A Risk-Neutral Parametric Liquidity Model for Derivatives

David Bakstein* Sam Howison†

Oxford Centre for Industrial and Applied Mathematics,
Mathematical Institute, 24–29 St. Giles', Oxford OX1 3LB, UK.

Abstract

We develop a parameterised model for liquidity effects arising from the trading in an asset. Liquidity is defined as the combination of an individual transaction cost and a price slippage impact, which is felt by all participants in the market. The chosen definition allows liquidity to be observable in a centralised order-book of an asset as is usually provided in most non-specialist exchanges. The discrete-time version of the model is based on the CRR binomial tree and in the appropriate continuous-time limits we derive various nonlinear partial differential equations. Both versions can be directly applied to the pricing and hedging of options, thereby, due to the nonlinear nature of liquidity, deriving natural bid-ask spreads that are based on the liquidity of the market for the underlying and the existence of super-replication strategies. We test and calibrate our model set-up empirically with high-frequency data of German blue chips and discuss further extensions to the model as well as applications like liquidity derivatives and portfolio trading.

Keywords: liquidity, option pricing, liquidity derivatives, portfolio trading

1 Introduction

Two of the underlying assumptions of, amongst others, the basic Black-Scholes or CAPM economies are firstly that markets are frictionless, and secondly that agents are price-takers, i.e. no single participant can affect asset prices in the market through her trading strategies. But real world markets substantially deviate from these assumptions, because for virtually all traded assets there exist both bid-ask spreads and a limited market depth. In general, the former serve as a revenue source as well as a risk insurance buffer for market makers, since they will buy low and sell high, and the latter is the volume of an asset available to buy or sell at a certain price. Together they represent the order book of an asset, which serves as market inventory allowing immediate execution. Usually, if many market makers and participants want to trade, bid-ask spreads tend to be narrow and market depth plentiful because of competition. Colloquially, the market is then said to be liquid.[1] Obviously market participants prefer markets or assets with high liquidity, because they

*contact: bakstein@maths.ox.ac.uk; work supported by the EPSRC, Charterhouse, Socrates and ESF

†contact: howison@maths.ox.ac.uk

[1]Other factors that might affect bid-ask spreads are the availability of information about the traded asset or the legislation of the market itself.

can get in and out of their positions quickly and cheaply. Hence, say, the equity of a large company that has a high free float and trades on a big exchange will, ceteris paribus, have narrower spreads, compared to a closely held small-cap.[2] The same will also apply to contingent claims written on that asset. Because the value of a derivative both originally in the Black-Scholes theory and in reality at least partly is derived from replicating trading strategies in its underlying, the contract's bid-ask spreads will be narrower, the more liquid the market for the hedging instrument. Generally, however, there is neither a consensus approach how to calculate liquidity premia of derivatives nor how to parameterise and measure the liquidity of a market or asset.

Qualitatively, liquidity or the lack of it causes two effects. Firstly, it has an impact on the transaction price. Whereas it may be possible to trade small quantities of an asset at the best possible price, which is close to the published mid-price, the larger the trade size the more levels of market depth (from one or more market makers) will have to be tapped and the further the average transaction price will deviate from the mid-price. Thus, in general, the average transaction price will be an increasing function of trade size. Secondly, liquidity is directly related to the degree of market slippage due to individual transactions. This means that, since every participant can observe the same market depth,[3] large trades of one agent may remove entire price layers and lead market makers to adjust their prices accordingly. In reality, it is common that asset prices are pushed, in some cases deliberately, in a certain direction by comparatively large trades (see e.g. [Tal]). But even if no trader has an explicit intention to do so some agents have to trade certain quantities of the underlying to execute a large portfolio trade or to hedge the exposure of their portfolios of derivatives. In the latter case if, as in the Black-Scholes theory, they try and Delta-hedge, then for options with non-smooth or even-discontinuous payoffs, the Delta and Gamma, i.e. the amount of the underlying they have to hold and approximately add or remove, respectively, become large close to expiry or close to a payoff discontinuity. Since, in reality, markets only have limited liquidity, these traders may thus move the value of the underlying in an undesired direction because the trade-induced slippage feeds back into their mark-to-market contract values. To avoid mis-hedging, the required quantities of the underlying thus have to be adjusted by a liquidity factor. This will affect the value of the position, since the latter is derived, by the Black-Scholes framework, from the risk-free amount that can be earned on a replicating portfolio. Beyond this if a trader has a good intuition of the liquidity of the market, then, instead of hedging a position, she may be inclined to liquidate the accumulated hedge quantity and thus push the market in a desired direction. However, normally, traders are not supposed to know the positions of other participants in the market, yet, if it became known or if a trader acted on behalf of a client on the one hand and had a proprietary book on the other, then she might exploit his knowledge. Taking this possibility into account, the initial premium required for a contingent claim from counterparties could be reduced due to this informational asymmetry.

