

Scaling limits for super-replication with transient price impact

PETER BANK¹ and YAN DOLINSKY²

¹*Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany.*
E-mail: bank@math.tu-berlin.de

²*Department of Statistics, Hebrew University, Mount Scopus, Jerusalem 91905, Israel.*
E-mail: yan.dolinsky@mail.huji.ac.il

We prove a scaling limit theorem for the super-replication cost of options in a Cox–Ross–Rubinstein binomial model with transient price impact. The correct scaling turns out to keep the market depth parameter constant while resilience over fixed periods of time grows in inverse proportion with the duration between trading times. For vanilla options, the scaling limit is found to coincide with the one obtained by PDE-methods in (*Math. Finance* 22 (2012) 250–276) for models with purely temporary price impact. These models are a special case of our framework and so our probabilistic scaling limit argument allows one to expand the scope of the scaling limit result to path-dependent options.

Keywords: binomial model; liquidity; scaling limit; super-replication; transient price impact

1. Introduction

Super-replication in continuous-time financial models with market frictions is well known to typically lead to trivial buy-and-hold results. For markets with proportional transaction costs, this was first established rigorously by Shreve et al. [18]. For discrete-time models, Kusuoka [14] used a dual description of super-replication costs to determine a regime that yields a non-trivial scaling limit for vanishing transaction costs. Such scaling limits were recently obtained in the multivariate case in [6], for fixed costs in [3], and for purely temporary nonlinear costs as specified by Çetin et al. [8] in Gökay and Soner [12], Dolinsky and Soner [10], Bank et al. [5].

The present paper yields such a scaling limit result for models with transient price impact where also past trades affect the spread at which present transactions are executed; see [2, 15, 16] for models of this type for optimal liquidation problems and [7] for an optimal investment study of such a model. The present paper is motivated by Bank and Dolinsky [4] which confirms the triviality of super-replication costs also for continuous-time models with transient price impact. We therefore introduce in this paper a discrete-time version of the model considered in [4] and, for the special case of a binomial Cox–Ross–Rubinstein reference model, compute the scaling limit of super-replication costs when market resilience becomes infinite.

It turns out that the resulting scaling limit coincides with the scaling limit obtained for binomial models with purely temporary price impact and modified market depth, as studied (for the geometric random walk case) in [5, 10, 12]. In this regard, it nicely complements the high-resilience asymptotics carried out by Roch and Soner [17] who prove convergence in probability of wealth dynamics.

Our approach for computing the scaling limit is purely probabilistic. The proof of the lower bound is done in two steps. In the first step, we establish a simple lower bound for the super-replication prices in terms of consistent price systems with “small” spread. The second step is to use Kusuoka’s techniques from [14] to construct, for a given martingale M on Wiener space with suitably regular volatility process, a sequence of consistent price systems for our binomial reference models with vanishing spread which converge in law to M . Kusuoka’s techniques are particularly useful here as they also allow us to control the approximation of the quadratic variation of M .

The proof of the upper bound is more complicated. First, we notice that the portfolio value in the transient price impact dominates from above the portfolio value in a quadratic costs setup with a modified market depth which can be viewed as a binomial version of the temporary price impact model introduced in [8]. The key step is then to establish an upper bound for the super-replication prices with such quadratic costs. Using the pathwise Doob inequalities of Acciaio et al. [1], we argue that it essentially suffices to super-replicate the payoff knocked out when the underlying fluctuates “too much”. For this “tamed” payoff, we identify a rich enough subclass of constrained trading strategies, for which super-replication costs remain unchanged asymptotically, but whose dual consistent price systems turn out to be tight. This new technique to obtain tightness in fully quadratic costs problems is key for our analysis and allows us to resolve an open question from [5, 10] who had to impose linear growth constraints on transaction costs and only allowed quadratic costs in an ever smaller region around zero. As a by-product of our probabilistic approach, we obtain an extension of the limit result of Gökay and Soner [12], who used PDE-techniques, from vanilla options to path-dependent options.

The paper is organized as follows. In Section 2, we formalize the super-replication problem with transient price impact and give a duality result. Section 3 formulates and discusses our scaling limit result and gives its proof.

2. Super-replication with transient price impact in discrete time

2.1. A discrete-time model with transient price impact

In this section, we develop and analyze a discrete-time version of the continuous-time financial model studied in [4] where the trades of a large investor affect an asset’s price in a transient manner. Specifically, we fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0, \dots, N}, \mathbb{P})$ and consider an adapted, real-valued process $P = (P_n)_{n=0, \dots, N}$ to describe the evolution of an asset’s fundamental value at times $n = 0, \dots, N$. In addition to this asset, a large investor has at her disposal a bank account that, for simplicity, bears no interest. She is endowed with an initial position of $X_0 \triangleq x_0 \in \mathbb{R}$ units of the asset and is free to choose her position $X_n \in \mathcal{F}_{n-1}$ in which she will confront the n th fundamental shock $\Delta P_n \triangleq P_n - P_{n-1}$, $n = 1, \dots, N$. We will let denote the collection of all these strategies X . In line with Huberman and Stanzl [13], the investor’s transactions have a linear permanent impact on the asset’s price beyond its fundamental value. So, the mid-price evolves according to

$$P_n^X \triangleq P_n + \iota X_n, \quad n = 0, \dots, N.$$

In addition, the investor’s transactions affect the half-spread, that is, the mark-up above (resp. below) the mid-price P^X at which the investor’s orders are filled when she buys (respectively sells) the asset. We model this quantity by

$$\begin{aligned} \zeta_0^X &\triangleq \zeta_0, \\ \zeta_n^X &\triangleq (1-r)\zeta_{n-1}^X + \frac{|X_n - X_{n-1}|}{\delta}, \quad n = 1, \dots, N. \end{aligned} \tag{2.1}$$

Here, $\zeta_0 \geq 0$ is the given initial half-spread. The investor’s trades widen the spread in inverse proportion to the market’s depth $\delta > 0$, assumed to be constant for simplicity. The constant $0 < r \leq 1$ measures the market’s resilience and describes the fraction by which the spread will diminish over a trading period. It is convenient (and quite appropriate) to assume that transactions affect mid-prices and spreads gradually, letting the first bits of the n th transaction $X_n - X_{n-1}$ be filled at the favorable pre-transaction mid-price $P_{n-1} + \iota X_{n-1}$ and at the pre-transaction spread $(1-r)\zeta_{n-1}^X$ while the last bits are filled at the less favorable post-transaction levels $P_{n-1} + \iota X_n = P_n^X - \Delta P_n$ and $(1-r)\zeta_{n-1}^X + |X_n - X_{n-1}|/\delta = \zeta_n^X$. As a result, the investor’s given cash position ξ^X evolves from its given initial level $\xi_0 \in \mathbb{R}$ according to

$$\begin{aligned} \xi_0^X &\triangleq \xi_0, \\ \xi_n^X &\triangleq \xi_{n-1}^X - \left(P_{n-1} + \frac{\iota}{2}(X_n + X_{n-1}) \right) (X_n - X_{n-1}) \\ &\quad - \left((1-r)\zeta_{n-1}^X + \frac{1}{2\delta}|X_n - X_{n-1}| \right) |X_n - X_{n-1}| \end{aligned} \tag{2.2}$$

at times $n = 1, \dots, N$. A more tangible description of the investor’s cash positions is given by the following lemma.

Lemma 2.1. *The investor’s cash position at time $n = 1, \dots, N$ is*

$$\begin{aligned} \xi_n^X &= \xi_0 - \sum_{m=1}^n P_{m-1}(X_m - X_{m-1}) - \frac{\iota}{2}(X_n^2 - x_0^2) - \kappa_n^X \\ &= \xi_0 + x_0 P_0 - X_n P_n + \sum_{m=1}^n X_m(P_m - P_{m-1}) - \frac{\iota}{2}(X_n^2 - x_0^2) - \kappa_n^X, \end{aligned} \tag{2.3}$$

where κ^X describes the liquidity costs

$$\kappa_n^X = (1-r) \sum_{m=1}^n \zeta_{m-1}^X |X_m - X_{m-1}| + \frac{1}{2\delta} \sum_{m=1}^n (X_m - X_{m-1})^2 \tag{2.4}$$

$$= \frac{\delta}{2} \left((\zeta_n^X)^2 + (1 - (1-r)^2) \sum_{1 \leq m < n} (\zeta_m^X)^2 - (1-r)^2 \zeta_0^2 \right) \tag{2.5}$$

with

$$\zeta_n^X = (1 - r)^n \zeta_0 + \frac{1}{\delta} \sum_{m=1}^n (1 - r)^{n-m} |X_m - X_{m-1}|, \quad n = 1, \dots, N.$$

Proof. Identity (2.3) follows readily from (2.2) where the representation (2.4) of κ_n^X is due to $\sum_{m=1}^n (X_m + X_{m-1})(X_m - X_{m-1}) = \sum_{m=1}^n X_m^2 - X_{m-1}^2 = X_n^2 - x_0^2$; (2.5) follows readily by expressing $|X_m - X_{m-1}|$ in terms of ζ_m^X and ζ_{m-1}^X as made possible by (2.1). \square

In particular, we see that the liquidity costs κ^X are a convex functional of the investor’s trading strategy $X \in \mathcal{X}$. This observation opens the door for convex duality methods that indeed will be key for our subsequent analysis.

2.2. Super-replication duality

Having established the investor’s wealth dynamics, we can now consider the problem to super-replicate a contingent claim specified by a payoff $H \in \mathcal{F}_N$ unaffected by the investor’s transactions. More precisely, we will try to characterize the super-replication costs

$$\pi(H) \triangleq \inf \{ \xi_0^X : \xi_N^X \geq H \text{ a.s. for some } X \in \mathcal{X} \text{ with } X_N = 0 \}.$$

For models with full resilience ($r = 1$) as in [8], a dual description of super-replication costs has been obtained in [10]. For models with limited resilience ($r \in (0, 1)$) such a description is given by the following lemma which complements its continuous-time analogue established in [4]:

Proposition 2.2. *If $r \in [0, 1)$, the super-replication costs of any contingent claim $H \geq 0$ have the dual description*

$$\pi(H) = \sup_{(\mathbb{Q}, M, \alpha)} \left\{ \mathbb{E}_{\mathbb{Q}}[H] - \frac{1}{2} \mathbb{E}_{\mathbb{Q}} \left[\sum_{n=1}^N |\alpha_n - \zeta_0|^2 \mu_n \right] - M_0 x_0 - \frac{\iota}{2} x_0^2 \right\}, \quad (2.6)$$

where

$$\mu_n \triangleq \delta(1 - (1 - r)^2)(1 - r)^{2n} \quad \text{for } n = 1, \dots, N - 1, \text{ and } \mu_N = \delta(1 - r)^{2N},$$

and where the supremum is taken over all triples (\mathbb{Q}, M, α) of measures $\mathbb{Q} \ll \mathbb{P}$, square-integrable \mathbb{Q} -martingales M and \mathbb{Q} -square-integrable, predictable processes α with

$$|P_{n-1} - M_{n-1}| \leq \frac{1}{\delta(1 - r)^n} \mathbb{E}_{\mathbb{Q}} \left[\sum_{m=n}^N \alpha_m \mu_m \mid \mathcal{F}_{n-1} \right], \quad n = 1, \dots, N.$$

In the case $r = 1$ corresponding to purely temporary impact, we have the simpler duality

$$\pi(H) = \sup_{(\mathbb{Q}, M)} \left\{ \mathbb{E}_{\mathbb{Q}}[H] - \frac{1}{2\delta} \mathbb{E}_{\mathbb{Q}} \left[\sum_{n=1}^N |P_{n-1} - M_{n-1}|^2 \right] - M_0 x_0 - \frac{\iota}{2} x_0^2 \right\}$$

with a supremum over probabilities $\mathbb{Q} \ll \mathbb{P}$ and all square-integrable \mathbb{Q} -martingales M .

