is* closed in a strong sense. The intuitive reason is that approximating a nonsmooth strategy would require trading at increasingly high speed, generating infinite costs, and preventing convergence to a finite payoff.

3.1. *The market bound.* Superlinear frictions in the sense of Assumption 2.3 lead to a striking boundedness property: for a fixed initial position, all payoffs of feasible strategies are bounded above by a single random variable $B < \infty$, the *market bound*, which depends on the friction G and on the price S, but *not* on the strategy. This property clearly fails in frictionless markets, where any payoff with zero initial capital can be scaled arbitrarily and, therefore, admits no uniform bound. In such markets, a much weaker boundedness property holds: Corollary 9.3 of Delbaen and Schachermayer (2006) shows that the set of payoff of x-admissible strategies is bounded in L^0 *if* the market is arbitrage-free in the sense of the condition (NFLVR), and a similar result holds with proportional transaction costs under the (RNFLVR) property [Guasoni, Lépinette and Rásonyi (2012)].

A central tool in this analysis is the function G^*, the Fenchel–Legendre conjugate of G, which we call *dual friction*. Its importance was first recognized by Dolinsky and Soner (2013), who used it to derive a superhedging result in discrete time. G^* is defined as[8]

$$G_t^*(y) := \sup_{x\in\mathbb{R}^d} \big(xy - G_t(x)\big), \qquad y \in \mathbb{R}^d, t \in [0, T], \tag{10}$$

[8]Note that the supremum can be taken over $\mathbb{Q}^d$, hence G^* is $\mathcal{O} \otimes \mathcal{B}(\mathbb{R}^d)$-measurable. Note also that, under Assumption 2.3, $G_t^*(\cdot)$ is a finite, convex function satisfying $G_t^*(x) \geq -K$ for all x, see the proof of Lemma 3.2.

and the typical case $d = 1$, $G_t(x) = \Lambda|x|^\alpha$ leads to $G_t^*(y) = \frac{\alpha-1}{\alpha}\alpha^{1/(1-\alpha)} \times \Lambda^{1/(1-\alpha)}|y|^{\alpha/(\alpha-1)}$ [in particular, $G_t^*(y) = y^2/(4\Lambda)$ for $\alpha = 2$]. The key observation is the following.

LEMMA 3.1. *Under Assumption 2.3, any $\phi \in \mathcal{A}$ satisfies*

$$V_T^0(z, \phi) \le z^0 + \int_0^T G_t^*(-S_t)\,dt < \infty \qquad a.s.$$

PROOF. Indeed, this follows from (4), the definition of G_t^*, and Lemma 3.2 below. □

LEMMA 3.2. *Under Assumption 2.3, the random variable $B := \int_0^T G_t^*(-S_t)\,dt$ is finite almost surely.*

PROOF. Consider first the case $d = 1$. Then, by direct calculation,

$$G_t^*(y) \le \sup_{r\in\mathbb{R}}\bigl(ry - H_t|r|^\alpha\bigr) = \frac{\alpha - 1}{\alpha}\alpha^{1/(1-\alpha)} H_t^{1/(1-\alpha)}|y|^{\alpha/(\alpha-1)}. \tag{11}$$

Noting that $\sup_{t\in[0,T]}|S_t|$ is finite a.s. by the càdlàg property of S, and knowing that $\inf_{t\in[0,T]} H_t$ is a positive random variable, it follows that

$$\sup_{t\in[0,T]} G_t^*(-S_t) < \infty \qquad \text{a.s.},$$

which clearly implies the statement. If $d > 1$, then note that

$$\begin{aligned} G_t^*(y) &\le \sup_{r\in\mathbb{R}^d}\left(\sum_{i=1}^d r^i y^i - H_t|r|^\alpha\right) \le \sum_{i=1}^d \sup_{r\in\mathbb{R}^d}\bigl(r^i y^i - (H_t/d)|r|^\alpha\bigr) \\ &\le \sum_{i=1}^d \sup_{x\in\mathbb{R}}\bigl(xy^i - (H_t/d)|x|^\alpha\bigr) \end{aligned} \tag{12}$$

and the conclusion follows from the scalar case. □

Since $B < \infty$ a.s., it is impossible to achieve a scalable arbitrage: though a trading strategy may realize an a.s. positive terminal value, one cannot get an arbitrarily large profit by scaling the trading strategy (i.e., by multiplying it with large positive constants) since bigger trading values also enlarge costs. Even if an arbitrage exists, amplifying it too much backfires, because the superlinear friction eventually overrides profits. Yet, arbitrage opportunities can exist in limited size (cf. Section 4 below).

Limited arbitrage opportunities also appear in the frictionless models of Fernholz, Karatzas and Kardaras (2005) and Karatzas and Kardaras (2007) through a completely different mechanism. These models allow for arbitrage opportunities that can lead to a possible intermediate loss before realizing a certain final gain, while requiring that wealth remains positive at all times. As a result, an arbitrage opportunity is scalable only insofar as its maximal intermediate loss is less than the initial capital committed to the arbitrage. By contrast, with superlinear frictions arbitrage is limited even though wealth may well become negative before gains are realized (cf. Definition 2.1), because the superlinear friction defeats attempts to scale an arbitrage linearly, by reducing and eventually eliminating its profitability for larger positions.

3.2. *Trading volume bound.* For $Q \sim P$, denote by $L^1(Q)$ the usual Banach space of $(d+1)$-dimensional, Q-integrable random variables; given a subset A of a Euclidean space, $L^0(A)$ denotes the set of (P-a.s. equivalence classes of) A-valued random variables, equipped with the topology of convergence in probability. $E_Q X$ denotes the expectation of a random variable X under Q. From now on, fix $1 < \beta < \alpha$, where α is as in Assumption 2.3. Let γ be the conjugate number of β, defined by

$$\frac{1}{\beta} + \frac{1}{\gamma} = 1.$$

The next definition identifies a class of reference probability measures with integrability properties that fit the friction G and the price process S well. Our main results (see Section 3.4) involve suprema of expectations of various functionals under families of probability measures equivalent to P. Ideally, *all* such measures should be taken (as in Theorem 3.11 below) but on infinite Ω this leads to integrability issues. Thus, we need to single out a family of probability measures which is large enough for the results of Section 3.4 to hold, but also small enough to ensure appropriate integrability properties. This is why we introduce the sets $\mathcal{P}$ and $\mathcal{P}(W)$ in Definition 3.3 below. $\mathcal{P}$ identifies a set of probabilities under which some shadow execution price has the martingale property, as explained in the proof of Theorem 5.5 and Lemma 5.6 below.

DEFINITION 3.3. $\mathcal{P}$ denotes the set of probabilities $Q \sim P$ such that

$$E_Q \int_0^T H_t^{\beta/(\beta-\alpha)} (1 + |S_t|)^{\beta\alpha/(\alpha-\beta)} \, dt < \infty.$$

$\tilde{\mathcal{P}}$ denotes the set of probability measures $Q \in \mathcal{P}$ such that

$$E_Q \int_0^T |S_t| \, dt < \infty \quad \text{and} \quad E_Q \int_0^T \sup_{|x| \le N} G_t(x) \, dt < \infty \qquad \text{for all } N \ge 1.$$

For a (possibly multivariate) random variable W, define

$$\mathcal{P}(W) := \{Q \in \mathcal{P} : E_Q|W| < \infty\}, \qquad \tilde{\mathcal{P}}(W) := \{Q \in \tilde{\mathcal{P}} : E_Q|W| < \infty\}.$$

Under Assumption 2.3, note that $\tilde{\mathcal{P}}(W) \neq \varnothing$ for all W by Dellacherie and Meyer [(1982), page 266]. The next lemma shows that, if a payoff has a finite negative part under some probability in $\mathcal{P}$, then its trading rate must also be (suitably) integrable. There is no analogue to such a result in frictionless markets, but transaction costs Guasoni, Lépinette and Rásonyi (2012), Lemma 5.5, lead to a similar property, whereby any admissible strategy must satisfy an upper bound on its total variation. In both cases, the intuition is that, with frictions, excessive trading causes unbounded losses. Hence, a bound on losses translates into one for trading volume. Lemma 3.4 is crucial to establish the closedness of the set of attainable payoffs (Proposition 3.5 below) as well as to prove the martingale property of shadow execution prices in utility maximization problems (see Lemma 5.6 in Section 5).

In the sequel, x_- denotes the negative part of $x \in \mathbb{R}$.