To capture the various effects of liquidity analytically the papers of [Jar1], [Schö], [Frey], [A&C1] and [H&S1] propose a number of models with similar components. Firstly, they introduce a reaction function that models the immediate impact of a trade and the average price paid per asset. It is also a function of both a liquidity scaling parameter and the trade size. A possible proxy for the former is explicitly given by [Krak] as the ratio of change in the price of the underlying asset to notional traded. This choice of estimator has the advantage that at the time of the trade the liquidity parameter is observable and predictable. Secondly, the papers by [A&C1], [H&S1] and [H&S2], further consider the permanent slippage effect on the asset, by making its new equilibrium

[2]In this case the more liquid stock should also trade at a fundamental premium. But in this paper we will consider equity valuation as an exogeneous factor.

[3]As is the case in virtually all European exchanges and also e.g. NASDAQ, we assume that all participants either deal in a centralised transparent order-driven or electronically-linked broker market, instead of a specialists' market as is e.g. the NYSE.

price a function of both the previous and the average transaction price. In our paper we employ a combination of these effects, which will make them observable given a particular order-book. In section 2 we derive the discrete-time version of the model, which is a combination of the binomial model of [CRR] and a nonlinear controlled process, and apply it to the valuation of options. In section 3 we derive various nonlinear partial differential equations (PDEs) in the continuous-time limit under special choices of parameters. In section 4 we empirically analyse a trading book for various stocks traded on the German Xetra system. We present a consistent definition of liquidity and calibrate our model to the data. In section 5 we mention various extensions to the model and present a number of applications including liquidity derivatives and portfolio trading. Section 6 summarises the paper and suggests further areas of research around our model.

2 The discrete-time model

The main building block for the pricing framework of derivatives and portfolio trades is a suitable model for the underlying assets or, more generally, the state variables. We commence our analysis in a discrete-time finite-horizon economy where trading in assets takes place at times $\{t_0, t_1, \ldots, t_n = T\}$. The state of the economy is given by the finite set of trajectories $\Omega = \{\omega_1, \ldots, \omega_m\}$ and the revelation of the true state by the filtration, i.e. increasing sequence of algebras, $(\mathcal{F}_t)_{t\in\{t_0,\ldots,T\}}$. The initial set of states is $\mathcal{F}_{t_0} = \Omega$, the eventual true state of the economy is revealed as $\mathcal{F}_T = \omega_j$, $\omega_j \in \Omega$. There are two assets, namely a risky "stock" $S_t(\omega)$ and a riskless "bond" B_t, whose respective processes are adapted to the filtration $(\mathcal{F}_t)_{t\in\{t_0,\ldots,T\}}$ and valued in $\mathbb{R}_+$.