Proof. Using the wealth dynamics of Lemma 2.1, the proof can be done similarly as in the continuous-time analogue in [4] and is therefore omitted. For $r = 1$ one can proceed as in [10] together with the Lagrange multiplier argument for the choice of martingale M from [4]. \square

So, super-replication costs in our model with price impact take the form of a convex risk measure. The structure of the costs' dual description is similar in spirit to the one observed for proportional transaction costs models: the payoff's assessment is made using consistent price systems with a martingale M that is in some sense close to the underlying's price process P . By contrast to these models with fixed spread, closeness is measured in our setting by a process α that needs to be chosen to balance greater flexibility in choosing M with higher penalties from the L^2 -distance to the initial spread arising in the Legendre-Fenchel representation (2.6) of the super-replication cost functional.

As illustrated in [4], super-replication prices in continuous-time often are trivially arising from simple buy-and-hold strategies that cannot be improved upon due to the most unlikely, but nonetheless still most relevant strong short-term fluctuations in the price of the hedging instrument that are typically possible in these models. Similar to Kusuoka's approach in [14] to discrete-time models with fixed spread, we thus need to re-scale price impact to ensure a non-trivial scaling limit for our model. This will be made precise in the next section.

3. Scaling limit of super-replication costs

In this section, we will derive a scaling limit result for the super-replication costs from the previous section, letting the number of trading periods N over the time span $[0, 1]$ tend to infinity while re-scaling the time between trades as $1/N$. For the price fluctuations we now focus on a binomial model where $\Omega = \{-1, +1\}^{\{1,2,\dots\}}$ with coordinate maps $\xi_n(\omega) = \omega_n$ indicating the upwards and downwards movements of the fundamental asset value for scenario $\omega = (\omega_n)_{n=1,2,\dots} \in \Omega$. The filtration $(\mathcal{F}_n)_{n=1,2,\dots}$ is generated by these coordinate maps and we assume \mathbb{P} to be the measure under which ξ_1, ξ_2, \dots are i.i.d. with $\mathbb{P}[\xi_n = -1] = \mathbb{P}[\xi_n = +1] = 1/2$. Assuming an additive model for the fundamental asset price we let, with the usual square root scaling,

$$P_t^N \triangleq p_0 + \frac{\sigma}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \xi_n, \quad 0 \leq t \leq 1, \tag{3.1}$$

where $p_0 \in \mathbb{R}$ is the initial fundamental asset price and $\sigma > 0$ the asset's volatility. The price impact parameters $\iota \geq 0$, $r \in (0, 1]$ and $\delta > 0$ are kept constant as we re-scale. As a result, the

same resilience effect is obtained over ever shorter time periods $1/N$, implying a high-resilience limit in our scaling.

The main result of this paper is a scaling limit theorem for the super-replication price of payoff profiles h for which we will need the following regularity assumption.

Assumption 3.1. The functional $h : D[0, 1] \rightarrow \mathbb{R}$ is nonnegative and Lipschitz continuous with respect to the Skorohod metric

$$d(p, q) \triangleq \inf_{\chi} \left\{ \sup_{0 \leq t \leq 1} |t - \chi(t)| + \sup_{0 \leq t \leq 1} |p(t) - q(\chi(t))| \right\}, \quad p, q \in D[0, 1],$$

where the infimum is over all strictly increasing continuous time changes $\chi : [0, 1] \rightarrow [0, 1]$ with $\chi(0) = 0$ and $\chi(1) = 1$.

Observe that the maps $p \rightarrow p(1)$, $p \rightarrow \sup_{0 \leq t \leq T} p(t)$ are Lipschitz continuous with respect to the above Skorohod metric. Hence, call options, put options and lookback options are covered in our setup; knock-out features, however, will typically lead to discontinuities not covered by our assumption.

This puts us in a position to state our limit theorem:

Theorem 3.2. For a payoff profile h satisfying Assumption 3.1, the super-replication costs $\pi^N(h(P^N))$ in the N -period model, $N = 1, 2, \dots$, have the high-resilience scaling limit

$$\begin{aligned} \lim_N \pi^N(h(P^N)) &= \sup_{v \in \mathcal{D}} \mathbb{E}_{\mathbb{P}^W} \left[h(P^v) - \frac{r\delta}{8\sigma^2(2-r)} \int_0^1 |v_t^2 - \sigma^2|^2 dt \right] \\ &\quad - P_0 x_0 - \frac{l}{2} x_0^2, \end{aligned}$$

where \mathcal{D} is the set of all bounded, nonnegative progressively measurable processes v on the Wiener space $(\Omega^W, \mathcal{F}^W, (\mathcal{F}_t^W)_{0 \leq t \leq 1}, \mathbb{P}^W)$ with Wiener process W and where

$$P_t^v \triangleq p_0 + \int_0^t v_s dW_s, \quad 0 \leq t \leq 1.$$

The preceding theorem identifies the scaling limit of our discrete-time super-replication prices in the form of a convex risk measure. This measure assigns to a model, identified through its volatility profile v , a penalty that is determined by its local variances's L^2 -distance from the reference variance σ^2 . Interestingly, this is also the scaling limit that emerges from price impact models with purely temporary impact, albeit with a different weight; see [5,10,12].

The connection between transient and temporary impact for high-resilience limits has been observed before in [17] who prove convergence in probability for the value processes. Our result complements this with a first rigorous result in the context of super-replication.

On a technical level, it is worth mentioning that, to the best of our knowledge, our proof below is the first purely probabilistic approach which allows one to obtain a scaling limit result

with fully quadratic temporary costs. As a result, we are able to cover also sufficiently regular, path-dependent options, thus extending beyond the vanilla option case covered by the viscosity solution techniques of Gökay and Soner [12]. The key challenge here is to find a setting where one can prove tightness for a suitable sequence of dual variables. This challenge is met by a judiciously chosen set of constrained hedging strategies in our proof of the upper bound.

3.1. Proof of the lower bound

In this section, we will prove

$$\liminf_N \pi^N(h(P^N)) \geq \sup_{v \in \mathcal{D}} \mathbb{E}_{\mathbb{P}^W} \left[h(P^v) - \frac{r\delta}{8\sigma^2(2-r)} \int_0^1 |v_t^2 - \sigma^2|^2 dt \right] - P_0 x_0 - \frac{l}{2} x_0^2. \tag{3.2}$$

Let us start by observing that, by the density arguments of Lemma 7.3 in [10], the above supremum coincides with the one taken over the class \mathcal{D}_0 of volatility profiles $v \in \mathcal{D}$ which are bounded away from zero and Lipschitz in the sense that for some constant $C > 0$ we have

$$v_t(\omega) \geq 1/C,$$

$$|v_t(\omega) - v_{t'}(\omega')| \leq C \left(|t - t'| + \sup_{s \in [0,1]} |\omega(s) - \omega'(s)| \right)$$

for $t, t' \in [0, 1]$, $\omega, \omega' \in C[0, 1]$. For any such v , the seminal paper [14] constructs probabilities with martingales “close” to the random walk P^N which in distribution converge to $P^v = p_0 + \int_0^\cdot v_s dW_s$ as summarized in the following lemma.

Lemma 3.3. *For any $v \in \mathcal{D}_0$, there is a sequence of probability measures \mathbb{Q}^N on (Ω, \mathcal{F}_N) and $(\mathcal{F}_n)_{n=0, \dots, N}$ -predictable processes $\alpha^N = (\alpha_n^N)_{n=1, \dots, N}$, $N = 1, 2, \dots$, such that for some constant $C > 0$ independent of N we have*

1. $|\alpha_n^N| \leq C, |\alpha_n^N - \alpha_{n-1}^N| \leq C/\sqrt{N}, n = 1, \dots, N;$
2. $M_0^N \triangleq P_0, M_n^N \triangleq P_{n/N}^N + \alpha_n^N \xi_n / \sqrt{N}, n = 1, \dots, N,$ is a \mathbb{Q}^N -martingale;
3. $\text{Law}((P_{[Nt]}^N, \alpha_{[Nt]}^N)_{0 \leq t \leq 1} \mid \mathbb{Q}^N) \rightarrow \text{Law}((P_t^v, (v_t^2 - \sigma^2)/(2\sigma))_{0 \leq t \leq 1} \mid \mathbb{P}^W)$ weakly on $D[0, 1]$ as $N \uparrow \infty$.

Proof. Adjusting for the additive setting considered here, this follows exactly as in Kusuoka’s original approach for the multiplicative geometric random walk setting from [14]. □

With the above approximation result and the representation of liquidity costs (2.4), (2.5) at hand, we are now in a position to prove (3.2). Indeed, take \mathbb{Q}^N and α^N as in the preceding

lemma and observe that, for any N -period strategy $X = (X_n)_{n=0,\dots,N}$ with $X_N = 0$, we can estimate

$$\begin{aligned}
 & -\mathbb{E}_{\mathbb{Q}^N} \left[\sum_{m=1}^N P_{(m-1)/N}^N (X_m - X_{m-1}) \right] \\
 &= -\mathbb{E}_{\mathbb{Q}^N} \left[\sum_{m=1}^N (P_{(m-1)/N}^N - M_N^N) (X_m - X_{m-1}) + M_N^N (X_N - x_0) \right] \\
 &= -\mathbb{E}_{\mathbb{Q}^N} \left[\sum_{m=1}^N (P_{(m-1)/N}^N - M_{m-1}^N) (X_m - X_{m-1}) \right] + M_0^N x_0 \\
 &\leq \mathbb{E}_{\mathbb{Q}^N} \left[\sum_{m=1}^N \frac{|\alpha_{m-1}^N|}{\sqrt{N}} |X_m - X_{m-1}| \right] + M_0^N x_0 \\
 &= \mathbb{E}_{\mathbb{Q}^N} \left[\sum_{m=1}^N \frac{|\alpha_{m-1}^N|}{\sqrt{N}} \delta (\zeta_m^X - (1-r)\zeta_{m-1}^X) \right] + P_0 x_0,
 \end{aligned}$$

where we used the martingale property of M^N along with $X_N = 0$ for the second identity and the second property of α^N listed in Lemma 3.3 for the estimate. Hence, for $X \in \mathcal{X}$ with $X_N = 0$ which super-replicates $h(P^N)$ in the sense that $\xi_N^X \geq h(P^N)$ we can estimate

$$\begin{aligned}
 & \mathbb{E}_{\mathbb{Q}^N} [h(P^N)] \\
 & \leq \mathbb{E}_{\mathbb{Q}^N} [\xi_N^X] \\
 & = \xi_0 - \mathbb{E}_{\mathbb{Q}^N} \left[\sum_{m=1}^N P_{(m-1)/N}^N (X_m - X_{m-1}) + \frac{l}{2} (X_N^2 - x_0^2) + \kappa_N^X \right] \\
 & \leq \xi_0 + P_0 x_0 + \frac{l}{2} x_0^2 \\
 & \quad + \delta \frac{\alpha_0^N}{\sqrt{N}} \xi_0 - \frac{1}{2} (\xi_0)^2 \\
 & \quad + \delta \mathbb{E}_{\mathbb{Q}^N} \left[\sum_{1 \leq m < N} \left(\frac{|\alpha_{m-1}^N|}{\sqrt{N}} - (1-r) \frac{|\alpha_m|}{\sqrt{N}} \right) \zeta_m^X - \frac{1 - (1-r)^2}{2} (\zeta_m^X)^2 \right] \\
 & \quad + \delta \mathbb{E}_{\mathbb{Q}^N} \left[\frac{\alpha_{N-1}^N}{\sqrt{N}} \zeta_N^X - \frac{1}{2} (\zeta_N^X)^2 \right].
 \end{aligned}$$