LEMMA 3.4. *Let $Q \in \mathcal{P}$ and $\phi \in \mathcal{A}$ be such that $E_Q \xi_- < \infty$, where*

$$\xi := -\int_0^T S_t \phi_t \, dt - \int_0^T G_t(\phi_t) \, dt.$$

Then

$$E_Q \int_0^T |\phi_t|^\beta (1 + |S_t|)^\beta \, dt < \infty.$$

PROOF. For ease of notation, set $T := 1$. Define $\phi_t(n) := \phi_t 1_{\{|\phi_t| \le n\}} \in \mathcal{A}$, $n \in \mathbb{N}$. As $n \to \infty$, clearly $\phi_t(n) \to \phi_t$ for all t and $\omega \in \Omega$, and the random variables

$$\xi_n := -\int_0^1 S_t \phi_t(n) \, dt - \int_0^1 G_t(\phi_t(n)) \, dt \tag{13}$$

$$= -\sum_{i=1}^d \int_0^1 S_t^i \phi_t^i(n) [1_{\{S_t^i \le 0, \phi_t^i \le 0\}} + 1_{\{S_t^i > 0, \phi_t^i \le 0\}} + 1_{\{S_t^i \le 0, \phi_t^i > 0\}} + 1_{\{S_t^i > 0, \phi_t^i > 0\}}] \, dt \tag{14}$$

$$- \int_0^1 G_t(\phi_t(n)) \, dt \tag{15}$$

converge to ξ a.s. by monotone convergence. [Note that each of the terms with an indicator converges monotonically, and that $G_t(0) \le G_t(x)$ for all x.] Hölder's inequality yields

$$\int_0^1 |\phi_t(n)|^\beta (1 + |S_t|)^\beta \, dt$$

$$= \int_0^1 |\phi_t(n)|^\beta H_t^{\beta/\alpha} \frac{1}{H_t^{\beta/\alpha}} (1 + |S_t|)^\beta \, dt \tag{16}$$

$$\leq \left[\int_0^1 |\phi_t(n)|^\alpha H_t\,dt\right]^{\beta/\alpha}\left[\int_0^1 \left(\frac{1}{H_t^{\beta/\alpha}}(1+|S_t|)^\beta\right)^{\alpha/(\alpha-\beta)} dt\right]^{(\alpha-\beta)/\alpha}$$

$$\leq \left[\int_0^1 G_t(\phi_t(n))\,dt\right]^{\beta/\alpha}\left[\int_0^1 \left(\frac{1}{H_t^{\beta/\alpha}}(1+|S_t|)^\beta\right)^{\alpha/(\alpha-\beta)} dt\right]^{(\alpha-\beta)/\alpha}.$$

All these integrals are finite by Assumption 2.3 and the càdlàg property of S. Now, set

$$m := \left[\int_0^1 \left(\frac{1}{H_t^{\beta/\alpha}}(1+|S_t|)^\beta\right)^{\alpha/(\alpha-\beta)} dt\right]^{(\alpha-\beta)/\alpha},$$

and note that, by Jensen's inequality,

$$\begin{aligned}\left|\int_0^1 S_t\phi_t(n)\,dt\right| &\leq \int_0^1 |\phi_t(n)|(1+|S_t|)\,dt \\ &\leq \left[\int_0^1 |\phi_t(n)|^\beta(1+|S_t|)^\beta\,dt\right]^{1/\beta}.\end{aligned}\tag{17}$$

Note also that if $x \geq 1$ and $x \geq 2^{\beta/(\alpha-\beta)}m^{\alpha/(\alpha-\beta)}$ then $x^{1/\beta} - (x/m)^{\alpha/\beta} \leq x - 2x = -x$. This observation, applied to

$$x := \int_0^1 |\phi_t(n)|^\beta(1+|S_t|)^\beta\,dt,$$

implies that $\xi_n \leq -x$ on the event $\{x \geq 2^{\beta/(\alpha-\beta)}m^{\alpha/(\alpha-\beta)} + 1\}$. Thus,

$$\int_0^1 |\phi_t(n)|^\beta(1+|S_t|)^\beta\,dt \leq (\xi_n)_- + 2^{\beta/(\alpha-\beta)}m^{\alpha/(\alpha-\beta)} + 1 \qquad \text{a.s.}$$

Letting n tend to ∞, it follows that

$$\int_0^1 |\phi_t|^\beta(1+|S_t|)^\beta\,dt \leq \xi_- + 2^{\beta/(\alpha-\beta)}m^{\alpha/(\alpha-\beta)} + 1,\tag{18}$$

which implies the claim, since $E_Q\xi_- < \infty$ by assumption, and $E_Q m^{\alpha/(\alpha-\beta)} < \infty$ from $Q \in \mathcal{P}$. □

3.3. *Closed payoff space.* The central implication of the previous result is that the class of multivariate payoffs superhedged by a feasible strategy, defined as $C := [\{V_T(0,\phi) : \phi \in \mathcal{A}\} - L^0(\mathbb{R}_+^{d+1})] \cap L^0(\mathbb{R}^{d+1})$, is closed in a rather strong sense; recall the componentwise definition of the $(d+1)$-dimensional random variable $V_T(0,\phi)$ in (3) and (4). Closedness is the key property for establishing superhedging results; see, for example, Section 9.5 of Delbaen and Schachermayer (2006) or Section 3.6 of Kabanov and Safarian (2009).

PROPOSITION 3.5. *Under Assumption 2.3, the set $C \cap L^1(Q)$ is closed in $L^1(Q)$ for all $Q \in \mathcal{P}$ such that $\int_0^T |S_t|\,dt$ is Q-integrable.*

PROOF. Take $T = 1$ for simplicity, and assume that $\rho_n := \xi_n - \eta_n \to \rho$ in $L^1(Q)$ where $\eta_n \in L^0(\mathbb{R}_+^{d+1})$ and $\xi_n = V_1(0, \psi(n))$ for some $\psi(n) \in \mathcal{A}$ are such that $\rho_n \in L^1(Q)$. Up to a subsequence, this convergence takes place a.s. as well.

Lemma 3.4 implies that $E_Q \int_0^1 |\psi_t(n)|^\beta (1 + |S_t|)^\beta \, dt$ must be finite for all n since $(\xi_n)_- \le (\rho_n)_-$ and the latter is in $L^1(Q)$. Applying (18) with the choice $\phi := \psi(n)$ yields

$$\int_0^1 |\psi_t(n)|^\beta (1 + |S_t|)^\beta \, dt \le (\rho_n)_- + 2^{\beta/(\alpha-\beta)} m^{\alpha/(\alpha-\beta)} + 1.$$

Now, since $Q \in \mathcal{P}$, and the sequence ρ_n is bounded in $L^1(Q)$ because it is convergent in $L^1(Q)$, it follows that

$$\sup_{n \ge 1} E_Q \int_0^1 |\psi_t(n)|^\beta (1 + |S_t|)^\beta \, dt < \infty. \tag{19}$$

Consider $\mathbb{L} := L^1(\Omega, \mathcal{F}, Q; \mathbb{B})$, the Banach space of $\mathbb{B}$-valued Bochner-integrable functions, where $\mathbb{B} := L^\beta([0,1], \mathcal{B}([0,1]), \mathrm{Leb})$ is a separable and reflexive Banach space. The functions $\psi_\cdot(n) : \Omega \to \mathbb{B}$ are easily seen to be weakly measurable, hence also strongly measurable by the separability of $\mathbb{B}$. By (19), the sequence $\psi_\cdot(n)$ is bounded in $\mathbb{L}$, so Lemma 15.1.4 in Delbaen and Schachermayer (2006) yields convex combinations

$$\tilde{\psi}_\cdot(n) = \sum_{j=n}^{M(n)} \alpha_j(n) \psi_\cdot(n),$$

which converge to some $\tilde{\psi}_\cdot \in \mathbb{L}$ a.s. in $\mathbb{B}$-norm.

By the bound in (19), $\sup_n E_Q \int_0^1 |\phi_t(n)|(1 + |S_t|) \, dt < \infty$. Now apply Lemma 9.8.1 of Delbaen and Schachermayer (2006) to the sequence $\tilde{\psi}_\cdot(n)$ in the space of $(d+1)$-dimensional random variables $L^1(\Omega \times [0,1], \mathcal{O}, \nu)$, where ν is the measure defined by

$$\nu(A) := \int_{\Omega \times [0,1]} 1_A(\omega, t) \big(1 + |S_t|\big) \, dt \, dQ(\omega)$$

for $A \in \mathcal{O}$ (which is finite by the choice of Q). This lemma yields convex combinations $\hat{\psi}_\cdot(n)$ of the $\tilde{\psi}_\cdot(n)$ such that $\hat{\psi}_\cdot(n)$ converges to $\psi_\cdot$ ν-almost everywhere and hence $P \times$ Leb-almost everywhere. This shows, in particular, that ψ is $\mathcal{O}$-measurable.

Since $\tilde{\psi}_\cdot(n)$ converge a.s. in $\mathbb{B}$-norm, also $\hat{\psi}_\cdot(n) \to \tilde{\psi}$ a.s. in $\mathbb{B}$-norm, so $\psi = \tilde{\psi}$, $P \times$ Leb-a.e. and hence we may and will assume that $\tilde{\psi}_\cdot(n)$ tends to ψ a.s. in $\mathbb{B}$-norm as well as $P \times$ Leb-a.e.