Resorting to the widely used binomial model of [CRR], we will model randomness, which represents other participants' trading in the stock, by making the risky asset go up by a fraction $u-1$ with probability q or down by a fraction $1-d$ with probability $1-q$ over one timestep. Therefore

$$S_{t_i}(\omega_j) = \begin{cases} uS_{t_{i-1}} & \text{if } \omega_j = \omega_u \text{ implies up-step,} \\ dS_{t_{i-1}} & \text{if } \omega_j = \omega_d \text{ implies down-step,} \end{cases} \tag{1}$$

where $u > d$. The bond on the other hand will always yield the riskless return r, namely

$$B_{t_{i+1}} = (1+r)B_{t_i}. \tag{2}$$

The two key properties of the model are, firstly, the absence of arbitrage in the market provided that $0 < d < 1+r < u$. This implies that expectations of the discounted risky asset have to be taken with respect to the risk-neutral probabilities q and form a martingale when taking the bond as numeraire:

$$E_Q[S_T] = S_{t_0} B_T / B_{t_0}, \qquad T > t_0. \tag{3}$$

Secondly, u, d and q can be calibrated such that S_t follows (risk-neutral) geometric Brownian motion

$$dS_t = rS_t dt + \sigma S_t dW_t, \tag{4}$$

in the continuous-time limit. Here σ is the asset's volatility and dW_t a Wiener process, i.e. the increments of standard Brownian motion. The same model is employed in the seminal paper by [B&S] as the model for the underlying asset.

On top of the random process for the underlying we construct a controlled process that represents the effect that a large or influential trader has on the market. This trader's holding process in the risky asset we denote by $(H_t(\omega))_{\forall t,\omega}$ and in the bond by $(\hat{H}_t(\omega))_{\forall t,\omega}$. Both processes are adapted to the filtration $(\mathcal{F}_t)_{t\in\{t_0,\ldots,T\}}$ and moreover one-step-ahead predictable with respect to it. The

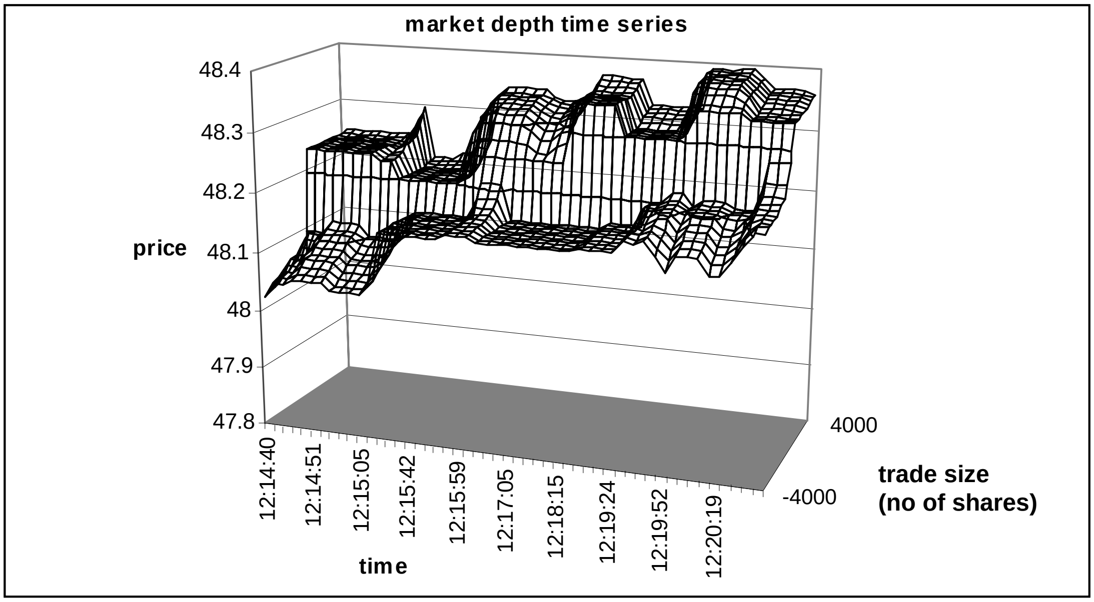


Figure 1: Average transaction price time series for BASF stock in a particular trading period derived from the electronic order book. At the beginning of the time period shown, liquidity is high on the ask side (the surface is nearly flat, so that 1 to 4000 stocks can be bought at approximately the same unit price) and less so on the bid side. The mid-price is the average of the values at the top and bottom of the 'cliff'.