Using the estimate $a\zeta - \frac{c}{2}\zeta^2 \leq a^2/(2c)$ in each of the last three lines and rearranging terms yields

$$\begin{aligned} & \xi_0 + P_0x_0 + \frac{\iota}{2}x_0^2 + \frac{\delta(\alpha_0^N)^2}{2N} + \delta\mathbb{E}_{\mathbb{Q}^N} \left[\frac{(\alpha_{N-1}^N)^2}{2N} \right] \\ & \geq \mathbb{E}_{\mathbb{Q}^N} [h(P^N)] - \delta\mathbb{E}_{\mathbb{Q}^N} \left[\sum_{1 \leq m < N} \frac{(|\alpha_{m-1}^N| - (1-r)|\alpha_m|)^2}{2(1-(1-r)^2)N} \right] \\ & \geq \mathbb{E}_{\mathbb{Q}^N} [h(P^N)] - \frac{\delta}{2(1-(1-r)^2)} \mathbb{E}_{\mathbb{Q}^N} \left[\frac{1}{N} \sum_{1 \leq m < N} (r|\alpha_{m-1}^N| + C/\sqrt{N})^2 \right], \end{aligned}$$

where in the last estimate we used the first property of α^N from Lemma 3.3. The same property also yields the uniform boundedness of α^N , $N = 1, 2, \dots$, and so the third property listed in Lemma 3.3 in conjunction with the regularity assumption 3.1 on h thus allows us to pass to the limit $N \uparrow \infty$ in the above estimate to conclude that

$$\begin{aligned} & \liminf_N \pi^N(h(P^N)) + P_0x_0 + \frac{\iota}{2}x_0^2 \\ & \geq \mathbb{E}_{\mathbb{P}^w} [h(P^v)] - \frac{\delta}{2(1-(1-r)^2)} \mathbb{E}_{\mathbb{P}^w} \left[\int_0^1 \left(r \frac{v_s^2 - \sigma^2}{2\sigma} \right)^2 ds \right] \\ & = \mathbb{E}_{\mathbb{P}^w} \left[h(P^v) - \frac{\delta r}{8(2-r)\sigma^2} \int_0^1 (v_s^2 - \sigma^2)^2 ds \right]. \end{aligned}$$

This yields the desired lower bound (3.2).

3.2. Proof of the upper bound

We will prove the upper bound first for the case

$$x_0 = \zeta_0 = 0$$

and reduce the general case to this one in the end.

For the upper bound

$$\limsup_N \pi^N(h(P^N)) \leq \sup_{v \in \mathcal{D}} \mathbb{E}_{\mathbb{P}^w} \left[h(P^v) - \frac{r\delta}{8\sigma^2(2-r)} \int_0^1 |v_t^2 - \sigma^2|^2 dt \right] \tag{3.3}$$

we first note that super-replication prices with transient impact are dominated by super-replication prices in a suitable model with purely temporary impact as in [8]:

Lemma 3.4. *For any $N = 1, 2, \dots$, we have*

$$\pi^N(h(P^N)) \leq \widehat{\pi}^N(h(P^N)),$$

where $\widehat{\pi}^N(h(P^N))$ is the super-replication price in the model with full resilience $\widehat{r} \triangleq 1$ and market depth $\widehat{\delta} \triangleq r\delta/(2-r)$.

Proof. Consider the cost term κ_n^X from (2.4) and observe that with $\zeta_0 = 0$,

$$\begin{aligned} & (1-r) \sum_{m=1}^n \zeta_{m-1}^X |X_m - X_{m-1}| \\ &= \frac{1-r}{\delta} \sum_{m=1}^n \sum_{1 \leq l < m} (1-r)^{m-1-l} |X_l - X_{l-1}| |X_m - X_{m-1}| \\ &\leq \frac{1-r}{\delta} \sum_{m=1}^n \sum_{1 \leq l < m} (1-r)^{m-1-l} \frac{1}{2} (|X_l - X_{l-1}|^2 + |X_m - X_{m-1}|^2) \\ &= \frac{1-r}{\delta} \sum_{m=1}^n \frac{1}{2} \left(\sum_{l=1}^{m-1} (1-r)^{m-1-l} + \sum_{l=m}^{n-1} (1-r)^{l-m} \right) |X_m - X_{m-1}|^2 \\ &\leq \frac{1-r}{\delta} \sum_{m=1}^n \left(\sum_{k=0}^{\infty} (1-r)^k \right) |X_m - X_{m-1}|^2 \\ &= \frac{1-r}{\delta r} \sum_{m=1}^n |X_m - X_{m-1}|^2. \end{aligned}$$

As a result the cost term κ_n^X in the original model can be estimated by

$$\kappa_n^X \leq \left(\frac{1-r}{\delta r} + \frac{1}{2\delta} \right) \sum_{m=1}^n |X_m - X_{m-1}|^2 = \widehat{\kappa}_n^X,$$

where $\widehat{\kappa}^X$ is the cost term for the fully resilient model with $\widehat{r} = 1$ and depth $\widehat{\delta} = r\delta/(2-r)$. The costs in this auxiliary model being higher, the super-replication of any claim cannot be less expensive than in the original model and we obtain our assertion. \square

For (3.3), it thus suffices to prove

$$\limsup_N \widehat{\pi}^N(h(P^N)) \leq \sup_{\nu \in \mathcal{D}} \mathbb{E}_{\mathbb{P}^\nu} \left[h(P^\nu) - \frac{\widehat{\delta}}{8\sigma^2} \int_0^1 |v_t^2 - \sigma^2|^2 dt \right].$$

For this asymptotic analysis, we will work with a family of space-time discretizations of our price process. Specifically, we let, for any $\varepsilon > 0$, the sequence of partitions $\tau^{N,\varepsilon} = (\tau_k^{N,\varepsilon})_{k=0,1,\dots,N}$, $N = 1, 2, \dots$, be given by $\tau_0^{N,\varepsilon} \triangleq 0$ and

$$\tau_k^{N,\varepsilon} \triangleq \inf \{ t \geq \tau_{k-1}^{N,\varepsilon} : |P_t^N - P_{\tau_{k-1}^{N,\varepsilon}}^N| \geq \varepsilon \text{ or } |t - \tau_{k-1}^{N,\varepsilon}| \geq \varepsilon^2 \} \wedge (1 - N^{-2/3})$$

for $k = 1, 2, \dots$. With $\tau^{N,\varepsilon}$ we associate the following discretization of P^N :

$$P_t^{N,\varepsilon} \triangleq \sum_{k=1,2,\dots} P_{\tau_{k-1}^{N,\varepsilon}}^{N,\varepsilon} 1_{[\tau_{k-1}^{N,\varepsilon}, \tau_k^{N,\varepsilon})}(t) + P_{1-N-2/3}^N 1_{[1-N-2/3, 1]}(t), \quad 0 \leq t \leq 1.$$

Our next lemma reveals that, under our regularity assumptions on h , the super-replication price of $h(P^N)$ in the N -step model is controlled by the super-replication price of a particular quadratic claim on $P^{N,\varepsilon}$ and a knock-out variant of the claim h applied to $P^{N,\varepsilon}$ that only generates a payoff if this underlying does not fluctuate “too much”:

Lemma 3.5. *Let $c = c(\lambda) > 0$ be such that $h(p) \leq \lambda^2(\|p - p_0\|_\infty^2 + c)$, $p \in D[0, 1]$ (observe that $c(\lambda)$ exists since Assumption 3.1 implies that $h(p)$ has at most linear growth in $\|p - p_0\|_\infty$). Then, for any $\varepsilon > 0$ and $\lambda \in (0, 1)$, the constant $K = K(\varepsilon, \lambda) \triangleq \lceil c/(\varepsilon\lambda)^2 \rceil + 1$ is large enough to ensure that for sufficiently large N we have*

$$\widehat{\pi}^N(h(P^N)) \leq 3L\varepsilon + (1 - \lambda)\widehat{\pi}^N(H^{N,\varepsilon,K}/(1 - \lambda)) + \lambda\widehat{\pi}^N(\lambda Q^{N,\varepsilon}), \tag{3.4}$$

where L is the Lipschitz constant from Assumption 3.1 and where

$$H^{N,\varepsilon,K} \triangleq h(P^{N,\varepsilon}) 1_{\{\tau_K^{N,\varepsilon} = 1 - N^{-2/3}\}},$$

$$Q^{N,\varepsilon} \triangleq \sup_{0 \leq t \leq 1} |P_t^{N,\varepsilon} - P_0|^2 + \sum_{k=1,2,\dots} (|P_{\tau_k^{N,\varepsilon}}^{N,\varepsilon} - P_{\tau_{k-1}^{N,\varepsilon}}^{N,\varepsilon}|^2 + |\tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}|). \tag{3.5}$$

Proof. From the definitions of L and $P^{N,\varepsilon}$ and the regularity of h , it follows that, for sufficiently large N ,

$$h(P^N) \leq 3L\varepsilon + h(P^{N,\varepsilon}). \tag{3.6}$$

For $K = K(\varepsilon, \lambda)$ as defined above, we have furthermore

$$h(p) \leq \lambda^2 \left(\sup_{0 \leq t \leq 1} |p(t) - p_0|^2 + K\varepsilon^2 \right), \quad p \in D[0, 1]. \tag{3.7}$$

From the definition of $\tau_k^{N,\varepsilon}$, $k = 0, 1, \dots$, we get in addition that

$$K\varepsilon^2 \leq \sum_{k=1}^K (|P_{\tau_k^{N,\varepsilon}}^{N,\varepsilon} - P_{\tau_{k-1}^{N,\varepsilon}}^{N,\varepsilon}|^2 + |\tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}|) \quad \text{on } \{\tau_K^{N,\varepsilon} < 1 - N^{-2/3}\}. \tag{3.8}$$

Combining (3.7) and (3.8) gives

$$h(P^{N,\varepsilon}) \leq H^{N,\varepsilon,K} + \lambda^2 Q^{N,\varepsilon}. \tag{3.9}$$

Convexity of the wealth dynamics (2.3) implies convexity of the super-replication cost functional, and so (3.9) yields

$$\widehat{\pi}^N(h(P^{N,\varepsilon})) \leq (1 - \lambda)\widehat{\pi}^N(H^{N,\varepsilon,K}/(1 - \lambda)) + \lambda\widehat{\pi}^N(\lambda Q^{N,\varepsilon}).$$

Together with (3.6), this implies (3.4). □

Our next lemma shows that the super-replication price of $\lambda Q^{N,\varepsilon}$ is easy to control (at least for small $\lambda \in (0, 1)$) and so its contribution to (3.4) vanishes as $\lambda \downarrow 0$:

