LIQUIDITY RISK, PRICE IMPACTS AND THE REPLICATION PROBLEM

ALEXANDRE F. ROCH

ABSTRACT. We extend a linear version of the liquidity risk model of Çetin et al. (2004) to allow for price impacts. We show that the impact of a market order on prices depends on the size of the transaction and the level of liquidity. We obtain a simple characterization of self-financing trading strategies and a sufficient condition for no arbitrage. We consider a stochastic volatility model in which the volatility is partly correlated with the liquidity process and show that, with the use of variance swaps, contingent claims whose payoffs depend on the value of the asset can be approximately replicated in this setting. The replicating costs of such payoffs are obtained from the solutions of BSDEs with quadratic growth and analytical properties of these solutions are investigated.

1. INTRODUCTION

In financial markets, liquidity either refers to the ease with which financial securities can be bought and sold or to the ability to trade without triggering important changes in asset prices. Liquidity becomes a risk factor when the magnitude of the impact of these phenomena changes randomly over time. Uncertainty regarding the level of liquidity in traded assets has been for a long time a critical issue for moderate to large traders. The cost of a given trading strategy in real world situations can be substantially high when large quantities of financial assets are traded due to the consequential impact of trading on prices, and the limited and uncertain future supply and demand. In this paper, we construct an arbitrage-free model which relates levels of liquidity to trade impacts and quantify liquidity costs of strategies used for hedging claims contingent on the value of the traded asset.

The literature on liquidity risk is large and can be mainly divided according to these two conceptual perspectives. In the first category of models, the price of an asset depends on the size of the transaction and the depth of the order book. The second category includes those commonly known as "large trader" models in which a large trader buys and sells such large quantities of assets that his trades affect the prices in a non-negligible way. The purpose of this paper is to combine both approaches in a unified framework and to study the problem of contingent claim replication.

Examples of recent papers in the first category include Çetin, Jarrow and Protter [6] and Çetin and Rogers [7]. Rogers and Singh [20] give a microeconomic argument for a price which depends on size and this is then reflected in the dynamics of self-financing strategies. They solve an optimal control problem in this context.

Bank and Baum [2], Frey [12] and Jarrow [15] are examples of papers in which the impact of the large trader is a function of its current holdings. In Alfonsi et al. [1], the authors relate the impact of trades to the shape of the order book and consider the problem of optimal liquidation by the large trader. On the other hand, Ly Vath et al. [17] study the problem of optimal portfolio selection for a large trader who has a price impact function and cost function of exponential form.

Our present model was in part inspired by the liquidity risk model of Çetin, Jarrow and Protter [6] (thereafter referred to as the CJP model). In the CJP model, liquidity is introduced by hypothesizing the existence of a supply curve $S(t,x)$ which gives, at a given time t, the price per share to pay for a stock in terms of the size x of the trade. In such a model, the trader observes the supply curve and acts as a price taker. In this setting, liquidity costs essentially depend on the quadratic variation of the trading strategy. The main drawback of this model is that liquidity risk can essentially be avoided by approximating a given self-financing trading strategy (s.f.t.s.) by a sequence $(X^n)_{n\geq 1}$ of continuous s.f.t.s. with finite variation (FV) which incur no liquidity costs. The prices of options are then unaffected by liquidity risk. This issue was cleverly dealt with in Çetin et al. [8] by adding constraints on the gamma of the hedging strategies. A liquidity premium is then reflected in option prices.

Our approach is to combine both notions of liquidity risk by hypothesizing the existence of a random linear supply curve and by studying the impact of trades on prices. One of the key observations made in this paper is that the magnitude of price impacts is directly related to the amount of liquidity of the asset. This leads to a simple characterization of self-financing trading strategies in which the profit is partly affected by the level of liquidity. The main goal of this paper is to study the effect of liquidity risk on the replicating costs of contingent claims. We consider a stochastic volatility model in which the volatility process depends in part on the level of liquidity. We will see that variance swaps are the simplest hedging tools in this setting.

The paper is organized as follows. In Section 2, we derive the impact of trading on prices using simple principles and show that changes in the price of an asset is directly affected by the amount of liquidity. We then use these observations to propose a model defined on the Brownian filtration and show it is arbitrage-free. A simple characterization of self-financing strategies is derived to help set up the replication problem. Section 3 is devoted to the main result of this paper, the replication of contingent claims using variance swaps and the characterization of replication costs in terms of backward stochastic differential equations with quadratic growth. Section 4 presents useful analytical properties of these solutions.

2. THE SETUP

We consider an economy consisting of a risky asset (typically a stock) which is traded through a limit order book, its associated contingent claims and a risk-free asset. We take the point of view of a hedger who observes the limit order book of the stock and makes market orders (also known as marketable limit orders). We start by describing the supply curve the hedger would expect to observe if he did not trade. We call it the *unaffected supply curve* and denote it by S. It represents the limit order book that results from *all other traders' limit and market orders*. It is a conceptual construction which is not directly observed. We will assume that the hedger's trades have a lasting impact on prices which will be added to S to obtain the *actual observed supply curve*, which we denote by S^0.

We are given a fixed maturity T and $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\leq t\leq T}, \mathbb{P})$ a filtered probability space which satisfies the usual conditions. We assume the interest rate is constant, and for simplicity we work with the discounted price processes. The (discounted) unaffected price process is an exogenously given adapted continuous process $S = (S_t(x))_{t\geq 0, x\in\mathbb{R}}$ (or sometimes written $S(t,x)$ for convenience). $S_t(x)$ is the price per share for a transaction of size x at time t that would be observed if the hedger did not trade before time t. The actual (discounted) quoted price per share that all market participants obtain for a transaction of size x at time t is denoted by $S^0_t(x)$. We start by assuming that the *unaffected supply curve* has the following linear structure:

$$S_t(x) \quad = \quad S_t + M_t x, \text{ for } x \in \mathbb{R} \tag{2.1}$$

where $(M_t)_{t\geq 0}$ and $(S_t)_{t\geq 0}$ are positive semimartingales. Note that the fact that this function is continuous at $x=0$ implies there is no bid-ask spread. While it is theoretically possible for $S_t(x)$ to be negative for some values of x, it is unlikely to happen in practice since the value of M_t is small. We assume there is a measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, under which the unaffected price process S is local martingale. As in the classical theory, this assumption will be sufficient to rule out arbitrage opportunities. See Theorem 2.5 below in this regard.

The assumption that the supply curve is linear is supported by the empirical study of Blais [3] for stocks that are frequently traded in large volumes. The study was based on a large data set of stocks traded on several different stock exchanges in the year 2003. See also Blais and Protter [4].

Before we specify the precise model for S and M on which we will focus, we start by detailing general characteristics that a liquidity risk model which include price impacts should reflect.

Equation (2.1) gives us a way to describe the limit order book. We represent it by a density function $\rho_t(z)$ which denotes the density of the number of shares being offered at price z at time t, i.e. $\int_{z_1}^{z_2}\rho_t(z)dz$ is the number of shares offered between prices z_1 and z_2. If a trader wants to buy x shares at time t through a market order then the price to pay is $\int_{S_t}^{z_x} z\rho_t(z)dz$ in which z_x solves $\int_{S_t}^{z_x}\rho_t(z)dz = x$. It is clear from the linear structure of the supply curve that for any $t\leq T$ the density equals $\rho_t(z) = \frac{1}{2M_t}$. In that case, $z_x = S_t + 2M_t x$ and the dollar outlay for x shares is

$$\frac{1}{2M_t}\int_{S_t}^{S_t+2M_t x} z dz = S_t x + M_t x^2 = x S_t(x).$$

Since ρ is a measure of liquidity, we can think of M as a measure of illiquidity. Indeed, the larger is M_t, the higher is the liquidity cost.

We let X_t denote the number of shares owned by the hedger at time t and $S_t^0(x)$ denote the actual asset price per share observed in the market, which includes the impact of the hedger's trading strategy, i.e. $S_t^0(x)$ implicitly depends on X. We define $S_t^0 = S_t^0(0)$ as the observed quoted price.

We now describe the impact that an arbitrary market order has on the limit order book. We will then use these observations to justify our specification of S and S^0. First, one should observe that if ΔX_t shares are bought at time t by a trader through a market order, then the corresponding part of the order book is used up. This would mean that immediately after the trade the limit order book would have a density of 0 for prices between S_t^0 and $S_t^0 + 2M_t\Delta X_t$ and ρ_t elsewhere since the lowest ask price would then be $S_t^0 + 2M_t\Delta X_t$ whereas the highest bid would remain the same. In this perspective, one can see that an implicit assumption made in the liquidity model of Çetin et al. [6] is that new limit orders to sell are placed immediately after a trade, thereby filling up the limit order book to its previous levels since it is assumed that trades have no impact on the supply curve. The new observed quoted price is the same as before and the impact on prices is non-existent in [6]. Although it is reasonable to assume that the limit order book fills up to its previous level after a trade, it is not clear whether the gap should be filled by bid or ask orders. For example, if the gap is filled entirely by bid orders after the purchase of ΔX_t, then the new quoted price is shifted upwards to $S_t^0 + 2M_t\Delta X_t$. In this case, the outcome is a full impact on prices.