latter point entails that the trader's portfolio can be rebalanced in between the random changes to the underlying asset. Now, if we assume that S_t represents the mid-market price at a generic time, then the most favourable prices to sell or buy the asset, i.e. the bid and the ask, will be below and above it, respectively. Also, if the quantity to be traded is large, then more than one quote has to be filled in order to complete the trade. This means that the average transaction price $\bar{S}_t$ is an increasing function of the trade size. We define its process $(\bar{S}_t(\omega))_{\forall t,\omega}$ as a function $f(S_{t_{i-1}}, H_{t_i} - H_{t_{i-1}}, \gamma, \lambda)$ of the observable spot $(S_t)_{\forall t,\omega}$, trade size $(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))_{\forall t,\omega}$ and liquidity parameters $\gamma, \lambda \geq 0$, where the former is a proxy for the width of the spread and the latter for the market depth. Intuitively, the reaction or price-impact function, in addition to being increasing with respect to the trade size, should have the properties that

$$\lim_{H_{t_i} - H_{t_{i-1}} \downarrow -\infty} f = 0, \qquad \lim_{H_{t_i} - H_{t_{i-1}} \uparrow \infty} f = \infty, \qquad f(S_{t_{i-1}}, 0, \gamma, \lambda) = S_{t_{i-1}}$$

and that

$$\lim_{H_{t_i} - H_{t_{i-1}} \downarrow \epsilon > 0} f > S_{t_i} + \epsilon, \qquad \lim_{H_{t_i} - H_{t_{i-1}} \uparrow -\epsilon < 0} f < S_{t_{i-1}} - \epsilon, \qquad \epsilon \text{ small.}$$

The first set of properties reflects the intuition that large sell or buy orders push the market down and up, respectively, while if no trading takes place the spot remains unchanged. The second set states that even if the traded quantity is small, there exists a positive bid-ask spread around the mid-market price. The latter is usually the one quoted in various information sources and its (two-dimensional) time series is employed in most financial applications like e.g. performance ratios or technical trading rules. But as figure 1 shows, when including a price impact function the true average transaction price time series has a dependence on the traded quantity and is thus three-

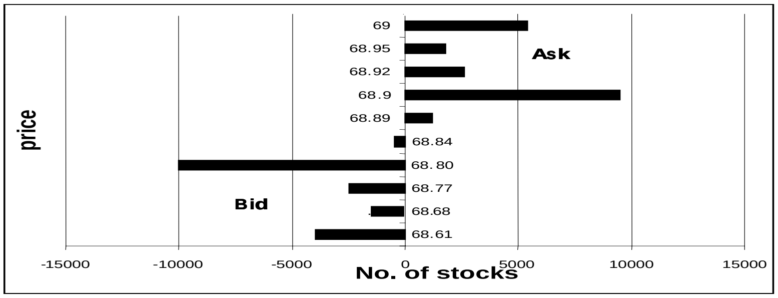


Figure 2: An example of an order book of Deutsche Telekom stock on 13 Jan 2000 at 12:14:41pm.

dimensional.[4] This reflects the fact that the mid-market price doesn't exist, i.e. nobody can transact at it.

One possible form of f that does capture the properties of the third dimension, as partly noted in [Jar1] and [Frey], is

$$\bar{S}_{t_{i-1}} = S_{t_{i-1}} \left(1 + \mathrm{sign}(H_{t_i} - H_{t_{i-1}})\gamma\right) e^{\lambda(H_{t_i} - H_{t_{i-1}})}, \tag{5}$$

where we suppressed the explicit dependence on the trajectory ω. In (5) the sign$(\cdot)$ models the bid-ask spreads and the exp$(\cdot)$ term the market depth, which represents the elasticity of a price to the quantity traded. Under this model the total cash flow and implicit transaction costs over a timestep are given by $(H_{t_{i-1}} - H_{t_i})\bar{S}_{t_{i-1}}$ and $(H_{t_{i-1}} - H_{t_i})(\bar{S}_{t_{i-1}} - S_{t_{i-1}})$, respectively.