Define $\tilde{\xi}_n := \sum_{j=n}^{M(n)} \alpha_j(n) \xi_j$ and $\tilde{\eta}_n := \sum_{j=n}^{M(n)} \alpha_j(n) \eta_j$. It holds that $\lim_{n\to\infty} \int_0^1 \tilde{\psi}_t(n) S_t \, dt = \int_0^1 \psi_t S_t \, dt$ almost surely, and also

$$\lim_{n\to\infty} \tilde{\xi}_n^i = \lim_{n\to\infty} \int_0^1 \tilde{\psi}_t^i(n) \, dt = \int_0^1 \psi_t^i \, dt \qquad \text{a.s. for } i = 1, \dots, d.$$

Hence, $\tilde{\eta}^i_n \to \eta^i$ a.s. with $\eta^i := \int_0^T \tilde{\psi}^i_t\,dt - \rho^i \in L^0(\mathbb{R}_+)$. By the convexity of G_t,

$$
\begin{aligned}
\rho^0 &= \lim_{n\to\infty}(\tilde{\xi}^0_n - \tilde{\eta}^0_n) \\
&\le \limsup_{n\to\infty}\Big[-\int_0^1 \tilde{\psi}_t(n)S_t\,dt - \int_0^1 G_t(\tilde{\psi}_t(n))\,dt - \tilde{\eta}^0_n\Big] \\
&= \limsup_{n\to\infty}\Big[-\int_0^1 \tilde{\psi}_t(n)S_t\,dt - \int_0^1 G_t(\psi_t)\,dt - \int_0^1 G_t(\tilde{\psi}_t(n))\,dt \\
&\qquad\qquad + \int_0^1 G_t(\psi_t)\,dt - \tilde{\eta}^0_n\Big] \\
&= -\int_0^1 \psi_t S_t\,dt - \int_0^1 G_t(\psi_t)\,dt \\
&\quad + \limsup_{n\to\infty}\Big[-\int_0^1 G_t(\tilde{\psi}_t(n))\,dt + \int_0^1 G_t(\psi_t)\,dt - \tilde{\eta}^0_n\Big].
\end{aligned}
$$

Now Fatou's lemma and $\tilde{\eta}_n \in L^0(\mathbb{R}^{d+1}_+)$ imply that the limit superior is in $-L^0(\mathbb{R}_+)$ [note that $G_t(\cdot)$ is continuous by convexity], hence there is $\eta^0 \in L^0(\mathbb{R}_+)$ such that

$$
\rho^0 = -\int_0^1 \psi_t S_t\,dt - \int_0^1 G_t(\psi_t)\,dt - \eta^0,
$$

which proves the proposition. □

COROLLARY 3.6. *Under Assumption 2.3, the set C is closed in probability.*

PROOF. Let $\rho_n \in C$ tend to ρ in probability. Up to a subsequence, convergence also holds almost surely. There exists $Q \in \mathcal{P}$ [see page 266 of Dellacherie and Meyer (1982)] such that $\rho, \sup_n|\rho - \rho_n|, \int_0^T |S_t|\,dt$ are all Q-integrable. Then $\rho_n \to \rho$ in $L^1(Q)$ as well, and Proposition 3.5 implies that $\rho \in C$. □

3.4. *Superhedging*. Finally, the main superhedging theorem. To the best of our knowledge, Theorem 3.7 is the first dual characterization in continuous time of hedgeable contingent claims with price-impact. Results in discrete time include Astic and Touzi (2007), Dolinsky and Soner (2013), Pennanen (2011a), Pennanen and Penner (2010). Our result is inspired, in particular, by Theorem 3.1 of Dolinsky and Soner (2013) for finite probability spaces.

Note that both terminal claims and initial endowments are multivariate, for a good reason. Due to the presence of price impact, positions in the safe asset and in various risky assets are not immediately convertible into each other at a fixed price. It is thus impossible to introduce, in a meaningful way, a one-dimensional wealth process representing holdings in units of a numéraire—multivariate book-keeping of positions is necessary.

TABLE 1
Summary of vector notation

$\mathbb{R}$	$\mathbb{R}^{d}$	$\mathbb{R}^{d+1}$
	$\bar{x} = (x^1/x^0, \dots, x^d/x^0)1_{\{x^0 \neq 0\}}$	$x = (x^0, x^1, \dots, x^d)$
	$\tilde{x} = (x^1, \dots, x^d)$	$\hat{x} = (1, x^1, \dots, x^d)$
c		$\check{c} = (c, 0, \dots, 0)$

In the multivariate notation below, inequalities among vectors are understood componentwise: $x \le y$ means that $x^i \le y^i$ for all i. Also, for a $(d+1)$-dimensional vector x, define $\bar{x}$ as the d-dimensional vector with $\bar{x}^i = (x^i/x^0)1_{\{x^0 \neq 0\}}$, $i = 1, \dots, d$, while $\hat{x}$ denotes the $(d+1)$-dimensional vector with components $\hat{x}^i = x^i$, $i = 1, \dots, d$ and $\hat{x}^0 = 1$. (See Table 1 for a summary of notation.)

THEOREM 3.7. *Let $W \in L^0(\mathbb{R}^{d+1})$, $z \in \mathbb{R}^{d+1}$ and Assumption 2.3 hold. There exists $\phi \in \mathcal{A}$ such that $V_T(z, \phi) \ge W$ a.s. if and only if*

$$Z_0 z \ge E_Q(Z_T W) - E_Q \int_0^T Z_t^0 G_t^*(\bar{Z}_t - S_t)\, dt, \tag{20}$$

for all $Q \in \mathcal{P}(W)$ and for all $\mathbb{R}_+^{d+1}$-valued bounded Q-martingales Z with $Z_0^0 = 1$ satisfying $Z_t^i = 0$, $i = 1, \dots, d$ on $\{Z_t^0 = 0\}$.

REMARK 3.8. Although the above theorem holds for general S, it has the interpretation of a superreplication result only if S (or at least S_T) has nonnegative components and, therefore, a positive number of units of risky positions has positive value. Otherwise, if S can take negative values, a larger number of units does not imply a position with higher value, but only a larger exposure.

Assume in the rest of this remark that S is nonnegative and one-dimensional (for simplicity). Take $\phi \in \mathcal{A}$ and consider the (optional) set $A := \{(\omega, t) : \phi_t(\omega) \neq 0, S_t(\omega) + G(\omega, t, \phi_t(\omega))/\phi_t(\omega) \ge 0\}$, which identifies the times at which execution prices are positive. Clearly, $V_T(z, \phi') \ge V_T(z, \phi)$ for $\phi'_t(\omega) := \phi_t(\omega)1_A$. Hence, in Theorem 3.7 one may replace $\mathcal{A}$ by

$$\mathcal{A}_+ := \{\phi \in \mathcal{A} : S_t(\omega) + G(\omega, t, \phi_t(\omega))/\phi_t(\omega) \ge 0 \text{ when } \phi_t(\omega) \neq 0\}.$$

In other words, the superreplication result continues to hold by considering only trading strategies with positive execution prices at all times, because any other strategy is dominated pointwise by a strategy that trades at the same rate when the execution price is positive, and otherwise does not trade. The class $\mathcal{A}_+$ is economically more appealing as it excludes the unintended consequence of (7) that $S_t(\omega) + G(\omega, t, \phi_t(\omega))/\phi_t(\omega) \to -\infty$ whenever $\phi_t(\omega) \to -\infty$.

The proof of Theorem 3.7 in fact yields also the following slightly different version, in terms of bounded martingales only.

THEOREM 3.9. *Let $W \in L^0(\mathbb{R}^{d+1})$, $z \in \mathbb{R}^{d+1}$ and let Assumption 2.3 hold. Fix a reference probability $Q \in \tilde{\mathcal{P}}(W)$. There exists $\phi \in \mathcal{A}$ such that $V_T(z,\phi) \geq W$ a.s. if and only if*

$$Z_0 z \geq E_Q(Z_T W) - E_Q \int_0^T Z_t^0 G_t^*(\bar{Z}_t - S_t)\,dt, \tag{21}$$

for all $\mathbb{R}_+^{d+1}$-valued bounded Q-martingales Z with $Z_0^0 = 1$ satisfying $Z_t^i = 0$, $i = 1, \dots, d$ on $\{Z_t^0 = 0\}$.

Defining $dQ'/dQ := Z_T^0$ one can state Theorem 3.9 in the following form, in which martingale probabilities Q are replaced by stochastic discount factors Z.

COROLLARY 3.10. *Let $W \in L^0(\mathbb{R}^{d+1})$, $z \in \mathbb{R}^{d+1}$ and Assumption 2.3 hold. Fix a reference probability $Q \in \tilde{\mathcal{P}}(W)$. There exists $\phi \in \mathcal{A}$ such that $V_T(z,\phi) \geq W$ a.s. if and only if*

$$\hat{Z}_0 z \geq E_{Q'}(\hat{Z}_T W) - E_{Q'} \int_0^T G_t^*(Z_t - S_t)\,dt, \tag{22}$$

for all $Q' \ll P$ with bounded dQ'/dQ and for all $\mathbb{R}_+^d$-valued Q'-martingales Z such that $(dQ'/dQ)Z_T$ is bounded.

Finally, in the case of a finite Ω Theorem 3.9 reduces to a simple version, without any integrability conditions.

THEOREM 3.11. *Let Ω be finite. Let $W \in L^0(\mathbb{R}^{d+1})$, $z \in \mathbb{R}^{d+1}$ and let Assumption 2.3 hold. Fix any reference probability $Q \sim P$. There exists $\phi \in \mathcal{A}$ such that $V_T(z,\phi) \geq W$ a.s. if and only if*

$$Z_0 z \geq E_Q(Z_T W) - E_Q \int_0^T Z_t^0 G_t^*(\bar{Z}_t - S_t)\,dt, \tag{23}$$

for all $\mathbb{R}_+^{d+1}$-valued Q-martingales Z with $Z_0^0 = 1$, and satisfying $Z_t^i = 0$, $i = 1, \dots, d$ on $\{Z_t^0 = 0\}$.