Lemma 3.6. *There exists $\lambda_0 > 0$ such that for any $\varepsilon > 0$, $\lambda \in [0, \lambda_0]$ and $N = 1, 2, \dots$ we have*

$$\widehat{\pi}^N(\lambda Q^{N,\varepsilon}) \leq \lambda(1 + 36\sigma^2).$$

Proof. Let $a, b, d, e > 0$ and consider a portfolio strategy with initial capital $\xi_0 = a$ and a (predictable) trading strategy of the form which for n from $[N\tau_{k-1}^{N,\varepsilon}] + 1$ to $[N\tau_k^{N,\varepsilon}]$, $k = 1, 2, \dots$, is given by

$$\begin{aligned} X_n = & -b \max_{i=0,\dots,k-1} (P_{\tau_i^{N,\varepsilon}}^N - p_0) + b \max_{0 \leq i \leq k-1} (p_0 - P_{\tau_i^{N,\varepsilon}}^N) \\ & - d(P_{\tau_{k-1}^{N,\varepsilon}}^N - p_0) + e(P_{(n-1)/N}^N - p_0), \end{aligned}$$

and which is 0 for n from $[N(1 - N^{-2/3})] + 1$ to N . In order to estimate the corresponding portfolio value at the maturity date, we apply Proposition 2.1 in [1] for $p = 2$. We notice that for $p = 2$ this proposition holds true for any sequence of real numbers, including negative numbers. We apply this pathwise Doob's inequality for

$$s_k \triangleq \pm (P_{\tau_k^{N,\varepsilon}}^N - p_0), \quad k = 1, 2, \dots$$

Moreover, we will use the elementary identity

$$\sum_{k=1}^j y_k (y_{k+1} - y_k) = \frac{1}{2} \left(y_{j+1}^2 - y_1^2 - \sum_{k=1}^j (y_{k+1} - y_k)^2 \right)$$

with $y_k = P_{\tau_{k-1}^{N,\varepsilon}}^N - p_0$ and also with $y_k = P_{(k-1)/N}^N - p_0$. By the well-known inequalities

$$(z_1 + z_2)^2 \leq 2(z_1^2 + z_2^2), \quad (z_1 + z_2 + z_3 + z_4)^2 \leq 4(z_1^2 + z_2^2 + z_3^2 + z_4^2),$$

the result then is

$$\begin{aligned} \xi_N^X = & a + \sum_{n=1}^N X_n (P_{n/N}^N - P_{(n-1)/N}^N) - \frac{1}{2\delta} \sum_{n=1}^N |X_n - X_{n-1}|^2 \\ \geq & a + \frac{b}{4} \left(\max_{0 \leq t \leq 1} P_t^{N,\varepsilon} - p_0 \right)^2 - b |P_1^{N,\varepsilon} - p_0|^2 \\ & + \frac{b}{4} \left(p_0 - \min_{0 \leq t \leq 1} P_t^{N,\varepsilon} \right)^2 - b |P_1^{N,\varepsilon} - p_0|^2 \\ & + \frac{d}{2} \sum_{k=1,2,\dots} |P_{\tau_k^{N,\varepsilon}}^N - P_{\tau_{k-1}^{N,\varepsilon}}^N|^2 - \frac{d}{2} |P_1^{N,\varepsilon} - p_0|^2 \end{aligned}$$

$$\begin{aligned}
 & + \frac{e}{2} |P_1^{N,\varepsilon} - P_0|^2 - \frac{e}{2} \sigma^2 \\
 & - \frac{2}{\delta} \left(e^2 \sigma^2 + (2b^2 + d^2) \sum_{k=1,2,\dots} |P_{\tau_k^{N,\varepsilon}}^N - P_{\tau_{k-1}^{N,\varepsilon}}^N|^2 \right) \\
 & - \frac{1}{2\delta} (2b + d + e)^2 \left(\frac{\sigma}{\sqrt{N}} + \max_{0 \leq t \leq 1} |P_t^{N,\varepsilon} - p_0| \right)^2.
 \end{aligned}$$

Here, the last two lines give an estimate for the transaction costs (including the liquidation costs) which correspond to our trading strategy.

It follows that

$$\begin{aligned}
 \xi_N^X & \geq a - \sigma^2 \left(\frac{e}{2} + \frac{2e^2}{\delta} + \frac{(2b + d + e)^2}{\delta} \right) \\
 & + \left(\frac{b}{4} - \frac{(2b + d + e)^2}{\delta} \right) \sup_{0 \leq t \leq 1} |P_t^{N,\varepsilon} - p_0|^2 \\
 & + \frac{e - 4b - d}{2} |P_1^{N,\varepsilon} - p_0|^2 \\
 & + \left(\frac{d}{2} - \frac{4b^2 + 2d^2}{\delta} \right) \sum_{k=1,2,\dots} |P_{\tau_k^{N,\varepsilon}}^N - P_{\tau_{k-1}^{N,\varepsilon}}^N|^2.
 \end{aligned}$$

Let $b = 8\lambda$, $d = 4\lambda$, $e = 4b + d = 36\lambda$ and $a = \lambda + e\sigma^2 = \lambda(1 + 36\sigma^2)$. Then for sufficiently small λ we get

$$\xi_N^X \geq \lambda + \lambda \sup_{0 \leq t \leq 1} |P_t^{N,\varepsilon} - P_0|^2 + \lambda \sum_{k=1,2,\dots} |P_{\tau_k^{N,\varepsilon}}^N - P_{\tau_{k-1}^{N,\varepsilon}}^N|^2 \geq \lambda Q^{N,\varepsilon}$$

and the result follows. □

The proof of the upper bound thus relies on an understanding how to super-replicate the claims $H^{N,\varepsilon,K}/(1 - \lambda)$. Notice that these claims depend on the values of their underlying at only a fixed number K of sampling times. Such claims turn out to allow for a particularly convenient duality estimate for their super-replication prices:

Lemma 3.7. *Let G be a claim of the form $G = g((\tau_k^{N,\varepsilon}, P_{\tau_k^{N,\varepsilon}}^N)_{k=0,\dots,K})$ for some measurable, bounded, nonnegative function $g = g((t_k, p_k)_{k=0,\dots,K})$. Then, for any $\varepsilon, \eta > 0$, we can find for sufficiently large N a probability \mathbb{Q}^N on (Ω, \mathcal{F}_N) (also depending on ε, η and g) such that for*

the filtration $(\mathcal{F}_k^{N,\varepsilon})_{k=0,\dots,K}$ generated by $(\tau_k^{N,\varepsilon}, P_{\tau_k^{N,\varepsilon}}^{N,\varepsilon})_{k=0,\dots,K}$ we have

$$\begin{aligned} \widehat{\pi}^N(G) &\leq \frac{1}{4}\eta\sigma^2\widehat{\delta} + \mathbb{E}_{\mathbb{Q}^N}[G] \\ &\quad - \frac{\widehat{\delta}}{8\sigma^2}\mathbb{E}_{\mathbb{Q}^N}\left[\sum_{k=1}^K\left(\frac{\mathbb{E}_{\mathbb{Q}^N}[(P_{\tau_k^{N,\varepsilon}}^N)^2 - (P_{\tau_{k-1}^{N,\varepsilon}}^N)^2 \mid \mathcal{F}_{k-1}^{N,\varepsilon}]}{\mathbb{E}_{\mathbb{Q}^N}[\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon} \mid \mathcal{F}_{k-1}^{N,\varepsilon}]} - \sigma^2\right)^2 (\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon})\right]. \end{aligned} \quad (3.10)$$

In addition, under \mathbb{Q}^N , $(P_{\tau_k^{N,\varepsilon}}^N)_{k=0,\dots,K}$ is close to being a martingale in the sense that

$$\mathbb{E}_{\mathbb{Q}^N}\left[\sum_{k=1}^K\left|\mathbb{E}_{\mathbb{Q}^N}[P_{\tau_k^{N,\varepsilon}}^N - P_{\tau_{k-1}^{N,\varepsilon}}^N \mid \mathcal{F}_{k-1}^{N,\varepsilon}]\right|\right] \leq (\|G\|_\infty + \eta)/\log N. \quad (3.11)$$

Proof. Fix $\varepsilon > 0$ and $N \in \{1, 2, \dots\}$. Rather than looking among all strategies in \mathcal{X} for a cost-effective super-hedge, we will consider a suitably constrained class. To this end, denote by \mathcal{A} the class of pairs $(\phi_k, \psi_k)_{k=1,\dots,K}$ of $(\mathcal{F}_k^{N,\varepsilon})_{k=0,\dots,K}$ -predictable processes such that $|\phi_k|, |\psi_k| \leq \log N$ for $k = 1, \dots, K$. Each such pair induces a strategy $(X_n^{(\phi, \psi)})_{n=1,\dots,N} \in \mathcal{X}$ in the N -step model that we can define piecewise on $[[N\tau_k^{N,\varepsilon}], [N\tau_{k+1}^{N,\varepsilon}]]$ for $k = 1, \dots, K$ as follows: If $\tau_k^{N,\varepsilon} < 1 - N^{-1/2}$, the duration $\tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}$ of the k th period is at least of order $N^{1/2}$ and we thus can subdivide the interval $[[N\tau_{k-1}^{N,\varepsilon}], [N\tau_k^{N,\varepsilon}]]$ into two parts. On the first (short) subinterval of length $N^{1/3}$ we trade at constant speed into a position holding $\phi_k + \psi_k P_{\tau_{k-1}^{N,\varepsilon}}^{N,\varepsilon}$ risky assets; in the periods n afterwards, we hold the position $\phi_k + \psi_k P_n^N$ until the stopping time $[N\tau_k^{N,\varepsilon}] + 1$ when the next iteration of this recipe proceeds with $k + 1$ instead of k while $k < K$. If $\tau_{k-1}^{N,\varepsilon} \geq 1 - N^{-1/2}$ or when we have completed the K th such iteration, we complete the construction of the strategy by liquidating the obtained position in $N^{1/3}$ steps and staying flat until the end.

Let us analyze the profits and losses and also the costs accruing from this strategy. For this, note that, due to the random walk dynamics (3.1), we have

$$\sum_{l < m \leq n} P_{\frac{m-1}{N}}^N (P_{\frac{m}{N}}^N - P_{\frac{m-1}{N}}^N) = \frac{1}{2} \left((P_{\frac{n}{N}}^N)^2 - (P_{\frac{l}{N}}^N)^2 - \sigma^2 \frac{n-l}{N} \right).$$

Moreover, note that P^N is uniformly bounded on $[0, \tau_K^{N,\varepsilon}]$ by $|p_0| + K\varepsilon + \sigma$, so that, in particular, $X^{(\phi, \psi)}$ is of size $O(\log N)$. Therefore, the profit and loss due to fluctuations in the fundamental value incurred by the above strategy is up to a term of order $O(\log N/N^{1/6})$ (accounting for the transition period of length $N^{1/3}$ when a position change of at most order $\log N$ is accomplished while the underlying moves in steps of order $1/\sqrt{N}$) given by

$$\sum_{k=1}^K \phi_k (P_{\tau_k^{N,\varepsilon}}^N - P_{\tau_{k-1}^{N,\varepsilon}}^N) + \sum_{k=1}^K \psi_k \frac{1}{2} \left((P_{\tau_k^{N,\varepsilon}}^N)^2 - (P_{\tau_{k-1}^{N,\varepsilon}}^N)^2 - \sigma^2 (\tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}) \right). \quad (3.12)$$

At the same time, the logarithmic bounds on the allowed positions ensure that the costs are

$$\widehat{\kappa}_N^{X^{(\phi, \psi)}} = \frac{1}{2\delta} \sum_{k=1}^K \sigma^2 \psi_k^2 (\tau_k^{N, \varepsilon} - \tau_{k-1}^{N, \varepsilon}) + O(\log^2 N/N^{1/3}), \tag{3.13}$$

where the O -term accounts for the $O(N^{1/3}(\log N/N^{1/3})^2) = O(\log^2 N/N^{1/3})$ costs for the gradual position build-up of the first $N^{1/3}$ steps of each period $k = 1, \dots, K$ and for the $O(\log^2 N/N^{1/3})$ costs resulting from the one possible jump at the end of this initial build-up which is at most of size $O(N^{1/3} \log N/\sqrt{N}) = O(\log N/N^{1/6})$; the running costs in the second leg of each trading period $k = 1, \dots, K$ are reflected by the sum in (3.13).