The empirical findings of Weber and Rosenow [21] showed that in practice the impact of trading on prices is important but can be less than the full impact described in the previous paragraph. In fact, they showed a negative correlation between returns and the volume of incoming limit orders which suggests that traders respond to buying market orders by adding new limit orders in the opposite direction. We model this phenomenon by introducing a parameter $\lambda\in[0,1]$ measuring the proportion of new bid orders (resp. ask orders) filling up the limit order book when a trade to buy (resp. sell) is made at time t. In effect, the effective impact on prices of a trade of size ΔX_t is to shift the quoted price to $S_t^0 + 2\lambda M_t \Delta X_t$, whereas the density level of the order book is unaffected.

We have to be careful how we define the observed price process in this setting. Indeed, when the hedger makes a trade at time t the price he pays is unaffected by the impact of this current trade whereas prices right after t will be. In this sense, S_t^0 will not be càdlàg in general, although S_{t+}^0 is and includes the impact of a trade at time t.

Suppose X is a simple trading strategy of the form $X = \sum_{k=0}^{k_n} \Delta_k^n X \mathbf{1}_{[\tau_k^n,\infty)}$ in which $\Delta_k^n X = X_{\tau_k^n} - X_{\tau_{k-1}^n}$ for $k = 1,\ldots,n$ and $\Delta_0^n X = X_0$. Then, the observed quoted price should satisfy

$$S_t^0 = S_t + 2\sum_{i=0}^{k-1} \lambda M_{\tau_i^n} \Delta_i^n X = S_t + 2\sum_{i=0}^{k-1} \lambda M_{\tau_{i-1}^n} \Delta_i^n X + 2\sum_{i=0}^{k-1} \lambda (\Delta_i^n M)(\Delta_i^n X)$$

for any $t \in (\tau_{k-1}^n, \tau_k^n]$. Note that the sum in the previous equation only goes up to $k-1$ since $S_{\tau_k^n}^0$, which represents the price per share for a trade of size 0, is not yet impacted by the trade at time τ_k^n. The right-limit version of this process is then given by

$$S_{t+}^0 = S_t + 2\sum_{i=0}^{k-1} \lambda M_{\tau_{i-1}^n} \Delta_i^n X + 2\sum_{i=0}^{k-1} \lambda (\Delta_i^n M)(\Delta_i^n X) \tag{2.2}$$

for any $t \in [\tau_{k-1}^n, \tau_k^n)$ when S is right-continuous. Following these observations, we define

$$S_{t+}^0 \quad = \quad S_t + 2\lambda \int_0^t M_{u-} dX_u + 2\lambda \int_0^t d[M,X]_u \tag{2.3}$$

for all $t \le T$, for a general semimartingale X. Furthermore, we define the observed quoted price by $S_t^0 = \lim_{s\uparrow t} S_{s+}^0$. By assuming that the level of liquidity ρ_t is unaffected by trades, we readily obtain that the supply curve is given by

$$S_t^0(x) = S_t^0 + M_t x \tag{2.4}$$

for all $0 < t \le T$ and $x \in \mathbb{R}$. We think of $1-\lambda$ as the fraction of the order book which is renewed after a market order so that in practice the actual impact on prices is λ times the full impact.

Equation 2.3 gives us a new understanding of the causes of volatility and its relation to illiquidity. As mentioned earlier, S is the price process which results from limit and market orders of all the other market participants. The equation suggests that the impact of the market orders of each market participant is proportional to the value of M. The volatility of S can then be expected to be correlated in part to M. (Another component of the volatility of S would be related to the volatility of limit orders.) In fact, many empirical works have shown that the level of liquidity is an important determinant of the variance of log-returns. The reader is referred to the works of Farmer et al. [11] and Weber and Rosenow [22] for a more detailed discussion. The observation that these authors make is that volatility is high when liquidity is low, and low when liquidity is high. Since M is a measure of illiquidity, we can expect the instantaneous variance of the log-returns of the stock price to be in part correlated with M. This is a key observation which will enable us to hedge derivatives. Indeed, in the next section, we will introduce variance swaps which will be used to hedge against the liquidity risk. Since volatility is one of the most correlated quantities to liquidity risk, this is a very natural choice. See Remark 3 in this regard.

Following these observations, we consider a stochastic volatility model for S:

$$dS_t \quad = \quad \Sigma_t S_t dW_{1,t}, \tag{2.5}$$

in which W_1 is a Brownian motion defined on the filtered probability space, and Σ_t is the stochastic volatility. Recall that we are working directly under a risk neutral measure $\mathbb{Q}$ for unaffected prices, hence S has no drift term. We model M and Σ as follows. Define V and U as the solutions of

$$\begin{aligned} dU_t &= \gamma(U_t+\eta)dt+\Phi(U_t)dW_{2,t}, \\ dV_t &= \alpha(V_t+a)dt+\Theta(V_t)dW_{3,t} \end{aligned}$$

in which $W=(W_{j,t})_{j\leq 3,t\leq T}$ is a three dimensional Brownian motion defined on the filtered probability space, and $\alpha,\gamma,\eta,a\in\mathbb{R}$. We define $\Sigma_t^2=U_t+V_t$ and let $M=\varepsilon\Gamma(U)$, in which Γ is strictly increasing and twice continuously differentiable. In practice, the process M takes small values compared to Σ, but is also an important component of the volatility process Σ. As a result, the constant ε is typically small.

We are using a three dimensional Brownian motion since there are three different sources of risk in this model, namely the stock price, the liquidity level and the volatility, which is, in practice, only partially dependent on the level of liquidity. The components of W are typically correlated and we denote by $R=\frac{1}{t}COV(W_t)$ the matrix of instantaneous correlation coefficients. We assume R is positive definite and we let L be the upper triangular matrix in the Cholesky decomposition such that $R^{-1}=L^{\top}L$. We then define $B=LW$. Then B is a three-dimensional Brownian motion with independent components. We denote the components of L^{-1} by

$$L^{-1}=\begin{pmatrix} \sigma_1 & \sigma_2 & \sigma_3 \\ 0 & \phi_2 & \phi_3 \\ 0 & 0 & \theta_3 \end{pmatrix}.$$

We assume the functions Θ and Φ are chosen so that the solutions of the above stochastic differential equations are well defined. For example, one can take $\Theta(v)=v^{\hat{\theta}}$ with $\hat{\theta}=0,\frac{1}{2}$ or 1. Examples of stochastic volatility models of this form are Heston [13] ($\hat{\theta}=\frac{1}{2}$), Hull and White [14] ($\hat{\theta}=1$), and Detemple and Osakwe [10]. Other expressions for Σ^2 could be used, however we have chosen this particular form for its mathematical tractability and its widespread use in theory and practice.

2.1. **Self-Financing Strategies and No Arbitrage.** In order to properly address the problem of replicating contingent claims, we give a characterization of self-financing strategies and establish under which condition our model is arbitrage-free. In our setting, the self-financing condition is as follows.

Definition 2.1. Let $\pi_n: t_0=\tau_0^n\leq\tau_1^n\leq\ldots\leq\tau_{k_n}^n=T$ be a sequence of random partitions tending to the identity. A pair of processes $(X_t,Y_t)_{t_0\leq t\leq T}$ is a self-financing trading strategy (s.f.t.s.) on $[t_0,T]$ if X is a semimartingale and Y is an optional process satisfying

$$Y_t = Y_{t_0-}-\Delta X_{t_0}S^0(t_0,\Delta X_{t_0})-\lim_{n\to\infty}\sum_{k=1}^{k_n}\Delta_k^n X S^0(\tau_k^n,\Delta_k^n X)1_{\{\tau_k^n\leq t\}} \tag{2.6}$$

where convergence is in ucp. (See Protter [19] for undefined terms.) Here, $\Delta_k^n X=X_{\tau_k^n}-X_{\tau_{k-1}^n}$ for $k=1,\ldots,n$.

X_t represents the number of shares of the asset owned by the hedger and Y_t is the position in the risk-free asset at time t. The interpretation is that the position in the risk-free asset at time t should be equal to the position at time t_0 minus the cost of all the trades between t_0 and t. Here, Y_{t_0-} is the value of the position in the risk-free asset before the trade at time t_0.

Remark. In the classical theory, the process X is predictable. We take X in the above definition to be a semimartingale for the stochastic integral in Equation 2.3 to be well defined. A consequence of Proposition 2.2 below is that the limit in Equation 2.6 is well-defined, and the definition of self-financing trading strategies is independent of the sequence of random partitions.