PROOF OF THEOREM 3.7. For a $(d+1)$-dimensional vector x, $\tilde{x}$ denotes the d-dimensional vector $\tilde{x}^i := x^i$, $i = 1, \dots, d$ (cf. Table 1). First, assume that $V_T(z,\phi) \geq W$. Take $Q \in \mathcal{P}(W)$ and a bounded Q-martingale Z with nonnegative components [more generally, it is enough to assume that $Z_T W$ is Q-integrable and that $Z_T \in L^\gamma(Q)$], satisfying $Z_t^i = 0$, $i = 1, \dots, d$ on $\{Z_t^0 = 0\}$.

Note that $E_Q|W| < \infty$ and $W^0 \le z + \int_0^T [-\phi_t S_t - G_t(\phi_t)]\,dt$ because $V_T(z,\phi) \ge W$, hence Lemma 3.4 implies

$$E_Q \int_0^T |\phi_t|^\beta (1+|S_t|)^\beta \,dt < \infty. \tag{24}$$

Again, since $V_T(z,\phi) \ge W$, it follows that

$$Z_T(W-z) \le \int_0^T \big[-Z_T^0 \phi_t S_t - Z_T^0 G_t(\phi_t) + \tilde{Z}_T \phi_t\big]\,dt. \tag{25}$$

By (24), Fubini's theorem applies and the properties of conditional expectations imply that

$$\begin{aligned}
E_Q(Z_T W) &\le z E_Q Z_T + E_Q \int_0^T \big[-Z_T^0 \phi_t S_t - Z_T^0 G_t(\phi_t) + \tilde{Z}_T \phi_t\big]\,dt \\
&= z Z_0 + \int_0^T E_Q\big(-Z_T^0 \phi_t S_t - Z_T^0 G_t(\phi_t) + \tilde{Z}_T \phi_t\big)\,dt \\
&= z Z_0 + \int_0^T E_Q\big(-Z_t^0 \phi_t S_t - Z_t^0 G_t(\phi_t) + \tilde{Z}_t \phi_t\big)\,dt \\
&= z Z_0 + \int_0^T E_Q\big(-Z_t^0 \phi_t S_t - Z_t^0 G_t(\phi_t) + Z_t^0 \bar{Z}_t \phi_t\big)\,dt \\
&\le z Z_0 + E_Q \int_0^T Z_t^0 G_t^*(\bar{Z}_t - S_t)\,dt,
\end{aligned}$$

which proves the first implication of this theorem.

To prove the reverse implication, suppose there is no ϕ such that $V_T(z,\phi) \ge W$, which means that $W - z \notin C$. Fix $Q \in \tilde{\mathcal{P}}(W)$. The set $C \cap L^1(Q)$ is closed in $L^1(Q)$ by Proposition 3.5. The Hahn–Banach theorem then provides a nonzero, bounded $(d+1)$-dimensional random variable $\tilde{Z}$ such that

$$E_Q\big[\tilde{Z}(W-z)\big] > \sup_{X \in C \cap L^1(Q)} E_Q[\tilde{Z} X]. \tag{26}$$

Since $-L^0(\mathbb{R}^{d+1}) \subset C$, $\tilde{Z} \ge 0$ a.s., otherwise the supremum would be infinity. Define now the (deterministic) processes $\psi(n,i)$ for all $n \in \mathbb{N}$ and $i = 1, \dots, d$ by setting $\psi_t^i(n,i) := n$, $\psi_t^j(n,i) = 0$, $j \ne i$ for all $t \in [0,T]$.

We claim that $E_Q \tilde{Z}^0 > 0$. Otherwise, for some $i > 0$ one should have $E_Q \tilde{Z}^i > 0$. By Assumption 2.3 $\psi(n,i) \in \mathcal{A}$. By the choice of Q, we even have $V_T(0,\psi(n,i)) \in C \cap L^1(Q)$ and $E_Q \tilde{Z} V_T(0,\psi(n,i)) = nT E_Q \tilde{Z}^i \to \infty$ as $n \to \infty$, which is impossible by (26). So we conclude that $E_Q \tilde{Z}^0 > 0$. Up to a positive multiple of Z, we may assume $E_Q \tilde{Z}^0 = 1$. Define $Z_t := E_Q[\tilde{Z}|\mathcal{F}_t]$, $t \in [0,T]$.

We also claim that, for all $i = 1, \dots, d$,

$$(P \times \mathrm{Leb})(A_i) = 0 \tag{27}$$
$$\text{where } A_i := \{(\omega, t) : Z_t^0(\omega) = 0\} \setminus \{(\omega, t) : Z_t^i(\omega) = 0\}.$$

If this were not the case for some i, define $\psi^i(n,i) := n1_{A_i}$, $\psi^j(n,i) := 0$, $j \neq i$. Clearly, $\psi(n,i) \in \mathcal{A}$ and $V_T(0, \psi(n,i)) \in C \cap L^1(Q)$ while $E_Q \tilde{Z} V_T(0, \psi(n,i)) \to \infty$, $n \to \infty$, which is absurd, proving (27).

By the measurable selection theorem applied to the measure space $(\Omega \times [0,T], \mathcal{O}, P \otimes \mathrm{Leb})$ [see Proposition III.44 in Dellacherie and Meyer (1978)], there is an optional process $\tilde{\chi}(n)$ such that

$$\tilde{\chi}_t(n)[\bar{Z}_t - S_t] - G_t(\tilde{\chi}_t(n)) \le G_t^*(\bar{Z}_t - S_t)$$

and

$$\tilde{\chi}_t(n)[\bar{Z}_t - S_t] - G_t(\tilde{\chi}_t(n)) \ge G_t^*(\bar{Z}_t - S_t) - \frac{1}{n} \ge -K - \frac{1}{n}, \tag{28}$$

for $(P \times \mathrm{Leb})$-almost every (ω, t). Here K denotes the bound for $\sup_{t \in [0,T]} G_t(0)$ from (9). Now define $\chi_t(n) := \tilde{\chi}_t(n) 1_{\{|\tilde{\chi}_t(n)| \le N(n)\}}$ where $N(n)$ is chosen such that $(P \times \mathrm{Leb})(|\tilde{\chi}_t(n)| > N(n)) \le 1/n^2$. By Assumption 2.3, $\chi(n) \in \mathcal{A}$ and by the choice of Q, $V_T(0, \chi(n)) \in C \cap L^1(Q)$. By construction,

$$\lim_{n \to \infty} \chi_t(n)[\bar{Z}_t - S_t] - G_t(\chi_t(n)) = G_t^*(\bar{Z}_t - S_t), \qquad (P \times \mathrm{Leb})\text{-a.e.}$$

Since Z, $\chi(n)$ are bounded and $Q \in \tilde{\mathcal{P}}$ we may use Fubini's theorem and the lower bound in (28) allows the use of Fatou's lemma, hence

$$\begin{aligned}
\liminf_{n \to \infty} E_Q Z_T V_T(0, \chi(n)) &= \liminf_{n \to \infty} E_Q \int_0^T \chi_t(n)[\tilde{Z}_T - Z_T^0 S_t] - Z_T^0 G_t(\chi_t(n))\, dt \\
&= \liminf_{n \to \infty} \int_0^T E_Q[\chi_t(n)[\tilde{Z}_t - Z_t^0 S_t] - Z_t^0 G_t(\chi_t(n))]\, dt \\
&= \liminf_{n \to \infty} E_Q \int_0^T \chi_t(n) Z_t^0 [\bar{Z}_t - S_t] - Z_t^0 G_t(\chi_t(n))\, dt \\
&\ge E_Q \int_0^T Z_t^0 G_t^*(\bar{Z}_t - S_t)\, dt.
\end{aligned}$$

From (26), we infer that

$$\begin{aligned}
z Z_0 &< \limsup_{n \to \infty} [E_Q(W Z_T) - E_Q Z_T V_T(0, \chi(n))] \\
&= E_Q(W Z_T) - \liminf_{n \to \infty} E_Q Z_T V_T(0, \chi(n)) \\
&\le E_Q(W Z_T) - E_Q \int_0^T Z_t^0 G_t^*(\bar{Z}_t - S_t)\, dt.
\end{aligned}$$

This completes the proof. $\square$

REMARK 3.12. The above proof also shows that the statements of Theorems 3.7 and 3.9 remain valid when the class of bounded martingales is replaced by the class of Q-martingales with $Z_T \in L^\gamma(Q)$ such that $Z_T W$ is Q-integrable.

For a real number c, denote by $\check{c}$ the $(d+1)$-dimensional vector $(c, 0, \dots, 0)^T$ (cf. Table 1). The next corollary specializes Theorem 3.7 to the situation in which a claim in cash is hedged from an initial cash position only.

COROLLARY 3.13. *Let $W \in L^0(\mathbb{R})$, $c \in \mathbb{R}$ and let Assumption 2.3 hold. There exists $\phi \in \mathcal{A}$ such that $V_T^0(\check{c}, \phi) \geq W$ a.s. and $V_T^i(\check{c}, \phi) \geq 0$, $i = 1, \dots, d$ if and only if*

$$c \geq E_Q(Z_T^0 W) - E_Q \int_0^T Z_t^0 G_t^*(\bar{Z}_t - S_t)\, dt, \tag{29}$$

for all $Q \in \mathcal{P}(W)$ and for all $\mathbb{R}_+^{d+1}$-valued bounded Q-martingales Z with $Z_0^0 = 1$ satisfying $Z_t^i = 0$, $i = 1, \dots, d$ on $\{Z_t^0 = 0\}$.