It follows that, for large enough N , we will have

$$\widehat{\pi}^N(G) \leq o(1) + \widetilde{\pi}^N(G), \tag{3.14}$$

where $\widetilde{\pi}^N(G)$ denotes the super-replication price of the claim G when restricting to strategies $X^{(\phi, \psi)}$ as above with profits and losses given by (3.12) and trading costs given by the sum in (3.13).

Observing that (3.12) is linear in ϕ and recalling the constraint $|\phi| \leq \log N$, we get from classical linear super-replication duality with convexly constrained strategy sets (cf. [11], Theorem 4.1 in connection with Example 2.3) that for any fixed ψ -component we have

$$\widetilde{\pi}^N(G) \leq \sup_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}} \left[G^\psi - \log N \sum_{k=1}^K |\mathbb{E}_{\mathbb{Q}} [P_{\tau_k^{N, \varepsilon}}^N - P_{\tau_{k-1}^{N, \varepsilon}}^N | \mathcal{F}_{k-1}^{N, \varepsilon}]| \right], \tag{3.15}$$

where the supremum is taken over the set of all measures \mathbb{Q} on (Ω, \mathcal{F}_N) and where

$$\begin{aligned} G^\psi \triangleq & G - \sum_{k=1}^K \psi_k \frac{1}{2} ((P_{\tau_k^{N, \varepsilon}}^N)^2 - (P_{\tau_{k-1}^{N, \varepsilon}}^N)^2 - \sigma^2 (\tau_k^{N, \varepsilon} - \tau_{k-1}^{N, \varepsilon})) \\ & + \frac{1}{2\delta} \sum_{k=1}^K \sigma^2 \psi_k^2 (\tau_k^{N, \varepsilon} - \tau_{k-1}^{N, \varepsilon}) \end{aligned}$$

denotes the claim that remains to be super-hedged by suitably choosing ϕ when the ψ -component is fixed.

For $\mathcal{E}(\mathbb{Q}, \psi)$ denoting the unconditional expectation in (3.15), it is readily checked that $\psi \mapsto \mathcal{E}(\mathbb{Q}, \psi)$ is convex for any fixed \mathbb{Q} and that $\mathbb{Q} \mapsto \mathcal{E}(\mathbb{Q}, \psi)$ is concave for any ψ fixed. Observing that the domains of \mathbb{Q} and ψ can easily be identified with convex and compact subsets in Euclidean space, we can thus invoke the Minimax Theorem (see, e.g., Theorem 45.8 in [19]) to obtain

$$\widetilde{\pi}^N(G) \leq \inf_{\psi} \sup_{\mathbb{Q}} \mathcal{E}(\mathbb{Q}, \psi) = \sup_{\mathbb{Q}} \inf_{\psi} \mathcal{E}(\mathbb{Q}, \psi).$$

In conjunction with (3.14), we therefore can find a \mathbb{Q}^N on (Ω, \mathcal{F}_N) such that

$$\widehat{\pi}^N(G) \leq o(1) + \inf_{\psi} \mathcal{E}(\mathbb{Q}^N, \psi). \quad (3.16)$$

In order to control the latter infimum, observe that the terms in $\mathcal{E}(\mathbb{Q}^N, \psi)$ involving ψ_k contribute

$$\begin{aligned} & \mathbb{E}_{\mathbb{Q}^N} \left[\frac{\sigma^2}{2\widehat{\delta}} \psi_k^2 (\tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}) - \psi_k \frac{1}{2} \left((P_{\tau_k^{N,\varepsilon}}^N)^2 - (P_{\tau_{k-1}^{N,\varepsilon}}^N)^2 - \sigma^2 (\tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}) \right) \right] \\ & \leq \mathbb{E}_{\mathbb{Q}^N} \left[\frac{\sigma^2}{2\widehat{\delta}} \psi_k^2 \mathbb{E}_{\mathbb{Q}^N} [\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon} \mid \mathcal{F}_{k-1}^{N,\varepsilon}] \right. \\ & \quad \left. - \frac{1}{2} \psi_k \mathbb{E}_{\mathbb{Q}^N} \left[(P_{\tau_k^{N,\varepsilon}}^N)^2 - (P_{\tau_{k-1}^{N,\varepsilon}}^N)^2 - \sigma^2 (\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}) \mid \mathcal{F}_{k-1}^{N,\varepsilon} \right] \right] \\ & \quad + \sup_{\psi} \left\{ -\frac{\sigma^2}{2\widehat{\delta}} \Psi^2 \eta + \frac{1}{2} \Psi \sigma^2 \eta \right\}, \end{aligned} \quad (3.17)$$

where $\eta > 0$ is arbitrary and where we used that ψ_k is $\mathcal{F}_{k-1}^{N,\varepsilon}$ -measurable. The minimum over such ψ_k in the last expectation is attained for

$$\psi_k^* = \frac{\widehat{\delta} \mathbb{E}_{\mathbb{Q}^N} \left[(P_{\tau_k^{N,\varepsilon}}^N)^2 - (P_{\tau_{k-1}^{N,\varepsilon}}^N)^2 - \sigma^2 (\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}) \mid \mathcal{F}_{k-1}^{N,\varepsilon} \right]}{2\sigma^2 \mathbb{E}_{\mathbb{Q}^N} [\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon} \mid \mathcal{F}_{k-1}^{N,\varepsilon}]},$$

which is uniformly bounded in N for $\eta > 0$ due to the uniform bound on P^N up to time τ_K^N . In particular, $|\psi_k^*| \leq \log N$ for sufficiently large N . The corresponding minimum is

$$\begin{aligned} & -\frac{\widehat{\delta}}{8\sigma^2} \mathbb{E}_{\mathbb{Q}^N} \left[\frac{\mathbb{E}_{\mathbb{Q}^N} \left[(P_{\tau_k^{N,\varepsilon}}^N)^2 - (P_{\tau_{k-1}^{N,\varepsilon}}^N)^2 - \sigma^2 (\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}) \mid \mathcal{F}_{k-1}^{N,\varepsilon} \right]^2}{\mathbb{E}_{\mathbb{Q}^N} [\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon} \mid \mathcal{F}_{k-1}^{N,\varepsilon}]} \right] \\ & = -\frac{\widehat{\delta}}{8\sigma^2} \mathbb{E}_{\mathbb{Q}^N} \left[\left(\frac{\mathbb{E}_{\mathbb{Q}^N} \left[(P_{\tau_k^{N,\varepsilon}}^N)^2 - (P_{\tau_{k-1}^{N,\varepsilon}}^N)^2 \mid \mathcal{F}_{k-1}^{N,\varepsilon} \right]}{\mathbb{E}_{\mathbb{Q}^N} [\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon} \mid \mathcal{F}_{k-1}^{N,\varepsilon}]} - \sigma^2 \right)^2 (\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}) \right]. \end{aligned}$$

Now, we just need to combine these contribution to $\mathcal{E}(\mathbb{Q}^N, \psi^*)$ with estimate (3.16) and the fact that the supremum in (3.17) is $\eta\sigma^2\widehat{\delta}/8$ to derive the claimed estimate (3.10).

For the remaining estimate (3.11), consider $\psi \equiv 0$ in the estimate (3.16) for $\widehat{\pi}^N(G)$. Since by absence of arbitrage at the same time $\widehat{\pi}^N(G) \geq 0$, we can conclude, at least for large enough N ,

$$0 \leq \eta + \mathbb{E}_{\mathbb{Q}^N} \left[G - \log N \sum_{k=1}^K \left| \mathbb{E}_{\mathbb{Q}^N} \left[P_{\tau_k^{N,\varepsilon}}^N - P_{\tau_{k-1}^{N,\varepsilon}}^N \mid \mathcal{F}_{k-1}^{N,\varepsilon} \right] \right| \right],$$

which gives (3.11). \square

The claims $H^{N,\varepsilon,K}/(1-\lambda)$ of (3.5) are of the form required for the previous lemma. This yields an upper bound for super-replication prices which, however, still depends on N and only involves a process which is almost a martingale along the times of its jumps. To get a more convenient upper bound, it will be useful to consider processes on a slightly expanded time horizon, namely on $[0, 1 + \lambda]$ rather than $[0, 1]$. The payoff function h can be rescaled to $h_{1+\lambda} : D[0, 1 + \lambda] \rightarrow \mathbb{R}$ simply by letting, for $p \in D[0, 1 + \lambda]$,

$$h_{1+\lambda}(p) \triangleq h([0, 1] \ni t \mapsto p(t(1 + \lambda))).$$

Now, consider, for $\varepsilon, \lambda > 0$ fixed and $K = K(\varepsilon, \lambda)$ as in Lemma 3.5, the class $\mathcal{D}^{\varepsilon,\lambda}$ of measurable processes D on some probability space $(\Omega^D, \mathcal{F}^D, \mathbb{P}^D)$ of the form

$$D_t = \sum_{k=1}^K D_{\theta_{k-1}} 1_{[\theta_{k-1}, \theta_k)}(t) + (D_{\theta_K} + \sigma W_{t-\theta_K}) 1_{[\theta_K, 1+\lambda]}(t) \tag{3.18}$$

such that, for $k = 1, \dots, K$, we have

$$D_0 = p_0, \quad |D_{\theta_k} - D_{\theta_{k-1}}| \leq 2\varepsilon, \tag{3.19}$$

$$\theta_0 = 0, \quad \frac{\lambda}{K} \leq \theta_k - \theta_{k-1} \leq \frac{\lambda}{K} + \varepsilon^2, \tag{3.20}$$

and

$$D_{\theta_{k-1}} = \mathbb{E}^D[D_{\theta_k} | \mathcal{F}_{\theta_{k-1}}^D], \tag{3.21}$$

where $(\mathcal{F}_t^D)_{0 \leq t \leq 1+\lambda}$ denotes the filtration generated by D and where W is a Brownian motion independent of $\mathcal{F}_{\theta_K}^D$ under \mathbb{P}^D . It will also be convenient to associate with each such D the process

$$\zeta_t^D \triangleq \sum_{k=1}^K \frac{\mathbb{E}_{\mathbb{P}^D}[(D_{\theta_k})^2 - (D_{\theta_{k-1}})^2 | \mathcal{F}_{\theta_{k-1}}^D]}{\mathbb{E}_{\mathbb{P}^D}[\theta_k - \theta_{k-1} | \mathcal{F}_{\theta_{k-1}}^D]} 1_{[\theta_{k-1}, \theta_k)}(t) + \sigma^2 1_{[\theta_K, 1+\lambda]}(t),$$

which, for later use, we observe is bounded for any fixed $\lambda > 0, \varepsilon < 1/\lambda$. Indeed, combining (3.21), (3.19) and (3.20) with $K = \lceil (c/\varepsilon\lambda)^2 \rceil + 1$ implies

$$|\zeta_t^D| \leq \frac{4\varepsilon^2}{\lambda/K} \vee \sigma^2 \leq \frac{4(c+1)}{\lambda^3} \vee \sigma^2. \tag{3.22}$$

With this notation, we get the following duality estimate:

Lemma 3.8. *For any $\varepsilon, \lambda > 0$ and with $K = K(\varepsilon, \lambda)$ as in Lemma 3.5, we have*

$$\begin{aligned} & \widehat{\pi}^N(H^{N,\varepsilon,K}/(1-\lambda)) \\ & \leq \left(2 + \frac{1}{4}\sigma^2\delta\right) \frac{\lambda}{K} + \sup_{D \in \mathcal{D}^{\varepsilon,\lambda}} \mathbb{E}_{\mathbb{P}^D} \left[\frac{h_{1+\lambda}(D)}{1-\lambda} - \frac{\widehat{\delta}}{8\sigma^2} \int_0^{1+\lambda} (\zeta_t^D - \sigma^2)^2 dt \right] \end{aligned} \tag{3.23}$$

for sufficiently large N .