Even though s.f.t.s. are defined in terms of S^0, they can be characterized in terms of the exogenously given processes M and S as follows:

Proposition 2.2. *Let $t_0 > 0$. If $(X_s, Y_s)_{t_0 \le s \le T}$ is a self-financing trading strategy then*

$$\begin{aligned} Y_t + X_t(S^0_{t+} - \lambda M_t X_t) &= Y_{t_0-} + X_{t_0-}(S^0_{t_0} - \lambda M_{t_0} X_{t_0-}) + \int_{t_0}^t X_{u-} dS_u \\ &\quad - \lambda \int_{t_0}^t X^2_{u-} dM_u - \int_{t_0}^t (1-\lambda) M_u d[X,X]_u \end{aligned} \tag{2.7}$$

for all $t_0 \le t \le T$.

Proof. Let $\pi_n : t_0 = \tau^n_0 \le \tau^n_1 \le \ldots \le \tau^n_{k_n} = t$ be a sequence of random partitions tending to the identity. The self-financing condition is

$$Y_t = Y_{t_0-} - \Delta X_{t_0} S^0(t_0, \Delta X_{t_0}) - \lim_{n\to\infty} \sum_{i=1}^{k_n} \Delta^n_i X \left(S^0_{\tau^n_i} + M_{\tau^n_i} \Delta^n_i X \right)$$

where the convergence is in ucp. We can expand the sum in the last equation to find $-\lim_{n\to\infty} \sum_{i=1}^{k_n} \Delta^n_i X \left(S^0_{\tau^n_i} + M_{\tau^n_i} \Delta^n_i X \right)$

$$\begin{aligned}
&= -\lim_{n\to\infty} \sum_{i=1}^{k_n} \left(X_{\tau^n_i} S^0_{\tau^n_i} - X_{\tau^n_{i-1}} S^0_{\tau^n_{i-1}} \right) + \lim_{n\to\infty} \sum_{i=1}^{k_n} X_{\tau^n_{i-1}} \Delta^n_i S^0 - \lim_{n\to\infty} \sum_{i=1}^{k_n} M_{\tau^n_i} (\Delta^n_i X)^2 \\
&= -X_t S^0_t + X_{t_0} S^0_{t_0} + \lim_{n\to\infty} \sum_{i=1}^{k_n} \left(X_{\tau^n_{i-1}} \Delta^n_i S + 2\lambda M_{\tau^n_{i-1}} X_{\tau^n_{i-1}} \Delta^n_{i-1} X \right) - \lim_{n\to\infty} \sum_{i=1}^{k_n} M_{\tau^n_i} (\Delta^n_i X)^2 \\
&= -X_t S^0_t + X_{t_0} S^0_{t_0} + 2\lambda M_{t_0} X_{t_0} \Delta X_{t_0} - 2\lambda M_t X_t \Delta X_t + \lim_{n\to\infty} \sum_{i=1}^{k_n} X_{\tau^n_{i-1}} \Delta^n_i S \\
&\quad + 2 \lim_{n\to\infty} \sum_{i=1}^{k_n} \lambda M_{\tau^n_i} X_{\tau^n_i} \Delta^n_i X - \lim_{n\to\infty} \sum_{i=1}^{k_n} \lambda M_{\tau^n_i} (\Delta^n_i X)^2 - \lim_{n\to\infty} \sum_{i=1}^{k_n} (1-\lambda) M_{\tau^n_i} (\Delta^n_i X)^2 \\
&= -X_t S^0_{t+} + X_{t_0} S^0_{t_0+} + \lim_{n\to\infty} \sum_{i=1}^{k_n} X_{\tau^n_{i-1}} \Delta^n_i S + \lim_{n\to\infty} \sum_{i=1}^{k_n} \lambda M_{\tau^n_i} \Delta^n_i X^2 \\
&\quad - \lim_{n\to\infty} \sum_{i=1}^{k_n} (1-\lambda) M_{\tau^n_i} (\Delta^n_i X)^2
\end{aligned}$$

$$
\begin{aligned}
&= -X_t S^0_{t+} + X_{t_0} S^0_{t_0+} + \lambda M_t X_t^2 - \lambda M_{t_0} X_{t_0}^2 + \lim_{n\to\infty} \sum_{i=1}^{k_n} X_{\tau^n_{i-1}} \Delta^n_i S - \lim_{n\to\infty} \sum_{i=1}^{k_n} \lambda X^2_{\tau^n_{i-1}} \Delta^n_i M \\
&\quad - \lim_{n\to\infty} \sum_{i=1}^{k_n} (1-\lambda) M_{\tau^n_i} (\Delta^n_i X)^2 \\
&= -X_t(S^0_{t+} - \lambda M_t X_t) + X_{t_0}(S^0_{t_0+} - \lambda M_{t_0} X_{t_0}) + \int_{t_0}^t X_{u-} dS_u \\
&\quad -\lambda \int_{t_0}^t X^2_{u-} dM_u - \int_{t_0}^t (1-\lambda) M_u d[X,X]_u
\end{aligned}
$$

by Theorem 21 (Chapter II) of Protter [19] since X is càdlàg. □

One can think of $Y_t + x(S^0_t - \lambda M_t x)$ as the liquidation value of a portfolio with x shares at time t. Indeed, take $t_0 = t$ and $\Delta X_t = X_{t-}$ in Equation 2.7. Then one finds that the cash value of a position X_{t-} at time $t-$ in the stock is equal to $\Delta Y_t = X_{t-}(S^0_t - \lambda M_t X_{t-}) - (1-\lambda) M_t (\Delta X_t)^2$ if it is liquidated at time t. Furthermore, if one uses a sequence X^n a continuous and FV processes converging to X (this can be done by Lemma 3.2), then the liquidation value converges to $X_{t-}(S^0_t - \lambda M_t X_{t-})$. Consequently, λM_t can be interpreted as the effective liquidity parameter.

Similar to the infinitely-liquid case ($M = 0$), Equation 2.7 states that the difference in the liquidation values between time t_0 and t is equal the cumulative gains in the risky asset $\int_{t_0}^t X_{u-} dS_u$, except that in this case there are added costs coming from the finite liquidity of the asset. First note that if $\lambda = 0$ we get a linear version of the CJP model. The integral with respect to M is related to the impact of trading. If $\lambda = 0$, the limit order book is automatically refilled after a market order, as in the CJP model. At the other extreme, when $\lambda = 1$ the impact of trading is at its fullest. It is interesting to notice that whatever the trading strategy used an investor always has a partial benefit from the asset becoming more liquid. Indeed, when M_t decreases, the associated integral is positive no matter what the sign of X_t is. To understand this, it is important to remember that the hedger's trades have a permanent impact on the quoted price which is proportional to the level of liquidity. If the liquidity is low when he purchases a share and high when he sells it, the price goes up higher after his purchase then it comes down after the sale. As a result, the hedger has a partial gain from this trade. This is a typical characteristic of large trader models. Note that unless the hedger uses a trading strategy with zero quadratic variation this is only a partial benefit because there is always a liquidity cost associated to his trades.

Using Proposition 2.2, for $y \in \mathbb{R}$, we define the set $\mathscr{X}_y$ of payoffs of maturity T attainable at price y by $\mathscr{F}_T$-measurable random variables Y_T of the type

$$
Y_T = y + \int_0^T X_{t-} dS_t - \lambda \int_0^T X^2_{t-} dM_t - \int_0^T (1-\lambda) M_t d[X,X]_t
$$

in which $(X_t)_{t\geq 0}$ is càdlàg with finite quadratic variation.

We will denote by $\mathscr{X} \stackrel{def}{=} \bigcup_{y\in\mathbb{R}} \mathscr{X}_y$ the set of all attainable payoffs. We use the following definition of admissibility.

Definition 2.3. Let $a \geq 0$. A s.f.t.s. $(X_t, Y_t)_{t\geq 0}$ is a-admissible if

$$
\int_0^t X_{s-} dS_s - \lambda \int_0^t X^2_{s-} dM_s - \int_0^t (1-\lambda) M_s d[X,X]_s \geq -a
$$

for all $t \leq T$. The s.f.t.s. $(X_t, Y_t)_{\{t\geq 0\}}$ is simply said to be admissible if it is a-admissible for some $a \geq 0$.

A strategy is admissible if its payoff is bounded from below. In particular, this definition rules out doubling strategies and is well known to be a key element in the definition of arbitrage opportunities. See Delbaen and Schachermayer [9] in this regard.