To understand the meaning of (29), it is helpful to consider its statement in the frictionless case, at least formally.[9] If S itself is a Q-martingale, then the penalty term with G^* vanishes with the choice of $Z_t^0 := 1$, $Z_t^i := S_t^i$, $i = 1, \dots, d$. It follows that, in order to super-replicate W, the initial endowment c must be greater than or equal to the supremum of $E_Q W$ over the set of equivalent martingale measures for S. This shows that our findings are formally consistent with well-known superhedging theorems for frictionless markets. The results are similarly consistent with superhedging theorems for proportional transaction costs [Kabanov and Safarian (2009)], formally obtained with $G_t(x) = \varepsilon S_t |x|$.

3.5. *Examples*. With the class of superlinear frictions considered in this article, typical contingent claims are virtually impossible to superreplicate with certainty at a fixed price, as we now show. For example, consider the problem of delivering a cash payoff equal to S_T (the price of the risky asset) at time T, starting from cash only. In a market without frictions, or with proportional transaction costs, one solution is to immediately buy the share and, therefore, the superreplication price is at most the (current) price of the asset (or a slightly higher multiple when transaction costs are present).

But this policy is not feasible with superlinear frictions, as block trades are forbidden. An approximate solution would be to buy at rate n over the period $[0, 1/n]$, but this policy incurs a positive probability that the asset price will rapidly increase in value, and in typical models, such as geometric Brownian motion, there is no certain upper bound on the execution price.

[9]The theorem does not apply to the frictionless case because $G = 0$ does not satisfy Assumption 2.3, and feasible strategies differ from admissible strategies.

This discussion motivates the following result.

EXAMPLE 3.14. *Let* $\mu \in \mathbb{R}$, $\sigma, S_0 > 0$, $S_t := S_0 e^{(\mu-\sigma^2/2)t+\sigma W_t}$, $G_t(x) = \frac{\lambda}{2} S_t x^2$, *where* W_t *is a Brownian motion* (*and* $\mathcal{F}_t$ *is its completed filtration made right-continuous*). *Then a cash payoff equal to* S_T *cannot be superreplicated from any initial capital*.

PROOF. In view of Theorem 3.7 above, it is enough to show that the right-hand side of inequality (20) takes arbitrarily large values for a suitable family of reference probabilities Q and martingales Z.

To this end, consider $Q = P$ and the family of exponential martingales Z parameterized by $x > 0$ and $n \in \mathbb{N}$, $n > 1/T$ with

$$\begin{aligned} Z_t^0 = \exp\Big\{ & -\sigma W_{t\wedge(T-1/n)} - \frac{\sigma^2}{2} t \wedge (T-1/n) \\ & + 1_{\{t \geq T-1/n\}} \Big((x-\sigma)(W_t - W_{T-1/n}) - \frac{(x-\sigma)^2}{2} \big(t - (T-1/n)\big) \Big) \Big\} \end{aligned} \tag{30}$$

and $Z_t^1 = S_0 Z_t^0$. [In plain English, Z_t^0 adds a drift of $-\sigma$ (to the Brownian motion) between 0 and $T - 1/n$, and a drift of $x - \sigma$ between $T - 1/n$ and T.] In the sequel, $C_1, C_2, \dots$ denotes various positive constants whose values do not depend either on x or on n.

Notice that, for $0 \leq t \leq T - 1/n$,

$$E Z_t^0 S_t = S_0 e^{(\mu-\sigma^2)t} \leq C_1$$

and for $T - 1/n \leq t \leq T$,

$$\begin{aligned} E Z_t^0 S_t &\leq C_1 e^{(x^2/2)(t-(T-1/n))+(\mu-\sigma^2/2)(t-(T-1/n))-((x-\sigma)^2/2)(t-(T-1/n))} \\ &\leq C_2 e^{\sigma x/n}. \end{aligned}$$

Similarly, for $0 \leq t \leq T - 1/n$,

$$E S_0^2 Z_t^0 / S_t = E S_0 e^{-2\sigma W_t-(\mu-\sigma^2/2)t-(\sigma^2/2)t} \leq C_3$$

and for $T - 1/n \leq t \leq T$,

$$\begin{aligned} E S_0^2 Z_t^0 / S_t &\leq C_3 e^{((x-2\sigma)^2/2)(t-(T-1/n))-(\mu-\sigma^2/2)(t-(T-1/n))-((x-\sigma)^2/2)(t-(T-1/n))} \\ &\leq C_4. \end{aligned}$$

We also have

$$E Z_T^0 S_T \geq S_0 e^{(\mu-\sigma^2)(T-1/n)} e^{(x^2/2)(1/n)+(\mu-\sigma^2/2)(1/n)-((x-\sigma)^2/2)(1/n)} \geq C_5 e^{\sigma x/n}.$$

Now set $x = x(n) = n \ln n/\sigma$. Since $G_t^*(y) = \frac{1}{2\lambda S_t} y^2$, for $W = (S_T, 0)$, which represents a cash payoff equal to the final stock price, it follows that

$$
\begin{aligned}
& E(Z_T W) - E\int_0^T Z_t^0 G_t^*(\bar{Z}_t - S_t)\,dt \\
&\quad = E\big[S_T Z_T^0\big] - \frac{1}{2\lambda}\int_0^T E\bigg[\frac{(Z_t^1)^2}{S_t Z_t^0} - 2Z_t^1 + Z_t^0 S_t\bigg]\,dt \\
&\quad = E\big[S_T Z_T^0\big] - \frac{1}{2\lambda}\int_0^T \bigg(E\bigg[\frac{S_0^2 Z_t^0}{S_t}\bigg] - 2S_0 + E\big[Z_t^0 S_t\big]\bigg)\,dt \qquad (31)\\
&\quad \geq C_5 n - \frac{1}{2\lambda}\int_0^{T-1/n}[C_3 - 2S_0 + C_1]\,dt - \frac{1}{2\lambda n}[C_4 - 2S_0 + C_2 n] \\
&\quad \geq C_5 n - C_6 \to \infty
\end{aligned}
$$

as $n \to \infty$. As a result, the right-hand side takes arbitrarily large values, implying an infinite superreplication price. $\square$

The previous proof uses Theorem 3.7 to obtain a dual characterization of superreplication prices. In fact, the same conclusion can be reached exploiting the market bound obtained in Lemma 3.1.

ALTERNATIVE PROOF. Observe that $G_t(x) = \frac{\lambda}{2} S_t x^2$ implies that $G_t^*(y) = \frac{y^2}{2\lambda S_t}$, whence the market bound is

$$
B = \int_0^T G_t^*(-S_t)\,dt = \frac{1}{2\lambda}\int_0^T S_t\,dt. \qquad (32)
$$

Thus, any strategy starting with initial capital x satisfies the bound

$$
V_T^0(x, \phi) \leq x + \int_0^T G_t^*(-S_t)\,dt \leq x + \frac{1}{2\lambda}\int_0^T S_t\,dt. \qquad (33)
$$

In particular, on the event $\{x + \frac{1}{2\lambda}\int_0^T S_t\,dt < S_T\}$, which has positive probability for any x (because Brownian motion has full support on the space of continuous functions starting at 0) superreplication fails for any strategy, and for any initial capital. $\square$

The previous example should be understood as follows: if a large position in the risky asset needs to be acquired, it is not possible a priori to guarantee a fixed execution price with certainty: price impact prevents the transaction to take place instantly, while over time intervening news may lead the price to arbitrary levels. Yet, the fact that even such a simple contract is not superreplicable with finite wealth raises the question of which contracts have a finite superhedging price, and the next example provides one.

EXAMPLE 3.15. *Let $S_t > 0$ a.s. for all t and $G_t(x) := \frac{\lambda}{2} S_t x^2$. Then, for all $k > 0$, the contract that at time T pays $\frac{1}{\lambda}\int_0^T (\sqrt{1+2k\lambda/S_t}-1)\,dt$ units of the risky asset is superreplicable from initial cash position kT.*

PROOF. The main idea is that this payoff is dominated by a *constant cash-flow* strategy, a strategy that buys the risky asset at the rate of one unit of the safe asset per unit of time (e.g., one dollar per second). To see this, recall the relation between the cash flow and the trading rate

$$dV_t^0 = -\phi_t S_t\,dt - \frac{\lambda}{2} S_t \phi_t^2\,dt. \tag{34}$$

Thus, a constant cash flow $dV_t^0 = -k\,dt$ corresponds to a buying rate

$$\phi_t = \frac{1}{\lambda}\left(-1 + \sqrt{1 + \frac{2\lambda k}{S_t}}\right), \tag{35}$$

which yields at time T exactly $\frac{1}{\lambda}\int_0^T (-1 + \sqrt{1+\frac{2\lambda k}{S_t}})\,dt$ units of the risky asset. In the frictionless limit ($\lambda \downarrow 0$), this strategy implies a buying rate of $\phi_t = k/S_t$, which yields $k\int_0^T 1/S_t\,dt$ units of the risky asset. □

In the above example note that, as k varies, the resulting family of payoffs is not linear, in that while each of the above payoffs are replicable, their multiples need not be. In particular, increasing the buying rate k does not scale the number of units of risky asset bought proportionally, except in the frictionless limit $\lambda = 0$. Note also that the above payoff is superreplicable because it promises a lower number of shares when the asset price is high. The square-root relation is of course linked to the quadratic price impact considered in this example.