Proof. Fix $\varepsilon, \lambda > 0$ and let $K \triangleq K(\varepsilon, \lambda)$ and $\eta \triangleq \lambda/K$. We will use the notation from Lemma 3.7 and assume henceforth that N is large enough for this lemma's assertion to hold true. With $G \triangleq H^{N,\varepsilon,K}$ and $\tau_k \triangleq \tau_k^{N,\varepsilon}, k = 0, \dots, K$, we furthermore define, again for $k = 0, \dots, K$,

$$\begin{aligned} \theta_k^N &\triangleq \tau_k^{N,\varepsilon} + k\eta, \\ D_{\theta_k^N}^N &\triangleq p_0 + \sum_{j=1}^k (P_{\tau_j^{N,\varepsilon}}^N - \mathbb{E}_{\mathbb{Q}^N} [P_{\tau_j^{N,\varepsilon}}^N | \mathcal{F}_{\theta_{j-1}^N}^{D^N}]), \end{aligned}$$

to specify via (3.18) a process $D = D^N$ along with a probability $\mathbb{P}^D \triangleq \mathbb{Q}^N$ that is contained in $\mathcal{D}^{\varepsilon,\lambda}$. Indeed, the martingale-like property (3.21) is immediate as are the constraints on the intervention times (3.20). The increment restriction (3.19) holds since $P_{\tau_k}^{N,\varepsilon}$ is within ε of the $\mathcal{F}_{\theta_{k-1}^N}^{D^N}$ -measurable quantity $P_{\tau_{k-1}}^{N,\varepsilon}$ due to the definition $\tau_k^{N,\varepsilon}$. Moreover, it is easy to check that

$$\max_{k=0,\dots,K} |P_{\tau_k}^{N,\varepsilon} - D_{\theta_k^N}^N| \leq \sum_{k=1}^K |\mathbb{E}_{\mathbb{Q}^N} [P_{\tau_k}^N - P_{\tau_{k-1}}^N | \mathcal{F}_{k-1}^{N,\varepsilon}]| \tag{3.24}$$

and

$$\max_{k=0,\dots,K} \left| \frac{\tau_k^{N,\varepsilon} + \lambda k/K}{1 + \lambda} - \tau_k^{N,\varepsilon} \right| \leq \lambda.$$

Therefore, on the event $\{\tau_K^{N,\varepsilon} = 1\}$ we can estimate the Skorohod-distance

$$d(P^{N,\varepsilon}, D_{(1+\lambda)\cdot}^N) \leq \lambda + \sum_{k=1}^K |\mathbb{E}_{\mathbb{Q}^N} [P_{\tau_k}^N - P_{\tau_{k-1}}^N | \mathcal{F}_{k-1}^{N,\varepsilon}]|$$

so that by Assumption 3.1 on h and (3.11) we get

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}^N} [H^{N,\varepsilon,K}] &\leq \mathbb{E}_{\mathbb{Q}^N} [h_{1+\lambda}(D^N)] \\ &\quad + L \mathbb{E}_{\mathbb{Q}^N} \left[\lambda + \sum_{k=1}^K |\mathbb{E}_{\mathbb{Q}^N} [P_{\tau_k}^N - P_{\tau_{k-1}}^N | \mathcal{F}_{k-1}^{N,\varepsilon}]| \right] \\ &\leq \mathbb{E}_{\mathbb{Q}^N} [h_{1+\lambda}(D^N)] + L(\lambda + \|H^{N,\varepsilon,K}\|_\infty + \eta) / \log N \\ &\leq \mathbb{E}_{\mathbb{Q}^N} [h_{1+\lambda}(D^N)] + O\left(\frac{1}{\log N}\right), \end{aligned} \tag{3.25}$$

where the latter estimate is due to the simple bound

$$H^{N,\varepsilon,K} \leq h(p_0) + Ld(P^{N,\varepsilon}, p_0) \leq h(p_0) + LK\varepsilon \quad \text{on } \{H^{N,\varepsilon,K} > 0\}.$$

Next, from (3.11), (3.24) and the fact that D^N, P^N are uniformly bounded in N we obtain

$$\begin{aligned} & \mathbb{E}_{\mathbb{Q}^N} \left(\int_0^{1+\lambda} (\zeta_t^{D^N} - \sigma^2)^2 dt \right) \\ &= \mathbb{E}_{\mathbb{Q}^N} \left(\sum_{k=1}^K \left(\frac{\mathbb{E}_{\mathbb{Q}^N} [(D_{\tau_k^{N,\varepsilon}}^N)^2 - (D_{\tau_{k-1}^{N,\varepsilon}}^N)^2 \mid \mathcal{F}_{k-1}^{N,\varepsilon}]}{\mathbb{E}_{\mathbb{Q}^N} [\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon} \mid \mathcal{F}_{k-1}^{N,\varepsilon}]} - \sigma^2 \right)^2 (\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}) \right) \\ &\leq \mathbb{E}_{\mathbb{Q}^N} \left(\sum_{k=1}^K \left(\frac{\mathbb{E}_{\mathbb{Q}^N} [(P_{\tau_k^{N,\varepsilon}}^N)^2 - (P_{\tau_{k-1}^{N,\varepsilon}}^N)^2 \mid \mathcal{F}_{k-1}^{N,\varepsilon}]}{\mathbb{E}_{\mathbb{Q}^N} [\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon} \mid \mathcal{F}_{k-1}^{N,\varepsilon}]} - \sigma^2 \right)^2 (\eta + \tau_k^{N,\varepsilon} - \tau_{k-1}^{N,\varepsilon}) \right) \\ &\quad + O\left(\frac{1}{\log N}\right). \end{aligned} \tag{3.26}$$

Using (3.25) and (3.26) in the estimate (3.10) provided by Lemma 3.7 it thus follows that $\hat{\pi}^N(H^{N,\varepsilon,K}/(1-\lambda))$ cannot be larger than the right-hand side of (3.23) for sufficiently large N . \square

Letting $\varepsilon \downarrow 0$ in the above expression will be made possible by the following tightness result:

Lemma 3.9. *For $\lambda > 0$ fixed, any sequence $D^m \in \mathcal{D}^{1/m,\lambda}$, $m = 1, 2, \dots$, contains a subsequence along which $\text{Law}(D^m, \int_0^{1+\lambda} \zeta_s^{D^m} ds \mid \mathbb{P}^{D^m})$ converges weakly on $D[0, 1 + \lambda]$ to $\text{Law}(M, \langle M \rangle \mid \mathbb{P}^M)$ for some continuous martingale $M = (M_t)_{0 \leq t \leq 1+\lambda}$ on a suitable probability space $(\Omega^M, \mathcal{F}^M, \mathbb{P}^M)$.*

Proof. Let $K^m \triangleq K(1/m, \lambda)$ as in Lemma 3.5 and denote by $(\theta_k^m)_{k=0, \dots, K^m}$ the times associated with D^m via (3.18)–(3.21); let furthermore $\mathbb{P}^m \triangleq \mathbb{P}^{D^m}$ denote the associated probability. For any $m = 1, 2, \dots$, we denote by $\{\hat{D}_t^m\}_{t=0}^{1+\lambda}$ the continuous linear interpolation of D^m ; observe that after time θ_K^m , this amounts to $\hat{D}_t^m \triangleq D_{\theta_K^m}^m + \sigma W_{t-\theta_K^m}$ where W is a Brownian motion independent of $\mathcal{F}_{\theta_K^m}^{D^m}$.

We will verify the Kolmogorov tightness criterion for these processes \hat{D}^m , $m = 1, 2, \dots$. So, take $m \in \{1, 2, \dots\}$ and fix $0 \leq t_1 < t_2 \leq 1 + \lambda$. Define the random times

$$\eta_i = K^m \wedge \min\{k \in \{0, \dots, K^m\} : \theta_k^m \geq t_i\}, \quad i = 1, 2.$$

The discrete-time process $\{D_{\theta_k^m}^m\}_{k=0, \dots, K^m}$ is a martingale with respect to the filtration generated by $(\theta_k^m, D_{\theta_k^m}^m)_{k=0, \dots, K^m}$ and η_1, η_2 are stopping times with respect to this filtration. Thus, from the Burkholder–Davis–Gundy inequality,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^m} [|D_{\eta_2}^m - D_{\eta_1}^m|^4] &\leq O(1) \mathbb{E}_{\mathbb{P}^m} \left[\left(\sum_{j=\eta_1+1, \dots, \eta_2} |D_{\theta_j^m}^m - D_{\theta_{j-1}^m}^m|^2 \right)^2 \right] \\ &\leq O(m^{-4}) \mathbb{E}_{\mathbb{P}^m} [|\eta_2 - \eta_1|^2], \end{aligned} \tag{3.27}$$

where the last inequality follows from (3.19) which ensures that the jumps of D^m are bounded by $2/m$.

Next, since the time between two subsequent jumps is at least $\lambda/K^m = O(1/m^2)$ and the jumps of D^m are bounded by $2/m$, we obtain that the size of the (random) slope for the linear interpolation process \hat{D}^m is at most of order $O(m)$. This together with the fact that the time between two subsequent jumps is less than or equal to $\lambda/K^m + 1/m^2$ yields

$$\begin{aligned} & |\hat{D}_{t_2}^m - \hat{D}_{t_1}^m| \\ & \leq |\hat{D}_{t_2 \wedge \theta_K^m}^m - \hat{D}_{t_1 \wedge \theta_K^m}^m| + \sigma |W_{t_2 \vee \theta_K^m} - W_{t_1 \vee \theta_K^m}| \\ & \leq 1_{\{t_2 > t_1 + 1/m^2\}} |D_{\theta_{\eta_2}^m}^m - D_{\theta_{\eta_1}^m}^m| \\ & \quad + 2O(m)((t_2 - t_1) \wedge (\lambda/K_m + 1/m^2)) + \sigma |W_{t_2 \vee \theta_K^m} - W_{t_1 \vee \theta_K^m}|. \end{aligned} \tag{3.28}$$

Using again that the time between two subsequent jumps is at least $\lambda/K_m = O(1/m^2)$, we obtain that if $t_2 > t_1 + 1/m^2$ then $\eta_2 - \eta_1 = O(m^2)(t_2 - t_1)$. Thus, from (3.27)–(3.28) and the elementary inequalities

$$\begin{aligned} (t_2 - t_1) \wedge (\lambda/K^m + 1/m^2) & \leq O(1/m)\sqrt{t_2 - t_1}, \\ (z_1 + z_2 + z_3)^4 & \leq 81(z_1^4 + z_2^4 + z_3^4), \end{aligned}$$

we obtain $\mathbb{E}_{\mathbb{P}^m}[|\hat{D}_{t_2}^m - \hat{D}_{t_1}^m|^4] = O((t_2 - t_1)^2)$ and tightness follows.