Definition 2.4. An arbitrage opportunity is an admissible s.f.t.s. whose payoff $Y_T \in \mathscr{L}_0$ satisfies

$$\mathbb{P}\{Y_T \geq 0\} = 1 \qquad \text{and} \qquad \mathbb{P}\{Y_T > 0\} > 0. \tag{2.8}$$

It is already known (see [6]) that the existence of a local martingale measure for S rules out arbitrage opportunities in the CJP model. In the presence of trade impacts, the equation for the payoff of a s.f.t.s. has an integral with respect to M. Since the integrand of this integral is always negative $(-\lambda X_{t-}^2)$, then the part of the profit coming from this integral will be negative on average if M is a submartingale under the risk neutral measure. This idea is made precise in the following theorem which gives a sufficient condition for no arbitrage.

Theorem 2.5. *If there exists a measure $\mathbb{Q} \sim \mathbb{P}$ under which S is a $\mathbb{Q}$-local martingale and M is a $\mathbb{Q}$-local submartingale, then there are no arbitrage opportunities.*

Proof. By the Doob-Meyer decomposition theorem there exists a $\mathbb{Q}$-local martingale $\widetilde{M}$ and an increasing predictable process A such that $M = \widetilde{M} + A$. Let $Z_t = \int_0^t X_{u-} dS_u - \lambda \int_0^t X_{u-}^2 dM_u - \int_0^t (1-\lambda) M_u d[X,X]_u$. Then $Z_t + \lambda \int_0^t X_{u-}^2 dA_u + \int_0^t (1-\lambda) M_u d[X,X]_u = \int_0^t X_{u-} dS_u - \lambda \int_0^t X_{u-}^2 d\widetilde{M}_u \geq -a$ since A is increasing and $\int_0^t (1-\lambda) M_u d[X,X]_u \geq 0$. Now, S and $\widetilde{M}$ are $\mathbb{Q}$-local martingales hence $Z_t + \lambda \int_0^t X_{u-}^2 dA_u + \int_0^t (1-\lambda) M_u d[X,X]_u$ is also a local martingale and because it is bounded from below it is a supermartingale. Therefore, Z is also a supermartingale and $\mathbf{E}_{\mathbb{Q}} Z_T \leq 0$. But, because $\mathbb{Q} \sim \mathbb{P}$, if Z_T were an arbitrage opportunity it would also satisfy Equation 2.8 with $\mathbb{Q}$ instead of $\mathbb{P}$ and $\mathbf{E}_{\mathbb{Q}} Z_T > 0$. $\square$

In the simplest case, when $\Gamma(x) = x$, it suffices to take γ and $\eta > 0$. In the case that $\Gamma(x) = x^2$, if $\Phi(m) = \sqrt{m}$ then we need $\gamma\eta \geq \frac{1}{4}$; if $\Phi(m) = m$ then we must have $\gamma \geq \frac{1}{4}$. In the remaining parts of the paper, all expectations are with respect to $\mathbb{Q}$.

3. The Replication Problem

We now turn to the problem of contingent claims replication. Because the presence of the processes M and Σ involve risks that cannot be hedged completely by solely trading the stock, not all payoffs are attainable when only the underlying asset is allowed to be traded. Because these two processes are components of the instantaneous variance of the log-returns of the stock, the natural hedging instruments to consider are variance swaps. We thus consider contingent claims denoted by G_i ($i = 1,2$) for which the payoff at time $T_i > T$ ($T_1 \neq T_2$) equals the difference between the realized variance over the time interval $[0, T_i]$ and a strike K_i, i.e.,

$$G_{i,T_i} = \int_0^{T_i} \Sigma_s^2 ds - K_i = \int_0^{T_i} (U_s + V_s) ds - K_i.$$

To rule out arbitrage opportunities, we assume the unaffected price processes G^i are $\mathbb{Q}$-martingales $(i = 1,2)$, i.e.

$$G_t^i = \mathbf{E}\left(\int_0^{T_i} \Sigma_s^2 ds - K_i \middle| \mathscr{F}_t \right)$$

for all $t \le T_i$. We further assume the G_i's have a linear supply curve, i.e. $G_{i,t}(x) = G_{i,t} + xM'_{i,t}$ for all x and $t \le T$. Since it is not infinitely liquid, trading G_i can affect its price and we denote by $\lambda_i M'_{i,t}$ its effective liquidity. Typically, changes in the supply curves of the G_i's will happen less often. Hence, to keep the problem tractable, we assume that $M'_{i,t} \equiv M'_{i,0}$ is some given positive constant for all $t \in [0,T]$. We will see that two of these swaps are sufficient to hedge against liquidity risks. Because we now have two more traded assets, $\chi_{1,t}$ denotes the number of shares of G_1 and $\chi_{2,t}$ the number of shares of G_2 in the portfolio at time t. We can easily extend the definition of s.f.t.s. to the case of three traded securities. As shown before, s.f.t.s. (X,χ,Y) satisfy

$$\begin{aligned} Y_T &= Y_t + \int_t^T X_{u-} dS_u + \sum_{i=1,2} \int_t^T \chi_{i,u-} dG_{i,u} - \lambda \int_t^T X_{u-}^2 dM_u \\ &\quad - \int_t^T (1-\lambda) M_u d[X,X]_u - \sum_{i=1,2} \int_t^T (1-\lambda_i) M'_{i,u} d[\chi_i,\chi_i]_u \end{aligned}$$

for $t_0 \le t \le T$, when $X_{t_0} = \chi_{1,t_0} = \chi_{2,t_0} = X_T = \chi_{1,T} = \chi_{2,T} = 0$.

The following proposition gives condition under which the three price processes S, G_1, G_2 are non-redundant. It justifies the choice of variance swaps by providing a simple explicit representation of the processes G_i. This result will then be used to solve the replication problem.

Proposition 3.1. *Suppose $\alpha \neq \gamma$ and Θ (resp. Φ) satisfies one of the following conditions:*

(1) $\Theta(v) = v^{\hat\theta}$ (resp. $\Phi(m) = m^{\hat\phi}$) for $\hat\theta \in [0,\frac{1}{2}]$ (resp. $\hat\phi \in [0,\frac{1}{2}]$),
(2) Θ (resp. Φ) is Lipschitz continuous.

Then, there exists a predictable process $\psi = (\psi_{i,j,t})_{1\le i,j\le 3, 0\le t\le T}$ in $\mathbb{R}^{2\times 2}$ such that

$$\begin{aligned} G_t^i &= \mathbf{E}\left(\int_0^{T_i} \Sigma_s^2 ds - K_i\right) + \sum_{j=1,2,3} \int_0^t \psi_{i,j,s} dB_{j,s} \text{ and} \\ S_t &= \sum_{j=1,2,3} \int_0^t \psi_{3,j,s} dB_{j,s} \end{aligned}$$

for all $t \le T$ and $i = 1,2$. Furthermore, $(\psi_{i,j,t})_{1\le i,j\le 3}$ is invertible for all t.

Proof. Consider the process $\tilde V_t := e^{-\alpha t}(V_t + a)$ for $t \le T$. Then,

$$d\tilde V_t = \sum_{i=1,2,3} e^{-\alpha t} \theta_i \Theta(V_t) dB_{i,t}.$$

(We let $\theta_1 = \theta_2 = \phi_1 = 0$.) In other words, $\tilde V$ is a local martingale. We first show that $\tilde V$ is in fact a martingale. Suppose Θ is Lipschitz continuous. By the Burkholder-Davis-Gundy Inequality, there exists a positive constant C such that

$$\begin{aligned} \mathbf{E} \sup_{t\le T} \tilde V_t^2 &\le C\mathbf{E} \int_0^T e^{-2\alpha t} \Theta(V_t)^2 dt \\ &\le C\mathbf{E} \int_0^T V_t^2 dt + C \le C \int_0^T \mathbf{E} V_t^2 dt + C < \infty \end{aligned}$$

by well known estimates of moments of solutions of stochastic differential equations with Lipschitz coefficients. On the other hand, if $\Theta(v)=v^{\hat\theta}$ for $\hat\theta\in[0,\frac{1}{2}]$, then

$$\begin{aligned}\mathbf{E}\sup_{t\le T}\tilde V_t^2 &\le C\int_0^T \mathbf{E}V_t^{2\hat\theta}dt \le C\int_0^T (\mathbf{E}V_t)^{2\hat\theta}dt\\ &\le C\int_0^T \left(e^{\alpha t}\mathbf{E}\tilde V_t\right)^{2\hat\theta}dt \le C\int_0^T \left(e^{\alpha t}\tilde V_0\right)^{2\hat\theta}dt<\infty\end{aligned}$$

since $\tilde V$ is a positive local martingale. Hence $\tilde V$ is a martingale. Similarly, we can show that the process $\tilde U$, defined by $\tilde U_t := e^{-\gamma t}(U_t+\eta)$ for $t\le T$, is a martingale when Φ satisfies Condition 1 or 2.