4. Arbitrage (of the second kind). Any positive payoff that is superhedged for strictly less than zero is an arbitrage. Such opportunities, which start from an insolvent position and, by clever trading, yield a solvent one, are known in the literature as arbitrage of the second kind, and date back to Ingersoll (1987) [see also Kabanov and Kramkov (1994) in the context of large financial markets]. This definition is used with markets frictions in Dermody and Rockafellar (1991, 1995), and, more recently, in Bouchard and Nguyen Huu (2013), Bouchard and Taflin (2013), Denis and Kabanov (2012), Pennanen (2014), Rásonyi (2009).

The superhedging results in the previous section hold regardless of having arbitrage opportunities or not. Consequently, they can be used to *detect* arbitrage: if we find a nonnegative payoff W satisfying (29) with some $c < 0$ then Corollary 3.13 ensures that an arbitrage opportunity exists.

DEFINITION 4.1. An arbitrage of the second kind is a strategy $\phi \in \mathcal{A}$, such that $V_T(\check{c}, \phi) \geq 0$ for some $c < 0$. Absence of arbitrage of the second kind (NA2) holds if no such opportunity exists.

Note that this definition requires that S has positive components. Otherwise, a nonnegative position in an asset with negative price [as $V_T(\check{c}, \phi) \geq 0$ stipulates] cannot be interpreted as solvent.

The following theorem is a direct consequence of Corollary 3.13 and Remark 3.12.

THEOREM 4.2. *Let Assumption 2.3 hold. Then* (*NA*2) *holds if and only if, for all* $\varepsilon > 0$, *there exists* $Q \in \mathcal{P}$ *and an* $\mathbb{R}_+^{d+1}$-*valued* Q-*martingale* Z *with* $Z_T \in L^\gamma(Q)$ *such that* $E_Q \int_0^T Z_t^0 G_t^*(\bar{Z}_t - S_t)\, dt < \varepsilon$.

A broad class of models enjoys the (NA2) property. Let $D \subset (0, \infty)^d$ be nonempty, open and convex. We denote by $C[t, T](D)$ (resp., $C_x[t, T](D)$) the set of continuous functions f from $[t, T]$ to D [resp., satisfying $f(t) = x$]. Both spaces are equipped with the Borel sets of the topology induced by the uniform metric. Recall that a continuous stochastic process S on $[t, T]$ can be understood as a $C[t, T](D)$-valued random variable, and its support is defined in this (metric) space.

DEFINITION 4.3. A process S has conditional full support in D (henceforth, CFS-D) if $S \in C[0, T](D)$ a.s. and

$$\operatorname{supp} P(S|_{[t,T]} \in \cdot | \mathcal{F}_t) = C_{S_t}[t, T](D) \qquad \text{a.s. for all } t \in [0, T].$$

THEOREM 4.4. *Let Assumption 2.3 hold with* $H_t := H$ *constant. If* S *has the* CFS-D *property, then* (NA2) *holds.*

PROOF. It follows from Theorem 2.6 of Maris and Sayit (2012) that for all ε there is $Q \sim P$ and a Q-martingale M_t evolving in $D \subset \mathbb{R}_+^d$ such that $|S_t - M_t| < \varepsilon$ a.s. for all t. Define $Z_t^i := M_t^i$ for $i = 1, \ldots, d$ and $Z_t^0 := 1$ for all t.

In Maris and Sayit (2012) [see also Guasoni, Rásonyi and Schachermayer (2008)] it is shown that S_T, and hence Z_T are in $L^2(Q)$. A closer inspection of the proof reveals that in fact there exist $Z_T \in L^p(Q)$ for arbitrarily large p. Take $p := \max\{\gamma, \alpha\beta/(\alpha - \beta)\}$. Then Q is easily seen to be in $\mathcal{P}$ and Z_T is in $L^\gamma(Q)$. The estimate (11) in Lemma 3.2 implies that

$$E_Q \int_0^T G_t^*(\bar{Z}_t - S_t)\, dt = E_Q \int_0^T G_t^*(M_t - S_t)\, dt \leq \int_0^T \ell(\varepsilon)\, dt \leq T\ell(\varepsilon)$$

for a continuous (deterministic) function ℓ, which clearly tends to 0 as $\varepsilon \to 0$. Now the claim follows by Theorem 4.2. $\square$

Theorem 4.4 has an immediate implication for fractional Brownian motion. The arbitrage properties of fractional Brownian motion have long been delicate: in a frictionless setting it admits arbitrage of the second kind [Rogers (1997)] but, with proportional transaction costs, it does not even have arbitrage of the first kind [Guasoni, Rásonyi and Schachermayer (2008)]. With price-impact, the above theorem implies that it does not admit arbitrage of the second kind, since it satisfies the CFS-D property [Guasoni, Rásonyi and Schachermayer (2008)]. Whether arbitrage of the first kind (a positive, and possibly strictly positive, payoff from nothing) exists is still an open question.

5. Utility maximization. This section discusses utility maximization in the model of Section 2. The first result (Theorem 5.1 below) shows that optimal strategies exist under a simple integrability assumption, which is easy to check in practice. In particular, optimal strategies exist regardless of arbitrage, since such opportunities are necessarily limited. Put differently, the budget equation is nonlinear. Therefore, one cannot add to an optimal strategy an arbitrage opportunity, and expect the resulting wealth to be the sum.

The second result establishes the first-order condition for utility maximization, which provides a simple criterion for optimality, and helps understand the differences with the corresponding results for frictionless markets and proportional transaction costs. In particular, it shows that the analogue of a shadow price for price-impact models is a hypothetical frictionless price for which the optimal strategy would coincide with the execution price of the same strategy in the original price-impact model. This notion reduces to that of shadow price for markets with proportional transaction costs.

Importantly, these results consider only utilities defined on the real line, such as exponential utility, but exclude power and logarithmic utilities, which are defined only for positive values. This setting is consistent with the definition of feasible strategies, which do not constrain wealth to remain positive. When establishing optimality of a given strategy in such a setting, one technical challenge is to show that the resulting wealth processes are martingales (or just supermartingales) with respect to appropriate reference measures (these are martingale measures in the frictionless case). Lemma 5.6 below implies such a property for any feasible strategy and hence forms the main ingredient of the proof of Theorem 5.5. Finally, since the focus is on utility functions defined on a single variable, and with price impact there is no scalar notion of portfolio value, the results below assume for simplicity that all strategies begin and end with cash only.

Let W be an arbitrary real-valued random variable (representing a random endowment) and $c \in \mathbb{R}$ the investor's initial capital.

THEOREM 5.1. *Let $U : \mathbb{R} \to \mathbb{R}$ be concave and nondecreasing, and let $E|U(c + B + W)| < \infty$ hold for the market bound $B = \int_0^T G_t^*(-S_t)\,dt$ in*

Lemma 3.2. *Under Assumption* 2.3, *there is* $\phi^* \in \mathcal{A}'(U,c)$ *such that*

$$EU\big(V_T^0(\check{c},\phi^*)+W\big)=\sup_{\phi\in\mathcal{A}'(u,c)} EU\big(V_T^0(\check{c},\phi)+W\big),$$

where $\mathcal{A}'(U,c)=\{\phi\in\mathcal{A}: V_T^i(\check{c},\phi)=0, i=1,\dots,d, EU_-(V_T^0(\check{c},\phi)+W)<\infty\}$.

This theorem applies, in particular, for U bounded above and W bounded below.

PROOF OF THEOREM 5.1. Corollary 3.6 implies that

$$C':=\check{c}+\big(C\cap\{X:X^i=0\text{ a.s.}, i=1,\dots,d\}\big)$$

is closed in probability.

Let $\phi(n)$ be a sequence in $\mathcal{A}'(U,c)$ with

$$\lim_{n\to\infty} EU\big(V_T^0(\check{c},\phi(n))+W\big)=\sup_{\phi\in\mathcal{A}'(U,c)} EU\big(V_T^0(\check{c},\phi)+W\big).$$

Since $V_T^0(\check{c},\phi(n))\le c+B$ a.s. for all n, by Lemma 9.8.1 of Delbaen and Schachermayer (2006) there are convex combinations such that $\sum_{j=n}^{M(n)}\alpha_j(n)V_T^0(\check{c},\phi(j))\to V$ a.s. for some $[-\infty,c+B]$-valued random variable V. By convexity of G, we have that for $\tilde{\phi}(n):=\sum_{j=n}^{M(n)}\alpha_j(n)\phi(j)$,

$$V_T^0\big(\check{c},\tilde{\phi}(n)\big)\ge\sum_{j=n}^{M(n)}\alpha_j(n)V_T^0\big(\check{c},\phi(j)\big),$$

so $\sum_{j=n}^{M(n)}\alpha_j(n)V_T(\check{c},\phi(j))\in C'$ for each n.

By the concavity of U,

$$EU\left(W+\sum_{j=n}^{M(n)}\alpha_j(n)V_T^0\big(\check{c},\phi(j)\big)\right)\ge\sum_{j=n}^{M(n)}\alpha_j(n)EU\big(V_T^0\big(\check{c},\phi(j)\big)+W\big).$$

Fatou's lemma implies that $EU(V)\ge\sup_{\phi\in\mathcal{A}'(u)} EU(V_T^0(\check{c},\phi)+W)$, in particular, V is finite-valued and hence $\check{V}\in C'$ by the convexity and closedness of C'. It follows that $V=V_T^0(\check{c},\phi^*)-Y^0$ for some $\phi^*\in\mathcal{A}'(U,c)$ and $Y\in L_+^0$. Clearly, $EU(V_T^0(\check{c},\phi^*)+W-Y^0)=\sup_{\phi\in\mathcal{A}'(U,c)} Eu(V_T^0(\check{c},\phi)+W)$. Necessarily, $EU(V_T^0(\check{c},\phi^*)+W)=\sup_{\phi\in\mathcal{A}'(U,c)} EU(V_T^0(\check{c},\phi)+W)$ as well.[10] This completes the proof. □

[10]Note that U can be constant on an (infinite) interval hence $Y^0\neq 0$ is possible.