From Prokhorov's theorem we conclude that there exists a subsequence (still denoted by m) and a continuous process $M = (M_t)_{0 \leq t \leq 1+\lambda}$ that converges in law to some continuous process M on a suitable probability space $(\Omega^M, \mathcal{F}^M, \mathbb{P}^M)$. The obvious inequality $\sup_{0 \leq t \leq 1+\lambda} |\hat{D}_t^m - D_t^m| \leq 2/m$ yields the same convergence also for D^m .

Let us argue next that M is a martingale with respect to its own filtration. Fix m , let $(\mathcal{F}_t^m)_{0 \leq t \leq 1+\lambda}$ be the usual (right continuous and complete) filtration generated by D^m and consider the (RCLL) martingale

$$\tilde{D}_t^m = \mathbb{E}_{\mathbb{P}^m}[D_{1+\lambda}^m \mid \mathcal{F}_t^m], \quad 0 \leq t \leq 1 + \lambda.$$

Recall that the time between two subsequent jumps is bounded from below. Hence,

$$\tilde{D}_{\theta_k^m}^m = \mathbb{E}_{\mathbb{P}^m}[D_{\theta_{K^m}^m}^m \mid \mathcal{F}_{\theta_k^m}^m] = D_{\theta_k^m}^m, \quad k = 0, 1, \dots, K^m, \tag{3.29}$$

where the last equality follows from (3.21). From (3.29) and the estimate

$$\max_{k=1, \dots, K^m} |D_{\theta_k^m}^m - D_{\theta_{k-1}^m}^m| \leq 2/m$$

we get $\|\tilde{D}^m - D^m\|_\infty \leq 4/m$. Thus, the martingales \tilde{D}^m , $m = 1, 2, \dots$, are uniformly integrable and converge weakly to M . From Theorem 5.3 in [20] we conclude that M is a (continuous) martingale.

Now, we prove that $(D^m, \int_0^\cdot \zeta_s^{D^m} ds)$ converges in law to $(M, \langle M \rangle)$. For any $m = 1, 2, \dots$, let the quadratic variation of the martingale \tilde{D}^m be denoted by $([\tilde{D}^m]_t)_{0 \leq t \leq 1+\lambda}$. Theorem 5.5 in [20] then yields the convergence in law of $(\tilde{D}^m, [\tilde{D}^m])$ to $(M, \langle M \rangle)$. Thus, in order to complete the proof, it is sufficient to establish that

$$\lim_{m \rightarrow \infty} \mathbb{E}_{\mathbb{P}^m} \left[\sup_{0 \leq t \leq 1+\lambda} \left| \int_0^t \zeta_s^{D^m} ds - [\tilde{D}^m]_t \right| \right] = 0. \tag{3.30}$$

To that end, note that by the Burkholder–Davis–Gundy inequality

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}^m} \left[\max_{k=1, \dots, K^m} ([\tilde{D}^m]_{\theta_k^m} - [\tilde{D}^m]_{\theta_{k-1}^m})^2 \right] \\ & \leq \sum_{k=1}^{K^m} \mathbb{E}_{\mathbb{P}^m} \left[([\tilde{D}^m]_{\theta_k^m} - [\tilde{D}^m]_{\theta_{k-1}^m})^2 \right] \\ & \leq O(1) \sum_{k=1}^{K^m} \mathbb{E}_{\mathbb{P}^m} \left[\sup_{\theta_{k-1}^m \leq t \leq \theta_k^m} |\tilde{D}_t^m - \tilde{D}_{\theta_{k-1}^m}^m|^4 \right] \\ & \leq O(1) K^m O(1/m^4) = O(1/m^2). \end{aligned} \tag{3.31}$$

Next, observe that $([\tilde{D}^m]_{\theta_k^m} - \int_0^{\theta_k^m} \zeta_s^{D^m} ds)_{k=0, \dots, K^m}$ is a martingale. Hence, applying first the Doob–Kolmogorov inequality and Ito’s isometry and finally also (3.20), (3.22), we conclude (3.31)

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}^m} \left[\max_{k=0, \dots, K^m} \left([\tilde{D}^m]_{\theta_k^m} - \int_0^{\theta_k^m} \zeta_s^{D^m} ds \right)^2 \right] \\ & \leq 4 \mathbb{E}_{\mathbb{P}^m} \left[\sum_{k=1}^{K^m} \left([\tilde{D}^m]_{\theta_k^m} - [\tilde{D}^m]_{\theta_{k-1}^m} + \int_{\theta_{k-1}^m}^{\theta_k^m} |\zeta_s^{D^m}| ds \right)^2 \right] \\ & \leq 8 \mathbb{E}_{\mathbb{P}^m} \left[\sum_{k=1}^{K^m} ([\tilde{D}^m]_{\theta_k^m} - [\tilde{D}^m]_{\theta_{k-1}^m})^2 \right] \\ & \quad + 8 K^m \|\zeta^{D^m}\|_\infty^2 (\lambda/K^m + 1/m^2)^2 \\ & = O(1/m^2). \end{aligned} \tag{3.32}$$

Finally, by combining (3.20), (3.22) and applying the Jensen inequality for (3.31)–(3.32) we get

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}^m} \left[\sup_{0 \leq t \leq 1+\lambda} \left| \int_0^t \zeta_s^{D^m} ds - [\tilde{D}^m]_t \right| \right] \\ & = \mathbb{E}_{\mathbb{P}^m} \left[\sup_{0 \leq t \leq \theta_{K^m}^m} \left| \int_0^t \zeta_s^{D^m} ds - [\tilde{D}^m]_t \right| \right] \end{aligned}$$

$$\begin{aligned}
&\leq \|\zeta^{D^m}\|_\infty (\lambda/K^m + 1/m^2) \\
&\quad + \mathbb{E}^{\mathbb{P}^m} \left[\max_{k=1, \dots, K^m} |[\tilde{D}^m]_{\theta_k^m} - [\tilde{D}^m]_{\theta_{k-1}^m}| \right] \\
&\quad + \mathbb{E}^{\mathbb{P}^m} \left[\max_{k=0, \dots, K^m} \left| \int_0^{\theta_k^m} \zeta_s^{D^m} ds - [\tilde{D}^m]_{\theta_k^m} \right| \right] \\
&= O(1/m)
\end{aligned}$$

and (3.30) follows. \square

We will need the following stability result.

Lemma 3.10. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be an arbitrary filtered probability space and suppose h satisfies Assumption 3.1. Then the function*

$$F(z) = \sup_M \mathbb{E} \left[z_1 h(M) - z_2 \int_0^1 \left(\frac{d\langle M \rangle_t}{dt} - z_3 \right)^2 dt \right], \quad z = (z_1, z_2, z_3) \in (0, \infty)^3,$$

where the supremum is taken over all continuous martingales $M = (M_t)_{0 \leq t \leq 1}$ starting in $M_0 = p_0$ that have a quadratic variation $\langle M \rangle$ which is absolutely continuous with bounded density $\frac{d\langle M \rangle_t}{dt}$.

Proof. It is sufficient to prove the statement for the function

$$\begin{aligned}
\hat{F}(z) &\triangleq F(z) + z_2 z_3^2 \\
&= \sup_M \mathbb{E} \left[z_1 h(M) - z_2 \int_0^1 \left(\frac{d\langle M \rangle_t}{dt} \right)^2 dt + 2z_2 z_3 \langle M \rangle_1 \right]. \tag{3.33}
\end{aligned}$$

From the Doob–Kolmogorov inequality, the Jensen inequality and the estimate $h(p) \leq \|p - p_0\|_\infty^2 + c$ for $c = c(1)$ (as defined in Lemma 3.5) we obtain

$$\begin{aligned}
&\mathbb{E} \left[z_1 h(M) - z_2 \int_0^1 \left(\frac{d\langle M \rangle_t}{dt} \right)^2 dt + 2z_2 z_3 \langle M \rangle_1 \right] \\
&\leq z_1 c + (4z_1 + 2z_2 z_3) \mathbb{E} \langle M \rangle_1 - z_2 \mathbb{E} \left[\int_0^1 \left(\frac{d\langle M \rangle_t}{dt} \right)^2 dt \right] \\
&\leq z_1 c + (4z_1 + 2z_2 z_3) \sqrt{\mathbb{E} \left[\int_0^1 \left(\frac{d\langle M \rangle_t}{dt} \right)^2 dt \right]} - z_2 \mathbb{E} \left[\int_0^1 \left(\frac{d\langle M \rangle_t}{dt} \right)^2 dt \right]. \tag{3.34}
\end{aligned}$$

On the other hand by taking $M \equiv p_0$ we obtain $\hat{F} \geq 0$ (recall that h is nonnegative). Hence, on the right hand side of (3.33) we can restrict the supremum to the set of martingales for which the right hand side of (3.34) is nonnegative.

We conclude that for a bounded set $O \subset \mathbb{R}^3$ with $\inf_{z \in O} z_2 > 0$ there exists $\Theta = \Theta(O)$ such that for any $z \in O$ we have

$$\hat{F}(z) = \sup_M \mathbb{E} \left[z_1 h(M) - z_2 \int_0^1 \left(\frac{d\langle M \rangle_t}{dt} \right)^2 dt + 2z_2 z_3 \langle M \rangle_1 \right],$$

where the supremum is taken over the class \mathcal{M}_Θ of continuous martingales as in the formulation of this lemma which satisfy in addition that

$$\mathbb{E} \left[\int_0^1 \left(\frac{d\langle M \rangle_t}{dt} \right)^2 dt \right] \leq \Theta.$$

In particular, we obtain that $\hat{F}(z) < \infty$. By applying again the Doob–Kolmogorov inequality and the Jensen inequality it follows that for any $z, \tilde{z} \in O$ we have

$$\begin{aligned} & |\hat{F}(z) - \hat{F}(\tilde{z})| \\ & \leq \sup_{M \in \mathcal{M}_\Theta} \left(\mathbb{E} \left[z_1 h(M) - z_2 \int_0^1 \left(\frac{d\langle M \rangle_t}{dt} \right)^2 dt + 2z_2 z_3 \langle M \rangle_1 \right] \right. \\ & \quad \left. - \mathbb{E} \left[\tilde{z}_1 h(M) - \tilde{z}_2 \int_0^1 \left(\frac{d\langle M \rangle_t}{dt} \right)^2 dt + 2\tilde{z}_2 \tilde{z}_3 \langle M \rangle_1 \right] \right) \\ & \leq |z_1 - \tilde{z}_1|(c + 4\sqrt{\Theta}) + |z_2 - \tilde{z}_2|\Theta + 2|z_2 z_3 - \tilde{z}_2 \tilde{z}_3|\sqrt{\Theta} \end{aligned}$$

and continuity follows. □

We now have all the pieces in place that we need for the

Completion of the proof of the upper bound

Fix $\lambda > 0$. For $m = 1, 2, \dots$, choose $D^m \in \mathcal{D}^{1/m, \lambda}$ that get within $1/m$ of the supremum in (3.23) for $\varepsilon = 1/m$. For this sequence, let M^λ be a continuous martingale on $(\Omega^\lambda, \mathcal{F}^\lambda, \mathbb{P}^\lambda)$ as in Lemma 3.9. By Skorohod’s representation theorem, we can find copies of $D^m, m = 1, 2, \dots$, and M^λ (which to alleviate notation we denote by the same symbols) with the same respective distributions but specified jointly on a suitable probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that almost surely $(D^m, \int_0^\cdot \zeta_s^{D^m} ds)$ converges uniformly to $(M^\lambda, \langle M^\lambda \rangle)$ as $m \uparrow \infty$. From (3.19) we have $M_0^\lambda = p_0$ and from (3.22) the quadratic variation $\langle M^\lambda \rangle$ is absolutely continuous and the volatility process $\frac{d\langle M^\lambda \rangle_t}{dt}$ is bounded.