Suppose $\gamma,\alpha\neq 0$. Then,

$$\begin{aligned}\mathbf{E}\left(\int_0^{T_i} U_s ds|\mathscr{F}_t\right) &= \int_0^t U_s ds+\mathbf{E}\left(\int_t^{T_i}\left(e^{\gamma s}\tilde U_s-\eta ds\right)|\mathscr{F}_t\right)\\ &= \int_0^t U_s ds+\int_t^{T_i} e^{\gamma s}\left(\mathbf{E}\left(\tilde U_s|\mathscr{F}_t\right)-\eta\right)ds\\ &= \int_0^t U_s ds+\int_t^{T_i}\left(e^{\gamma s}\tilde U_t-\eta\right)ds\\ &= \int_0^t U_s ds+\tilde U_t\left(\frac{e^{\gamma T_i}-e^{\gamma t}}{\gamma}\right)-\eta(T_i-t)\\ &= \left(\frac{e^{\gamma T_i}-1}{\gamma}\right)\tilde U_0-\eta T_i+\int_0^t\left(\frac{e^{\gamma T_i}-e^{\gamma s}}{\gamma}\right)d\tilde U_s.\end{aligned}$$

In particular,

$$\begin{aligned}G_t^i &= \mathbf{E}\left(\int_0^{T_i} U_s ds|\mathscr{F}_t\right)+\mathbf{E}\left(\int_0^{T_i} V_s ds|\mathscr{F}_t\right)-K_i\\ &= \left(\frac{e^{\gamma T_i}-1}{\gamma}\right)\tilde U_0-\eta T_i+\left(\frac{e^{\alpha T_i}-1}{\alpha}\right)\tilde V_0-aT_i-K_i\\ &\quad+\int_0^t\left(\frac{e^{\gamma T_i}-e^{\gamma s}}{\gamma}\right)d\tilde U_s+\int_0^t\left(\frac{e^{\alpha T_i}-e^{\alpha s}}{\alpha}\right)d\tilde V_s\\ &= \left(\frac{e^{\gamma T_i}-1}{\gamma}\right)\tilde U_0-\eta T_i+\left(\frac{e^{\alpha T_i}-1}{\alpha}\right)\tilde V_0-aT_i-K_i+\sum_{j=1,2,3}\int_0^t \psi_{i,j,s}dB_{j,s}\end{aligned}$$

in which

$$\psi_{i,j,t}=\left(\frac{e^{\gamma T_i}-e^{\gamma t}}{\gamma}\right)\phi_j\Phi(U_t)+\left(\frac{e^{\alpha T_i}-e^{\alpha t}}{\alpha}\right)\theta_j\Theta(V_t)$$

for $i=1,2$ and $j=1,2,3$. Define $\psi_{3,j,t}=\sigma_j\Sigma_t S_t$ for $j=1,2,3$. Note that $\psi_{i,3,t}=0$ for $i=1,2$. Since

$$(\psi_{i,j,t})_{1\le i\le 2,2\le j\le 3}=\begin{pmatrix}\frac{e^{\gamma T_1}-e^{\gamma t}}{\gamma} & \frac{e^{\alpha T_1}-e^{\alpha t}}{\alpha}\\ \frac{e^{\gamma T_2}-e^{\gamma t}}{\gamma} & \frac{e^{\alpha T_2}-e^{\alpha t}}{\alpha}\end{pmatrix}\begin{pmatrix}\phi_2\Phi(U_t) & \phi_3\Phi(U_t)\\ 0 & \theta_3\Theta(V_t)\end{pmatrix}$$

is invertible, so is ψ_t. In the case that α (resp. γ) is equal to zero, the term $\frac{e^{\alpha T_i}-e^{\alpha t}}{\alpha}$ (resp. $\frac{e^{\gamma T_i}-e^{\gamma t}}{\gamma}$) in the above matrix is replaced by T_i-t, and the matrix is also invertible when $\alpha\neq\gamma$. $\square$

Remark. The fact that the matrix ψ can be explicitly obtained and shown to be invertible is the main benefit of using variance swaps to complete the market. Similar calculations for non-linear contingent claims like put and call options on the stock or the realized variance would have been much more difficult, if not impossible, to obtain. As a result, such non-linear contingent claims would make the hedging much more difficult in practice. Note that the processes U and V need not be martingales under the risk neutral measure, i.e. $\alpha,\gamma \neq 0$. Consequently, $\int_0^t \Sigma_s^2 ds$ is not a martingale and $G_{i,t} \neq \int_0^t \Sigma_s^2 ds - K_i$ for $i = 1,2$. If that were the case, one of the two variance swaps would be redundant.

From now on, we assume that $\alpha \neq \gamma$ and that Θ and Φ satisfy the assumptions of the previous proposition.

The next lemma implies that the best way of trading is always to use FV continuous s.f.t.s. to avoid liquidity costs coming from the quadratic variation of X. In this sense, trades should always be done at the quoted price $S(t,0)$. Note that even though some of the liquidity costs in Equation 2.7 are eliminated when using continuous FV strategies, liquidity risk has not been completely eliminated from the model since the integral $\int_t^T X_{u-}^2 dM_u$ is still present. That is the main difference between our setup and the CJP model.

If S is a special semimartingale with canonical decomposition $S = N + A$, i.e. in which N is a local martingale and A is a predictable and FV process, then the $\mathscr{H}^2$ norm of S is defined as

$$\|S\|_{\mathscr{H}^2} = \|\sqrt{[N]_T}\|_{L^2} + \|\int_0^T |dA_s|\|_{L^2}.$$

Lemma 3.2. *Let S be a special semimartingale and X be predictable and integrable with respect to S. There exists a sequence $\{X^n\}_n$ of bounded continuous processes with finite variation such that $X_0^n = X_T^n = 0$ and X^n converges to X in $\mathscr{H}^2$. In particular, $\int X^n dS \to \int X dS$ in $\mathscr{H}^2$.*

Proof. The statement is proved in the proof of Lemma 4.1 of Çetin et al. [6]. □

We will see that, because of the quadratic variation term in the equation of s.f.t.s., it is not possible to replicate exactly in general. Since continuous processes with finite variation have zero quadratic variation, the previous lemma will prove to be useful for the replication problem. Following Çetin et al. [6], we make the following definition.

Definition 3.3. $H \in L^1$ can be approximately replicated if there exists a sequence $(X^n, \chi^n, Y^n)_{n\geq 1}$ of s.f.t.s. such that $Y_T^n \to H$ in L^1.

In the presence of trade impacts, the process S^0 implicitly depends on X and its value at the maturity is

$$S_{T+}^0 = S_T + 2\lambda \int_0^T M_{u-} dX_u + 2\lambda \int_0^T d[M,X]_u = S_T - 2\lambda \int_0^T X_{u-} dM_u$$

when $X_T = 0$ and $X_0 = 0$. (Here we use the time $T+$ and $X_T = 0$ to make sure that the hedging strategy is liquidated before the payoff is calculated to avoid discrepancies between the observed asset price before and after the maturity.) The true replication problem involves finding a s.f.t.s. (X,Y) that replicates a terminal condition which itself depends on X. Instead, for each $x \in \mathbb{R}$, we consider the replication of the terminal condition given by $xh(\widetilde{S}_T^x)$ with $\widetilde{S}_T^x := S_T - 2\lambda \int_0^T x\hat{X}_{u-} dM_u$ in which $\hat{X}$ is the solution of the replication problem in the case $\lambda = 0$, $\varepsilon = 0$ and $x = 1$. Jarrow [15] used a similar approach and interpreted $\hat{X}_t$ as the market's perception of the option's "delta" X_t. In the expression for $\widetilde{S}_T^x$, x denotes the number of units to be replicated. Hence, the proposed approximation for the true delta for the replication of x units is xX_t. Proposition 4.2 in the next section gives an upper bound of the error introduced by this approximation. Let us begin by giving an overview of the replication problem in this simplified setting.

3.1. **Contingent Claims Replication Without Trade Impact and Liquidity Costs.** When $\lambda = 0$ and $\varepsilon = 0$, the s.f.t.s. $(X_s, Y_s)_{t \le s \le T}$ that replicates a payoff $H \in L^1$ satisfies

$$H = Y_t + \int_t^T X_{u-} dS_u + \sum_{i=1,2} \int_t^T \chi_{i,u-} dG_{i,u}. \tag{3.1}$$

Also, $S^0 \equiv S$.