REMARK 5.2. Theorem 5.1 can also be proved with

$$\mathcal{A}''(U,c) = \{\phi \in \mathcal{A} : V_T^i(\check{c},\phi) \geq 0, i = 1,\dots,d, EU_-\big(V_T^0(\check{c},\phi) + W\big) < \infty\}$$

in lieu of $\mathcal{A}'(U,c)$. Note that the two optimization problems are *not* equivalent, due to illiquidity.

REMARK 5.3. Let us assume that S is nonnegative and one-dimensional. We may replace $\mathcal{A}'(U,c)$ in Theorem 5.1 by

$$\mathcal{A}'_+(U,c) := \{\phi \in \mathcal{A} : S_t(\omega) + G\big(\omega,t,\phi_t(\omega)\big)/\phi_t(\omega) \geq 0 \text{ when } \phi_t(\omega) \neq 0,$$
$$V_T^i(\check{c},\phi) \geq 0, i = 1,\dots,d, EU_-\big(V_T^0(\check{c},\phi) + W\big) < \infty\},$$

that is, we may restrict our class of strategies to those for which the instantaneous execution price is nonnegative, as in Remark 3.8 above.

REMARK 5.4. The proofs of Theorem 5.1 and Proposition 3.5 use Lemmata 9.8.1 and 15.1.4 in Delbaen and Schachermayer (2006). They could be replaced, with minor modifications, with Komlós's theorem [Komlós (1967)] and its extensions [Balder (1989), v. Weizsäcker (2004)].

While the previous result shows the existence of optimal strategies, the next theorem provides a sufficient conditions for a strategy's optimality, through a variant of the usual first-order condition.

THEOREM 5.5. *Let Assumption 2.3 hold, and*

(a) *let U be concave, continuously differentiable, with U' strictly decreasing, and*

$$U(x) \leq -C|x|^\delta, \qquad x \leq 0, \tag{36}$$

for some $C > 0$ and $\delta > 1$;

(b) *denoting by $\tilde{U}$ the convex conjugate function of U, that is,*

$$\tilde{U}(y) := \sup_{x\in\mathbb{R}}\{U(x) - xy\}, \qquad y > 0;$$

(c) *let W be a bounded random variable;*

(d) *let $Q \in \mathcal{P}$ be such that*

$$dQ/dP \in L^\eta(P), \tag{37}$$

where $(1/\eta) + (1/\delta) = 1$;

(e) *let $G_t(\cdot)$ be $P \times$ Leb-a.s. continuously differentiable in x and $G'_t(\cdot)$ is strictly increasing;*

(f) *let Z be a càdlàg process with $Z_T \in L^{\gamma'}(Q)$ for some $\gamma' > \gamma$ and let ϕ^* be a feasible strategy such that, for some $y^* > 0$, the following conditions hold*:
   (i) *Z is a Q-martingale*;
   (ii) *$U'(V_T^0(x, \phi^*) + W) = y^*(dQ/dP)$ a.s.*;
   (iii) *$Z_t = S_t + G_t'(\phi_t^*)$ a.s. in $P \times$ Leb-a.e.*

Then the strategy ϕ^* is optimal for the problem

$$\max_{\phi \in \mathcal{A}'(U,c)} E\big[U\big(V_T^0(x, \phi) + W\big)\big]. \tag{38}$$

PROOF. For any $(\phi_t)_{t \geq 0} \in \mathcal{A}'(U, c)$ the final payoff equals

$$V_T^0(x, \phi) = x - \int_0^T S_t \phi_t \, dt - \int_0^T G_t(\phi_t) \, dt. \tag{39}$$

Let Z_t be as in the statement of the theorem, and rewrite the above payoff as

$$V_T^0(x, \phi) = x - \int_0^T Z_t \phi_t \, dt + \int_0^T (Z_t - S_t)\phi_t \, dt - \int_0^T G_t(\phi_t) \, dt.$$

By definition of G_t^*, it follows that

$$V_T^0(x, \phi) \leq x - \int_0^T Z_t \phi_t \, dt + \int_0^T G_t^*(Z_t - S_t) \, dt, \tag{40}$$

and equality holds if $Z_t - S_t = G_t'(\phi_t)$, $P \times$ Leb-a.s., that is, when (iii) holds.

It follows from Lemma 5.6 that

$$0 \leq E_Q\bigg[\bigg(x - V_T^0(x, \phi) + \int_0^T G_t^*(Z_t - S_t) \, dt\bigg)\bigg]. \tag{41}$$

Thus, for any payoff $V_T^0(x, \phi) + W$ and any $y > 0$ the following holds:

$$\begin{aligned} &E\big[U\big(V_T^0(x, \phi) + W\big)\big] \\ &\quad \leq E\bigg[U\big(V_T^0(x, \phi) + W\big) \\ &\qquad + y(dQ/dP)\bigg(x - V_T^0(x, \phi) + \int_0^T G_t^*(Z_t - S_t) \, dt\bigg)\bigg] \\ &\quad \leq E\bigg[\tilde{U}\big(y(dQ/dP)\big) + y(dQ/dP)\bigg(\int_0^T G_t^*(Z_t - S_t) \, dt + W\bigg)\bigg] + yx. \end{aligned} \tag{42}$$

If (iii) is satisfied then there is equality in (40) above. If, in addition, (ii) is satisfied then both inequalities in (42) are equalities for $y = y^*$. Thus, if conditions (i), (ii) and (iii) hold for ϕ^* then, by (42),

$$\begin{aligned} &E\big[U\big(V_T^0(x, \phi^*) + W\big)\big] \\ &\quad = E\bigg[\tilde{U}\big(y^*(dQ/dP)\big) + y^*(dQ/dP)\bigg(\int_0^T G_t^*(Z_t - S_t) \, dt + W\bigg)\bigg] + y^* x. \end{aligned}$$

For all $\phi \in \mathcal{A}'(U,c)$

$$E\big[U\big(V_T^0(x,\phi)+W\big)\big]$$
$$\leq E\bigg[\tilde{U}\big(y^*(dQ/dP)\big)+y^*(dQ/dP)\bigg(\int_0^T G_t^*(Z_t-S_t)\,dt+W\bigg)\bigg]+y^*x,$$

again by (42). Hence, the strategy ϕ^* is indeed optimal. □

LEMMA 5.6. *Under the assumptions of the previous theorem, any* $\phi \in \mathcal{A}'(U,c)$ *satisfies*

$$E_Q\int_0^T \phi_t Z_t\,dt = 0.$$

PROOF. Assume $T=1$. Define

$$\Phi_t^+ := \int_0^t (\phi_s)_+\,ds, \qquad \Phi_t^- := \int_0^t (\phi_s)_-\,ds.$$

We show that $E_Q\int_0^1 Z_t\,d\Phi_t^+ - E_Q\int_0^1 Z_t\,d\Phi_t^- = 0$.

Since $\phi \in \mathcal{A}'(U,c)$, (36), (37) and Hölder's inequality imply that $E_Q[V_1^0(x,\phi)]_- < \infty$, hence Lemma 3.4 implies that

$$E_Q\int_0^1 |\phi_t|^\beta\big(1+|S_t|\big)^\beta\,dt < \infty,$$

a fortiori,

$$E_Q\big(\Phi_1^+\big)^\beta = E_Q\bigg(\int_0^1 (\phi_t)_+\,dt\bigg)^\beta < \infty. \tag{43}$$

Define $\Phi_t^+(n) := \Phi^+(k_n(t)/n)$ where

$$k_n(t) := \max\bigg\{i \in \mathbb{N}: \frac{i}{n} \leq t\bigg\}$$

and observe that $d\Phi_t^+(n) \to d\Phi_t^+$ a.s. in the sense of weak convergence of measures on $\mathcal{B}([0,1])$. As Z_t is a.s. càdlàg, its trajectories have countably many points of discontinuity (a.s.). By $d\Phi_t^+ \ll \mathrm{Leb}$, this implies

$$Y_n^+ := \int_0^1 Z_t\,d\Phi_t^+(n) \to \int_0^1 Z_t\,d\Phi_t^+ =: Y^+,$$

almost surely. Furthermore,

$$\begin{aligned}&\bigg|\int_0^1 Z_t\,d\Phi_t^+(n)\bigg|\\ &\quad = \bigg|\sum_{k=1}^n Z_{k/n}\big[\Phi_{k/n}^+(n)-\Phi_{(k-1)/n}^+(n)\big]\bigg| \leq \sup_t |Z_t|\Phi_1^+,\end{aligned} \tag{44}$$

where $\sup_{t\in[0,T]}|Z_t| \in L^{\gamma'}(Q)$ by assumption and $\Phi_1^+ \in L^\beta(Q)$ by (43). It follows by Hölder's inequality that the sequence Y_n^+ is Q-uniformly integrable, so $E_Q Y_n^+ \to E_Q Y^+$, $n \to \infty$. From (44), we get, noting that $\Phi_0^+(n) = 0$,