Now, use the regularity of h as imposed by Assumption 3.1 to conclude

$$\lim_m \mathbb{E}[h_{1+\lambda}(D^m)] = \mathbb{E}[h_{1+\lambda}(M^\lambda)]$$

by dominated convergence. Moreover, we can estimate

$$\liminf_m \mathbb{E} \left[\int_0^{1+\lambda} (\zeta_t^{D^m} - \sigma^2)^2 dt \right] \geq \mathbb{E} \left[\int_0^{1+\lambda} \left(\frac{d\langle M^\lambda \rangle_t}{dt} - \sigma^2 \right)^2 dt \right]. \tag{3.35}$$

Indeed, observing that the ζ^{D^m} are uniformly bounded for $m > 1/\lambda$ (cf. (3.22)), we can apply Lemma A1.1 in [9] to get $\tilde{\zeta}^m \in \text{conv}(\zeta^{D^m}, \zeta^{D^{m+1}}, \dots)$, $m = 1, 2, \dots$, converging $\mathbb{P} \otimes dt$ -almost everywhere to some process ζ . In fact, $\zeta = d\langle M^\lambda \rangle / dt$ because by dominated convergence $\int_0^\cdot \zeta_t dt = \lim_m \int_0^\cdot \tilde{\zeta}_t^m dt = \lim_m \int_0^\cdot \zeta_t^{D^m} dt = \langle M^\lambda \rangle$. As a consequence, the estimate (3.35) holds by the convexity of $\zeta \mapsto \mathbb{E}[\int_0^{1+\lambda} (\zeta_t - \sigma^2)^2 dt]$ and Fatou's lemma.

As $K(1/m, \lambda) \rightarrow \infty$ for $m \uparrow \infty$, it now follows from Lemma 3.8 that, for any $\lambda > 0$,

$$\begin{aligned} & \limsup_m \limsup_N \widehat{\pi}^N (H^{N, 1/m, K(1/m, \lambda)} / (1 - \lambda)) \\ & \leq \mathbb{E} \left[\frac{h_{1+\lambda}(M^\lambda)}{1 - \lambda} - \frac{\widehat{\delta}}{8\sigma^2} \int_0^{1+\lambda} \left(\frac{d\langle M^\lambda \rangle_t}{dt} - \sigma^2 \right)^2 dt \right] \\ & = \mathbb{E} \left[\frac{h(\tilde{M}^\lambda)}{1 - \lambda} - \frac{\widehat{\delta}}{8\sigma^2(1 + \lambda)} \int_0^1 \left(\frac{d\langle \tilde{M}^\lambda \rangle_t}{dt} - \sigma^2(1 + \lambda) \right)^2 dt \right], \end{aligned} \quad (3.36)$$

where \tilde{M}^λ is the martingale given by $\tilde{M}_t^\lambda = M_{(1+\lambda)t}^\lambda$, $0 \leq t \leq 1$.

By applying Lemma 3.10 we see that for $\lambda \downarrow 0$ the expectation in (3.36) cannot be larger than $\sup_M \mathbb{E}[h(M) - \frac{\widehat{\delta}}{8\sigma^2} \int_0^1 (\frac{d\langle M \rangle_t}{dt} - \sigma^2)^2 dt]$ where the supremum is taken over all the continuous martingales $M = (M_t)_{0 \leq t \leq 1}$ considered in Lemma 3.10. Using the randomization technique of Lemma 7.2 in [10], we thus find that

$$\lim_{\lambda \downarrow 0} \limsup_m \limsup_N \widehat{\pi}^N (H^{N, 1/m, K(1/m, \lambda)} / (1 - \lambda))$$

is dominated by the supremum on the right hand side of (3.3). In view of the estimate (3.4) from Lemma 3.5 in conjunction with Lemma 3.6, this implies the desired upper bound (3.3) for the case $x_0 = 0$ and $\zeta_0 = 0$.

For the case where $x_0 \neq 0$ or $\zeta_0 > 0$ we need to establish that

$$\begin{aligned} \limsup_N \pi^N (h(P^N)) & \leq \sup_{\nu \in \mathcal{D}} \mathbb{E}_{\mathbb{P}^\nu} \left[h(P^\nu) - \frac{r\delta}{8\sigma^2(2-r)} \int_0^1 |v_t^2 - \sigma^2|^2 dt \right] \\ & \quad - P_0 x_0 - \frac{\iota}{2} x_0^2. \end{aligned} \quad (3.37)$$

For the N -step market we use the first $N^{1/3}$ steps to liquidate with a constant speed the initial number of shares x_0 . The result is that after $N^{1/3}$ steps, the portfolio value will be $P_0 x_0 + \frac{\iota}{2} x_0^2 + O(N^{-1/6})$ and the spread will be bounded by $\zeta_0 + \frac{x_0}{\delta}$. The number of shares is zero.

In the next $N^{1/3}$ steps we do not trade at all, and so the spread will become of order $O((1-r)N^{1/3})$. Observe that for any $\tilde{\delta} > \delta$, we have that for sufficiently large N ,

$$\delta(z + O((1-r)N^{1/3}))^2 \leq \tilde{\delta}z^2 + (1-r)N^{1/4} \quad \text{for all } z \geq 0.$$

From Lemma 2.1 we conclude that the limsup of the original prices $\pi^N (h(P^N))$ is less than or equal to the limsup of the superhedging prices which correspond to the market depth $\tilde{\delta} > \delta$, the

same resilience r and an initial position $\tilde{x}_0 = \tilde{\zeta}_0 = 0$ minus $P_0x_0 + \frac{1}{2}x_0^2$. Thus, by taking $\tilde{\delta} \downarrow \delta$ and applying (3.3) (for $\tilde{\delta}$ instead of δ) and by using Lemma 3.10, we obtain (3.37). Let us notice that we should apply (3.3) for a shift in time of the original price process P^N . Since the shift in time is of order $O(N^{1/3})$ and h is Lipschitz continuous, the difference between the original payoff $h(P^N)$ and the modified one is of order $O(N^{1/3}N^{-1/2}) = O(N^{-1/6})$ which is vanishing in the limit $N \rightarrow \infty$.

Acknowledgments

Y. Dolinsky was partially supported by the ISF Grant 160/17.

References

- [1] Acciaio, B., Beiglböck, M., Penkner, F., Schachermayer, W. and Temme, J. (2013). A trajectorial interpretation of Doob's martingale inequalities. *Ann. Appl. Probab.* **23** 1494–1505. MR3098440 <https://doi.org/10.1214/12-aap878>
- [2] Alfonsi, A., Fruth, A. and Schied, A. (2010). Optimal execution strategies in limit order books with general shape functions. *Quant. Finance* **10** 143–157. MR2642960 <https://doi.org/10.1080/14697680802595700>
- [3] Bank, P. and Dolinsky, Y. (2019). Super-replication with fixed transaction costs. *Ann. Appl. Probab.* **29** 739–757. MR3910016 <https://doi.org/10.1214/17-AAP1372>
- [4] Bank, P. and Dolinsky, Y. (2019). Continuous-time duality for superreplication with transient price impact. *Ann. Appl. Probab.* **29** 3893–3917. MR4047995 <https://doi.org/10.1214/19-AAP1498>
- [5] Bank, P., Dolinsky, Y. and Gökay, S. (2016). Super-replication with nonlinear transaction costs and volatility uncertainty. *Ann. Appl. Probab.* **26** 1698–1726. MR3513603 <https://doi.org/10.1214/15-AAP1130>
- [6] Bank, P., Dolinsky, Y. and Perkkio, A.-P. (2017). The scaling limit of superreplication prices with small transaction costs in the multivariate case. *Finance Stoch.* **21** 487–508. MR3626623 <https://doi.org/10.1007/s00780-016-0320-4>
- [7] Bank, P. and Voß, M. (2019). Optimal investment with transient price impact. *SIAM J. Financial Math.* **10** 723–768. MR3995032 <https://doi.org/10.1137/18M1182267>
- [8] Çetin, U., Jarrow, R.A. and Protter, P. (2004). Liquidity risk and arbitrage pricing theory. *Finance Stoch.* **8** 311–341. MR2213255 <https://doi.org/10.1007/s00780-004-0123-x>
- [9] Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.* **300** 463–520. MR1304434 <https://doi.org/10.1007/BF01450498>
- [10] Dolinsky, Y. and Soner, H.M. (2013). Duality and convergence for binomial markets with friction. *Finance Stoch.* **17** 447–475. MR3066984 <https://doi.org/10.1007/s00780-012-0192-1>
- [11] Föllmer, H. and Kramkov, D. (1997). Optional decompositions under constraints. *Probab. Theory Related Fields* **109** 1–25. MR1469917 <https://doi.org/10.1007/s004400050122>
- [12] Gökay, S. and Soner, H.M. (2012). Liquidity in a binomial market. *Math. Finance* **22** 250–276. MR2897385 <https://doi.org/10.1111/j.1467-9965.2010.00462.x>
- [13] Huberman, G. and Stanzl, W. (2004). Price manipulation and quasi-arbitrage. *Econometrica* **72** 1247–1275. MR2064713 <https://doi.org/10.1111/j.1468-0262.2004.00531.x>
- [14] Kusuoka, S. (1995). Limit theorem on option replication cost with transaction costs. *Ann. Appl. Probab.* **5** 198–221. MR1325049

- [15] Obizhaeva, A.A. and Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *J. Financ. Mark.* **16** 1–32. <https://doi.org/10.1016/j.finmar.2012.09.001>
- [16] Predoiu, S., Shaikhet, G. and Shreve, S. (2011). Optimal execution in a general one-sided limit-order book. *SIAM J. Financial Math.* **2** 183–212. [MR2775411 https://doi.org/10.1137/10078534X](https://doi.org/10.1137/10078534X)
- [17] Roch, A. and Mete Soner, H. (2013). Resilient price impact of trading and the cost of illiquidity. *Int. J. Theor. Appl. Finance* **16** Art. ID 1350037. [MR3117871 https://doi.org/10.1142/S0219024913500374](https://doi.org/10.1142/S0219024913500374)
- [18] Soner, H.M., Shreve, S.E. and Cvitanić, J. (1995). There is no nontrivial hedging portfolio for option pricing with transaction costs. *Ann. Appl. Probab.* **5** 327–355. [MR1336872](https://doi.org/10.1214/aop/1176913111)
- [19] Strasser, H. (1985). *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. *De Gruyter Studies in Mathematics* **7**. Berlin: de Gruyter. [MR0812467 https://doi.org/10.1515/9783110850826](https://doi.org/10.1515/9783110850826)
- [20] Whitt, W. (2007). Proofs of the martingale FCLT. *Probab. Surv.* **4** 268–302. [MR2368952 https://doi.org/10.1214/07-PS122](https://doi.org/10.1214/07-PS122)

Received August 2019 and revised December 2019