First, note that Equation 3.1 is equivalent to the following linear backward stochastic differential equation (BSDE):

$$Y_t = H - \sum_{j=1}^{3} \int_t^T \left(\sigma_j \Sigma_s X_s S_s + \chi_{1,s} \psi_{1,j,s} + \chi_{2,s} \psi_{2,j,s}\right) dB_{j,s}, \tag{3.2}$$

$0 \le t \le T$. Setting

$$Z_{j,t} = \chi_{1,s} \psi_{1,j,s} + \chi_{2,s} \psi_{2,j,s} + X_s \psi_{3,j,s} \tag{3.3}$$

for $j = 1,2,3$, the BSDE can be written as

$$Y_t = H - \sum_{j=1}^{3} \int_t^T Z_{j,s} dB_{j,s} \quad (0 \le t \le T). \tag{3.4}$$

When $H \in L^2$, BSDE 3.4 has a unique solution (Z,Y) in $M^2 \times M^2$ (see Pardoux and Peng [18] for example). Since ψ_t is invertible, we can define $X_s = \frac{Z_{1,s}}{\sigma_1 \Sigma_s S_s}$ $(t \le s \le T)$ and χ_s by inverting Equation 3.3. Then $(X_s, \chi_s, Y_s)_{t \le s \le T}$ is the solution of 3.1.

3.2. **The Replication Problem With Liquidity Risk.** From now on, we denote by $(\hat{X}, \hat{\chi}, \hat{Y})$ the solution of 3.2 with terminal condition $H = h(S_T)$. Recall that $\widetilde{S}_T^x := S_T - 2x\lambda \int_0^T \hat{X}_u dM_u$. The main result of this section is the following theorem.

Theorem 3.4. *Let $h : \mathbb{R}^+ \to \mathbb{R}$ be Lipschitz continuous. Then $xh(\widetilde{S}_T^x)$ can be approximately replicated for all $x \in \mathbb{R}$.*

Proof. Let $L > 0$ and $N > 0$. Let $x \in \mathbb{R}$ and h satisfy the conditions of the theorem, and define $h^N(y) = h(y)$ if $|y| \le N$ and $h^N(y) = h(N)$ otherwise. Since h is continuous on $[-N,N]$, h^N is bounded. Denote this bound by C_N. Define $H_T^N = xh^N(\widetilde{S}_T^x)$ and

$$\tau_L = \inf\{0 \le u \le T : S_u \le \frac{1}{L} \text{ or } \Sigma_u \ge L \text{ or } \Sigma_u \le \frac{1}{L}\}.$$

Consider the following BSDE:

$$Y_t = H^{N,L} - \int_t^{\tau_L} X_s dS_s + \lambda \int_t^{\tau_L} X_s^2 dM_s - \sum_i \int_t^{\tau_L} \chi_{i,s} dG_{i,s} \tag{3.5}$$

for $0 \le t \le \tau_L$ in which $H^{N,L} = \mathbf{E}\left(H_T^N | \mathscr{F}_{\tau_L}\right)$. It can be re-written as

$$H^{N,L} = Y_t - \lambda \int_t^{\tau_L} Z_{1,u}^2 \Lambda_u du + \sum_{i=1}^{3} \int_t^{\tau_L} Z_{i,u} dB_{i,u} \tag{3.6}$$

with

$$Z_{i,u} \quad = \quad \sigma_i \Sigma_u S_u X_u - \phi_i \zeta(M_t) X_u^2 + \sum_{j=1,2} \psi_{i,j,t} \chi_{j,u} \tag{3.7}$$

for $i = 2,3$, $Z_{1,u} = \sigma_1 \Sigma_u S_u X_u$ and $\Lambda_u = \frac{\mu(M_u)}{\sigma_1^2 \Sigma_u^2 S_u^2}$, in which

$$\mu(x) = \varepsilon\gamma(\Gamma^{-1}(x) + \eta)\Gamma'(\Gamma^{-1}(x)) + \frac{1}{2}\varepsilon\Gamma''(\Gamma^{-1}(x))\Phi(\Gamma^{-1}(x))^2$$

and $\zeta(x) = \varepsilon\Phi(\Gamma^{-1}(x))^2\Gamma'(\Gamma^{-1}(x))$. Note that the change of variable from (X, χ_1, χ_2) to (Z_1, Z_2, Z_3) is one-to-one because ψ_t is invertible. Since $\frac{\mu(M_u)}{\Sigma_u^2 S_u^2}$ is bounded on $[0, \tau_L]$ and $H^{N,L} \in L^\infty(\mathscr{F}_{\tau_L})$, there exists a pair $(Z,Y)_{0 \le t \le \tau_L}$ of predictable processes satisfying BSDE 3.6 by Theorem 2 of Briand and Hu [5]. Extend these processes to $[0,T]$ by setting $Y_t = Y_{\tau_L}$ and $Z_t = 0$ for $t \ge \tau_L$.

Define X and χ in terms Z with Equation 3.7. For $m \ge 0$, define $\overline{X}^m = X 1_{\{|X| \le m\}}$ and similarly for $\overline{\chi}^m$. Furthermore, let $\overline{Z}^m$ be given by Equation 3.7 with X and χ replaced by $\overline{X}^m$ and $\overline{\chi}^m$. By Lemma 3.2, there exists a sequence $\{(\overline{X}^{m,n}, \overline{\chi}^{m,n})\}_n$ of bounded continuous processes with finite variation converging to $(\overline{X}^m, \overline{\chi}^m)$ in $\mathscr{H}^2$. Define $\overline{Z}^{m,n}$ in terms of $(\overline{X}^{m,n}, \overline{\chi}^{m,n})$, then $\overline{Z}^{m,n} \to \overline{Z}^m$ in $\mathscr{H}^2$ as $n \to \infty$. Since $\int \overline{Z}^{m,n} dB \to \int \overline{Z}^m dB$, we also have that $\int_t^{\tau_L} |Z_s^{m,n}|^2 ds \to \int_t^{\tau_L} |Z_s^m|^2 ds$ in L^1. Letting

$$\overline{Y}_{\tau_L}^{m,n} = Y_0 - \lambda \int_0^{\tau_L} (\overline{Z}_{1,u}^{m,n})^2 \Lambda_u du + \sum_{i=1}^3 \int_0^{\tau_L} \overline{Z}_{i,u}^{m,n} dB_{i,u} \text{ and}$$

$$\overline{Y}_{\tau_L}^{m} = Y_0 - \lambda \int_0^{\tau_L} (\overline{Z}_{1,u}^{m})^2 \Lambda_u du + \sum_{i=1}^3 \int_0^{\tau_L} \overline{Z}_{i,u}^{m} dB_{i,u},$$

we find $\overline{Y}_{\tau_L}^{m,n} \to \overline{Y}_{\tau_L}^m$ in L^1 as $n \to \infty$. Furthermore, $(\overline{X}^{m,n}, \overline{\chi}^{m,n}, \overline{Y}^{m,n})$ is a s.f.t.s. since it satisfies Equation 3.5 and $[\overline{X}^{m,n}, \overline{X}^{m,n}] = [\overline{\chi}^{m,n}, \overline{\chi}^{m,n}] = 0$.

Since $\overline{Y}_{\tau_L}^m \to H^{N,L}$ as $m \to \infty$, we can find a sequence $(X^{n,L,N}, \chi^{n,L,N}, Y^{n,L,N})_{n \ge 1}$ of s.f.t.s. for each L and N such that $Y_{\tau_L}^{n,L,N} \to H^{N,L} = \mathbf{E}\left(H_T^N | \mathscr{F}_{\tau_L}\right)$ in L^1.

Since $\mathbf{E}\left(H_T^N | \mathscr{F}_{\tau_L}\right) \to H_T^N$ as $L \to \infty$ a.s. by martingale convergence, we also have convergence in L^1 by the Dominated Convergence Theorem. Finally since H_T^N converges to $xh(\widetilde{S}_T^x)$ when N goes to infinity, we can easily find a s.f.t.s. sequence $(X^n, \chi^n, Y^n)_{n \ge 1}$ such that $Y_T^n \to xh(\widetilde{S}_T^x)$ in L^1 as $n \to \infty$. □

The economic interpretation of Theorem 3.4 is that the availability of variance swaps for trading makes the market approximately complete in the sense that any contingent claim with a Lipschitz payoff function can be approximately replicated.

4. Analytical properties of the approximate solutions

In the presence of price impacts, the replicating cost of x units of a contingent claim is not in general x times the replicating cost of 1 unit. When h be a Lipschitz continuous function, recall that for each x an approximating s.f.t.s. for the approximate replication of $xh(\widetilde{S}_T^x)$ is obtained from the solution of BSDE 3.5, which we denote by (X^x, χ^x, Y^x) to emphasize the dependence on x, with the terminal condition $\mathbf{E}\left(xh^N(\widetilde{S}_T^x)\middle|\mathscr{F}_{\tau_L}\right)$ for N and L large. The theorems in this section give analytical properties of these approximate solutions for fixed L and N. To alleviate the notation, we omit the L's and N's in all the expressions in this section (e.g. $\tau=\tau_L, h=h^N$, etc ...) when there is no possible confusion. For each $t\le\tau$ and each $x\in\mathbb{R}$, we define $H_t(x)=\frac{1}{x}Y_t^x$ as the replicating cost per unit for x units of the claim with payoff function h. Furthermore, we let $H_t(0)=\lim_{x\to 0}H_t(x)$. The next theorem states that this limit exists and is given by the solution of the replication problem without trade impacts and liquidity costs of Section 3.1. Recall that $(\hat{X},\hat{\chi},\hat{Y})_{0\le t\le T}$ denotes the solution of the BSDE 3.1 with terminal condition $h(S_T)$.