$$\begin{aligned} E_Q Y_n^+ &= E_Q\left[\sum_{l=0}^{n-1}(Z_{l/n} - Z_{(l+1)/n})\Phi_{l/n}^+(n)\right] + E_Q Z_1 \Phi_1^+(n) \\ &= E_Q Z_1 \Phi_1^+(n), \end{aligned} \tag{45}$$

by the Q-martingale property of Z. Analogously, as $n \to \infty$,

$$E_Q Y_n^- = E_Q Z_1 \Phi_1^-(n) \to E_Q Y^-,$$

where Y_n^- is defined analogously to Y_n^+ using $d\Phi_t^-$ instead of $d\Phi_t^+$ and

$$Y^- := \int_0^1 Z_t \, d\Phi_t^-.$$

Since $\Phi_1(n) = \Phi_1 = 0$, (45) implies that $E_Q(Y_n^+ - Y_n^-) = 0$ for all n, whence also

$$E_Q(Y^+ - Y^-) = E_Q \int_0^T \phi_t Z_t \, dt = 0,$$

completing the proof. $\square$

Acknowledgments. For helpful comments, we thank Bruno Bouchard, Yan Dolinsky, Ioannis Karatzas, Kostas Kardaras, Johannes Muhle-Karbe, Teemu Pennanen, Walter Schachermayer, Mete Soner, and seminar participants at the Oberwolfach Workshop in Stochastic Analysis in Finance and Insurance.

REFERENCES

ALMGREN, R. and CHRISS, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk* **3** 5–40.

ASTIC, F. and TOUZI, N. (2007). No arbitrage conditions and liquidity. *J. Math. Econom.* **43** 692–708. MR2331745

BALDER, E. J. (1989). Infinite-dimensional extension of a theorem of Komlós. *Probab. Theory Related Fields* **81** 185–188. MR0982652

BANK, P. and KRAMKOV, D. (2011a). A model for a large investor trading at market indifference prices. I: Single-period case. Preprint. Available at arXiv:1110.3224.

BANK, P. and KRAMKOV, D. (2011b). A model for a large investor trading at market indifference prices. II: Continuous-time case. Preprint. Available at arXiv:1110.3229.

BERTSIMAS, D. and LO, A. (1998). Optimal control of execution costs. *Journal of Financial Markets* **1** 1–50.

BLACK, F. (1986). Noise. *J. Finance* **41** 529–543.

BOUCHARD, B. and NGUYEN HUU, A. (2013). No marginal arbitrage of the second kind for high production regimes in discrete time production-investment models with proportional transaction costs. *Math. Finance* **23** 366–386. MR3034082

BOUCHARD, B. and TAFLIN, E. (2013). No-arbitrage of second kind in countable markets with proportional transaction costs. *Ann. Appl. Probab.* **23** 427–454. MR3059265

ÇETIN, U., JARROW, R. A. and PROTTER, P. (2004). Liquidity risk and arbitrage pricing theory. *Finance Stoch.* **8** 311–341. MR2213255

ÇETIN, U. and ROGERS, L. C. G. (2007). Modeling liquidity effects in discrete time. *Math. Finance* **17** 15–29. MR2281790

ÇETIN, U., SONER, H. M. and TOUZI, N. (2010). Option hedging for small investors under liquidity costs. *Finance Stoch.* **14** 317–341. MR2670416

CONT, R., KUKANOV, A. and STOIKOV, S. (2014). The price impact of order book events. *Journal of Financial Econometrics* **12** 47–88.

DELBAEN, F. and SCHACHERMAYER, W. (2006). *The Mathematics of Arbitrage*. Springer, Berlin. MR2200584

DELLACHERIE, C. and MEYER, P.-A. (1978). *Probabilities and Potential. North-Holland Mathematics Studies* **29**. North-Holland, Amsterdam. MR0521810

DELLACHERIE, C. and MEYER, P.-A. (1982). *Probabilities and Potential. B: Theory of Martingales. North-Holland Mathematics Studies* **72**. North-Holland, Amsterdam. MR0745449

DENIS, E. and KABANOV, Y. (2012). Consistent price systems and arbitrage opportunities of the second kind in models with transaction costs. *Finance Stoch.* **16** 135–154. MR2872651

DERMODY, J. C. and ROCKAFELLAR, R. T. (1991). Cash stream valuation in the face of transaction costs and taxes. *Math. Finance* **1** 31–54.

DERMODY, J. C. and ROCKAFELLAR, R. T. (1995). Tax basis and nonlinearity in cash stream valuation. *Math. Finance* **5** 97–119.

DOLINSKY, Y. and SONER, H. M. (2013). Duality and convergence for binomial markets with friction. *Finance Stoch.* **17** 447–475. MR3066984

DUFOUR, A. and ENGLE, R. F. (2000). Time and the price impact of a trade. *J. Finance* **55** 2467–2498.

FERNHOLZ, R., KARATZAS, I. and KARDARAS, C. (2005). Diversity and relative arbitrage in equity markets. *Finance Stoch.* **9** 1–27. MR2210925

GARLEANU, N. and PEDERSEN, L. (2013). Dynamic trading with predictable returns and transaction costs. *J. Finance* **68** 2309–2340.

GUASONI, P., LÉPINETTE, E. and RÁSONYI, M. (2012). The fundamental theorem of asset pricing under transaction costs. *Finance Stoch.* **16** 741–777. MR2972240

GUASONI, P., RÁSONYI, M. and SCHACHERMAYER, W. (2008). Consistent price systems and face-lifting pricing under transaction costs. *Ann. Appl. Probab.* **18** 491–520. MR2398764

INGERSOLL, J. E. (1987). *Theory of Financial Decision Making*. Rowman & Littlefield, Lanham, MD.

JOUINI, E. and KALLAL, H. (1995). Martingales and arbitrage in securities markets with transaction costs. *J. Econom. Theory* **66** 178–197. MR1338025

KABANOV, YU. M. and KRAMKOV, D. O. (1994). Large financial markets: Asymptotic arbitrage and contiguity. *Teor. Veroyatn. Primen.* **39** 222–229. MR1348197

KABANOV, Y. and SAFARIAN, M. (2009). *Markets with Transaction Costs: Mathematical Theory*. Springer, Berlin. MR2589621

KALLSEN, J. and MUHLE-KARBE, J. (2010). On using shadow prices in portfolio optimization with transaction costs. *Ann. Appl. Probab.* **20** 1341–1358. MR2676941

KARATZAS, I. and KARDARAS, C. (2007). The numéraire portfolio in semimartingale financial models. *Finance Stoch.* **11** 447–493. MR2335830

KOMLÓS, J. (1967). A generalization of a problem of Steinhaus. *Acta Math. Acad. Sci. Hungar.* **18** 217–229. MR0210177

KYLE, A. (1985). Continuous auctions and insider trading. *Econometrica* **29** 1315–1335.

MARIS, F. and SAYIT, H. (2012). Consistent price systems in multiasset markets. *Int. J. Stoch. Anal.* Art. ID 687376, 14. MR2964813

PENNANEN, T. (2011a). Superhedging in illiquid markets. *Math*. *Finance* **21** 519–540. MR2830432

PENNANEN, T. (2011b). Arbitrage and deflators in illiquid markets. *Finance Stoch*. **15** 57–83. MR2764443

PENNANEN, T. (2014). Optimal investment and contingent claim valuation in illiquid markets. *Finance Stoch.* To appear.

PENNANEN, T. and PENNER, I. (2010). Hedging of claims with physical delivery under convex transaction costs. *SIAM J*. *Financial Math*. **1** 158–178. MR2658577

RÁSONYI, M. (2009). Arbitrage under transaction costs revisited. In *Optimality and Risk—Modern Trends in Mathematical Finance* 211–225. Springer, Berlin. MR2648605

ROGERS, L. C. G. (1997). Arbitrage with fractional Brownian motion. *Math*. *Finance* **7** 95–105. MR1434408

ROGERS, L. C. G. and SINGH, S. (2010). The cost of illiquidity and its effects on hedging. *Math*. *Finance* **20** 597–615. MR2731409

SCHACHERMAYER, W. (2004). The fundamental theorem of asset pricing under proportional transaction costs in finite discrete time. *Math*. *Finance* **14** 19–48. MR2030834

SCHIED, A. and SCHÖNEBORN, T. (2009). Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. *Finance Stoch*. **13** 181–204. MR2482051

V. WEIZSÄCKER, H. (2004). Can one drop L^1-boundedness in Komlós's subsequence theorem? *Amer. Math*. *Monthly* **111** 900–903. MR2104694

DEPARTMENT OF MATHEMATICS AND STATISTICS
BOSTON UNIVERSITY
111 CUMMINGTON STREET
BOSTON, MASSACHUSETTS 02215
USA
AND
SCHOOL OF MATHEMATICAL SCIENCES
DUBLIN CITY UNIVERSITY
GLASNEVIN, DUBLIN 9
IRELAND
E-MAIL: guasoni@bu.edu

MTA ALFRÉD RÉNYI MATHEMATICAL INSTITUTE
REÁLTANODA UTCA 13-15
1053 BUDAPEST
HUNGARY
AND
SCHOOL OF MATHEMATICS
UNIVERSITY OF EDINBURGH
KING'S BUILDINGS
EDINBURGH, EH9 3JZ
UNITED KINGDOM
E-MAIL: rasonyi.miklos@renyi.mta.hu