Theorem 4.1. *$H_t(0)=\hat{Y}_t=\mathbf{E}(h(S_T)|\mathscr{F}_t)$ and $\frac{1}{x}X^x\to\hat{X}$ in $L^2(d\mathbb{Q}\times dt)$ as $x\to 0$.*

Proof. For each x, we let $(Z^x,Y^x)_{0\le t\le\tau_L}$ be the solution of BSDE 3.6 with terminal condition $\mathbf{E}\left(xh(\widetilde{S}_T^x)\middle|\mathscr{F}_\tau\right)$. Using the notation of the proof of Theorem 3.4, we have that Λ_u is bounded on $[0,\tau]$, which means there exists a constant $C>0$ such that $\Lambda_u(Z_{1,u}^x)^2\le C|Z_u^x|^2$. Take $|x|<\frac{1}{4\lambda CC_N}$. First note that since $\left\|\mathbf{E}\left(h(\widetilde{S}_T^x)\middle|\mathscr{F}_\tau\right)\right\|_\infty\le C_N$ we know by the maximum principle (see [16], Proposition 2.1) that $|Y_s^x|\le\ |x|C_N\le\frac{1}{4\lambda C}$ for all $0\le s\le\tau$. Let $H^x=\mathbf{E}\left(h(\widetilde{S}_T^x)\middle|\mathscr{F}_\tau\right)$. In the proof of Theorem 3.4 we have shown that

$$xH^x=Y_t^x-\lambda\int_t^\tau\Lambda_u(Z_{1,u}^x)^2du+\int_t^\tau Z_u^x dB_u,$$

thus

$$\begin{aligned}
x^2(H^x)^2 &= (Y_t^x)^2-2\int_t^\tau\left(\lambda\Lambda_u(Z_{1,u}^x)^2Y_u^x-\frac{1}{2}|Z_u^x|^2\right)du+2\int_t^\tau Y_u^xZ_u^xdB_u\\
&\ge (Y_t^x)^2+\int_t^\tau(1-2\lambda CY_u^x)|Z_u^x|^2du+2\int_t^\tau Y_u^xZ_u^xdB_u\\
&\ge (Y_t^x)^2+\int_t^\tau\frac{1}{2}|Z_u^x|^2du+2\int_t^\tau Y_u^xZ_u^xdB_u.
\end{aligned}$$

We have that $\mathbf{E}(\int_t^\tau\Lambda_u(Z_{1,u}^x)^2du|\mathscr{F}_t)$

$$\le\quad \mathbf{E}(\int_t^\tau C|Z_u^x|^2du|\mathscr{F}_t)\le 2C\mathbf{E}(x^2(H^x)^2|\mathscr{F}_t)\le 2Cx^2C_N^2$$

by taking expectations. Since $Y_t^x=\mathbf{E}\left(xH^x+\int_t^\tau\Lambda_u(Z_{1,u}^x)^2du\middle|\mathscr{F}_t\right)$, we find

$$|\frac{1}{x}Y_t^x-\mathbf{E}(H^x|\mathscr{F}_t)|\quad\le\quad x2CC_N^2. \tag{4.1}$$

Since $h(\widetilde{S}_T^x)\to h(S_T)$ a.s. as $x\to 0$, we have that $\mathbf{E}(H^x|\mathscr{F}_t)=\mathbf{E}(h(\widetilde{S}_T^x)|\mathscr{F}_t)$ converges to $\mathbf{E}(h(S_T)|\mathscr{F}_t)$ a.s. as $x\to 0$ by the Dominated Convergence Theorem. Letting x go to zero in Equation 4.1, we have $H_t(0)=\hat{Y}_t=\mathbf{E}(h(S_T)|\mathscr{F}_t)$.

For the second part of the theorem, let $(\hat{Z},\hat{Y})_{0\le t\le T}$ be the solution of

$$\hat{Y}_t\quad=\quad h(S_T)-\sum_{j=1}^{3}\int_t^T\hat{Z}_{j,s}dB_{j,s}\quad(0\le t\le T).$$

Then $\hat{Z}_{j,s} = \left(\sigma_j \Sigma_s \hat{X}_s S_s + \hat{\chi}_{1,s}\psi_{1,j,s} + \hat{\chi}_{2,s}\psi_{2,j,s}\right)$ for $j = 1,2$ and $\hat{Z}_{1,s} = \sigma_1 \Sigma_s \hat{X}_s S_s$. Moreover, $\mathbf{E}\int_0^\tau |\frac{1}{x}Z_u^x - \hat{Z}_u|^2 du$

$$\begin{aligned} &= \mathbf{E}|\mathbf{E}(h(S_T)|\mathscr{F}_\tau) - H^x|^2 - \left(\mathbf{E}h(S_T) - \frac{1}{x}Y_0^x\right)^2 + 2\lambda\mathbf{E}\int_0^\tau \Lambda_u \frac{1}{x^2}(Z_{1,u}^x)^2(Y_u^x - x\hat{Y}_u)du \\ &\leq \mathbf{E}|h(S_T) - h(\widetilde{S}_T^x)|^2 + \frac{4\lambda CC_N}{x}\mathbf{E}\int_0^\tau |Z_u^x|^2 du \end{aligned}$$

which goes to 0 as $x \to 0$. Recall that $|X_t^x| = \left|\frac{Z_{1,t}^x}{\sigma_1 \Sigma_t S_t}\right| \leq \frac{1}{\sigma_1 L^2}\left|Z_{1,t}^x\right|$ on $[0,\tau]$ and that the same inequality holds for $\hat{X}_t$ and $\hat{Z}_{1,t}$ for any $0 \leq t \leq \tau$. Thus we find that $\frac{1}{x}X^x$ converges to $\hat{X}$ in $L^2(d\mathbb{Q}\times dt)$. $\square$

The next proposition gives an estimate of the error introduced by using $\widetilde{S}^x$ instead of S^0.

Proposition 4.2. *If h is Lipschitz continuous then $\mathbf{E}\left|S_{T+}^0 - \widetilde{S}_T^x\right|^2 = O(x^3)$ as $x \to 0$. In particular, $\mathbf{E}\left|h(S_{T+}^0) - h(\widetilde{S}_T^x)\right|^2 = O(x^3)$ as $x \to 0$.*

Proof. In terms of Z^x, the process S^0 can be decomposed as

$$S_{T+}^0 = S_T + \int_0^\tau \frac{\mu(M_s)}{\sigma_1 \Sigma_s S_s} Z_{1,s}^x ds + \sum_i \int_0^\tau \frac{\phi_i \zeta(M_s)}{\sigma_1 \Sigma_s S_s} Z_{1,s}^x dB_{i,s},$$

since $Z_s^x = 0$ for s outside $[0,\tau]$, whereas

$$\widetilde{S}_T^x = S_T + \int_0^\tau \frac{\mu(M_s)}{\sigma_1 \Sigma_s S_s} x\hat{Z}_{1,s} ds + \sum_i \int_0^\tau \frac{\phi_i \zeta(M_s)}{\sigma_1 \Sigma_s S_s} x\hat{Z}_{1,s} dB_{i,s}.$$

In the proof of the previous theorem, we found

$$\begin{aligned} \mathbf{E}\int_0^\tau |\frac{1}{x}Z_u^x - \hat{Z}_u|^2 du &\leq \mathbf{E}|h(S_T) - h(\widetilde{S}_T^x)|^2 + \frac{4\lambda CC_N}{x}\mathbf{E}\int_0^\tau |Z_u^x|^2 du \\ &\leq 2\lambda x^2 \mathbf{E}\left|\int_0^\tau \hat{X}_u dM_u\right|^2 + 8\lambda CC_N^3 x = O(x) \end{aligned}$$

as $x \to 0$. Then, for some positive constant $\hat{C}$,

$$\begin{aligned} \mathbf{E}\left|S_{T+}^0 - \widetilde{S}_T^x\right|^2 &\leq \hat{C}\mathbf{E}\int_0^\tau \left|x\hat{Z}_u - Z_u^x\right|^2 du \\ &\leq x^2 f(x) \end{aligned}$$

in which $f(x) = O(x)$ as $x \to 0$. $\square$

Under the additional assumption that h is differentiable we have that $H_t(x)$ is also differentiable at $x = 0$ and its derivative can be computed in terms of the solution of the replication problem without trade impacts. The interpretation of $H_t'(0)$ is analogous to the liquidity premium per share M_t of the stock. It gives the additional cost per unit for the replication of the contingent claim due to illiquidity when the number of units replicated is small, i.e. $H_t(x) \approxeq H_t(0) + H_t'(0)x$ when x is small. This is comparable to the price of the stock per share $S_t(x) = S_t(0) + M_t x$.

Proposition 4.3. *Let $0 \le t \le \tau$. If h is differentiable everywhere except at a finite number of points, then $H_t(x)$ is a.s. differentiable at $x = 0$ and*

$$H_t'(0) \;=\; \lambda \mathbf{E}\left(\int_t^\tau \mu(M_s)\hat{X}_s^2 ds\Big|\mathscr{F}_t\right) - 2\lambda \mathbf{E}\left(h'(S_T)\mathbf{1}_{\{S_T\le N\}}(\int_t^\tau \hat{X}_s dM_s)\Big|\mathscr{F}_t\right).$$

Proof. For $x > 0$ small enough, we have that

$$\begin{aligned}
&\Big|\tfrac{1}{x}\left(\tfrac{Y_t^x}{x} - \hat{Y}_t\right) - \lambda \mathbf{E}\left(\textstyle\int_t^\tau \mu(M_s)\hat{X}_s^2 ds\Big|\mathscr{F}_t\right) \\
&\quad +2\lambda \mathbf{E}\left(h'(S_T)\mathbf{1}_{\{S_T\le N\}}(\textstyle\int_t^\tau \hat{X}_s dM_s)\Big|\mathscr{F}_t\right)\Big| \\
&= \left|\frac{1}{x}\left(\frac{Y_t^x}{x} - \hat{Y}_t\right) - \lambda \mathbf{E}\left(\int_t^\tau \Lambda_s \hat{Z}_{1,s}^2 ds\Big|\mathscr{F}_t\right) + 2\lambda \mathbf{E}\left(h'(S_T)\mathbf{1}_{\{S_T\le N\}}(\widetilde{S}_T^x - S_T)\Big|\mathscr{F}_t\right)\right| \\
&\le \frac{2\lambda}{x}\mathbf{E}\left(\left|h^N(\widetilde{S}_T^x) - h^N(S_T) - h'(S_T)\mathbf{1}_{\{S_T\le N\}}(\widetilde{S}_T^x - S_T)\right|\Big|\mathscr{F}_t\right) \\
&\quad +\lambda\left|\mathbf{E}\left(\int_t^\tau \Lambda_s\left(\frac{Z_s^x}{x}\right)^2 ds\Big|\mathscr{F}_t\right) - \mathbf{E}\left(\int_t^\tau \Lambda_s \hat{Z}_s^2 ds\Big|\mathscr{F}_t\right)\right| \\
&\le \frac{2\lambda}{x}\mathbf{E}\left(\left|h^N(\widetilde{S}_T^x) - h^N(S_T) - h'(S_T)\mathbf{1}_{\{S_T\le N\}}(\widetilde{S}_T^x - S_T)\right|\Big|\mathscr{F}_t\right) \\
&\quad +\lambda\left|\mathbf{E}\left(\int_t^\tau \Lambda_s\left(\frac{(Z_s^x)^2}{x^2} - \hat{Z}_s^2\right) ds\Big|\mathscr{F}_t\right)\right|.
\end{aligned}$$

We know that the second term in the last expression goes to zero when $x \to 0$. On the other hand,

$$\begin{aligned}
\lim_{x\to 0}\frac{1}{x}\left(h^N(\widetilde{S}_T^x) - h^N(S_T)\right) &= \lim_{x\to 0}\frac{1}{x}\left(h^N(S_T + x\int_t^\tau \hat{X}_s dM_s) - h^N(S_T)\right) \\
&= h'(S_T)\mathbf{1}_{\{S_T\le N\}}\int_t^\tau \hat{X}_s dM_s \text{ a.s.}
\end{aligned}$$

since h^N is differentiable everywhere except at a finite number of points. Furthermore, note that

$$\frac{1}{x}\left|h^N(\widetilde{S}_T^x) - h^N(S_T)\right| \le \hat{C}\left|\int_t^\tau \hat{X}_s dM_s\right|$$

in which $\hat{C}$ is the Lipschitz constant of h. We then get the result by the Dominated Convergence Theorem. □

5. Conclusion

This paper extends the liquidity risk model of Çetin et al. [6] by hypothesizing the existence of a supply curve that evolves randomly in time and by studying the impact of trades on the supply curve. This leads to a new characterization of self-financing trading strategies and a sufficient condition for no arbitrage. We show the direct connection between stochastic volatility and illiquidity. As a result, contingent claims whose payoffs depend on the value of the asset can be approximately replicated with the use of variance swaps. The replicating costs of such payoffs are obtained from the solutions of BSDEs with quadratic growth. We show that the marginal cost and the liquidity premium of contingent claims can be easily computed from the solution of the replication problem without trade impacts.

6. Acknowledgements

This work was done at the Center for Applied Mathematics at Cornell University and is part of the author's Ph.D. thesis under the guidance of Philip Protter and Robert Jarrow. The author is also very thankful to the two anonymous referees and the associate editor for their constructive comments, and to the participants of the Cornell Liquidity Risk Conference and the First European Summer School in Financial Mathematics, and the Bachelier seminar where the results of this paper were presented.

References

[1] Alfonsi, A., Schied, A., Shulz, A.: Optimal execution strategies in limit order books with general shape functions. Working Paper (2007)

[2] Bank, P., Baum, D.: Hedging and Portfolio Optimization in Financial Markets with a Large Trader. Mathematical Finance **14**(1), 1–18 (2004)

[3] Blais, M.: Liquidity and Modelling the Stochastic Supply Curve for a Stock Price. PhD Thesis, Cornell University (2006)

[4] Blais, M., Protter, P. : The Supply Curve for Stock Prices, working paper (2009)

[5] Briand, P., Hu, Y.: BSDE With Quadratic Growth and Unbounded Terminal Value. Probab. Theory Relat. Fields **136**, 604–618 (2006)

[6] Çetin, U., Jarrow, R., Protter, P.: Liquidity Risk and Arbitrage Pricing Theory. Finance and Stochastics **8**, 311–341 (2004)

[7] Çetin, U., Rogers, L. C. G.: Modelling Liquidity Effects in Discrete Time. Mathematical Finance **17**(1), 15–29 (2007)

[8] Çetin, U., Soner, H. M., Touzi, N.: Options Hedging under Liquidity Costs. Working Paper (2006)

[9] Delbaen, F., Schachermayer, W.: The fundamental theorem of asset pricing for unbounded stochastic processes, Mathematische Annalen **312** 215–250 (1998)

[10] Detemple, J., Osakwe C.: The Valuation of Volatility Options. European Finance Review **4**, 21–50 (2000)

[11] Farmer, J.D., Gillemot, L., Lillo, F., Mike, S., Sen, A.: What really causes large price changes? Quantitative Finance **4**(4), 383–397 (2004)

[12] Frey, R.: Perfect option hedging for a large trader. Finance and Stochastics **2**, 115–141 (1998)

[13] Heston, S. L.: A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options. Rev. Financ. Stud. **6**, 327–343 (1993)

[14] Hull, J., White, A.: The Pricing of Options on Assets with Stochastic Volatilities. The Journal of Finance **42**(2), 281–300 (1987)

[15] Jarrow, R.: Derivative Security Markets, Market Manipulation, and Option Pricing Theory. Journal of Finance and Quantitative Analysis **29**(2), 241–261 (1994)

[16] Kobylanski, M.: Backward Stochastic Differential Equations and Partial Differential Equations with Quadratic Growth. Annals of Probability **28**(2), 558-602 (2000)

[17] Ly Vath, V., Mnif, M., Pham, H.: A model of optimal portfolio selection under liquidity risk and price impact. Finance and Stochastics **11**, 51–90 (2007)

[18] Pardoux, E., Peng, S. G.: Adapted solution of a backward stochastic differential equation. Systems & Control Letters **14**, 55–61 (1990)

[19] Protter, P.: Stochastic Integration and Differential Equations. Springer Verlag (2004)

[20] Rogers, L. C. G., Singh S.: The Cost of Illiquidity and its Effects on Hedging. Working Paper (2007)

[21] Weber, P., Rosenow, B.: Order book approach to price impact. Quantitative Finance **5**(4), 357–364 (2005)

[22] Weber, P., Rosenow, B.: Large stock price changes: volume or liquidity. Quantitative Finance **6**(1), 7–14 (2006)

ETH Zürich, Departement Mathematik, 8092 Zürich, Switzerland, alexandre.f.roch@gmail.com