But in addition to the pure transaction cost effect, there is a market impact effect that is felt by all participants. If the size of the trade was large the best quotes have been removed from the order-book, which, in a centralised order-driven exchange, effectively represents the market.[5] Thus some layers are no longer available to any of the other market participants and the latter will adjust their new quotes accordingly. Hence, in effect, the market has been moved. Depending whether the transaction is a buy or a sell, the average transaction price is below or above the last price traded, unless only one level of market depth was filled. In any case the market impact, i.e. the post-trade new asset price that was last traded, is directly observable given an order-book. Mathematically a convenient model for this effect is to make the new equilibrium log-price a combination of the two previous equilibrium and average transaction log-prices, namely their geometric average. Adding this permanent effect to the temporary reaction (5) we obtain the price dynamics

$$S_{t_{i-1}} \qquad \text{(mid-market price)} \tag{6}$$

$$\rightarrow \quad \bar{S}_{t_{i-1}} = S_{t_{i-1}} \left(1 + \mathrm{sign}(H_{t_i} - H_{t_{i-1}})\gamma\right) e^{\lambda(H_{t_i} - H_{t_{i-1}})} \qquad \text{(average transaction price)} \tag{7}$$

$$\rightarrow \quad S^{\alpha}_{t_{i-1}} \bar{S}^{1-\alpha}_{t_{i-1}} = S_{t_{i-1}} \left(1 + \mathrm{sign}(H_{t_i} - H_{t_{i-1}})\gamma\right)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_i} - H_{t_{i-1}})} \qquad \text{(price slippage)} \tag{8}$$

$$\rightarrow \quad \begin{cases} uS_{t_{i-1}} \left(1 + \mathrm{sign}(H_{t_i} - H_{t_{i-1}})\gamma\right)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_i} - H_{t_{i-1}})} = S_{t_i}(\omega_u) \\ dS_{t_{i-1}} \left(1 + \mathrm{sign}(H_{t_i} - H_{t_{i-1}})\gamma\right)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_i} - H_{t_{i-1}})} = S_{t_i}(\omega_d) \end{cases} \qquad \text{(random change)} \tag{9}$$

One interpretation of the new parameter α is given in the papers of [A&C1], [A&C2], [H&S1] and [H&S2], which also model a permanent price update effect that is a function of both the previous equilibrium S_{t_i} and the average transaction price $\bar{S}_{t_i}$, given as a convex combinations, i.e. $0 \leq \alpha \leq 1$.

[4] If liquidity were perfect, the third dimension would be flat.

[5] Usually there is also a broker market on top of the order-book, but most brokers should will a dual presence.

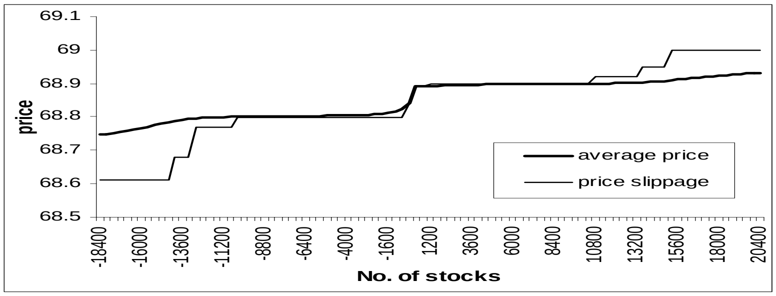


Figure 3: The exact average transaction price and resulting price slippage (last price traded) as a function of no. of stocks traded for the order book of figure 2. The flat parts of the price slippage curve represent the price layers of the order book.

Their explanation of α is that large trades may not contain fundamental new information and hence push the market to an untenable price level. The latter adjusts itself immediately as the order-book is refilled with updated quotes. However this effect is rather intuitive and not directly observable prior to a trade. At the limits, if $\alpha = 0$ then the new equilibrium price will be exactly the last average transaction price, corresponding to the case that only one layer of market depth was tapped. For $\alpha = 1$, since there is no subsequent manipulation effect, we have a pure transaction costs model similar to those of [B&V], [BLPS] and [ENU]. This would imply that other market participants didn't believe that the trade bore new fundamental information.

In our interpretation of α as an observable slippage parameter we have a choice, whether (8) should represent the new last price traded or the new mid-market price derived therefrom. In the former case α would be non-positive because we assume that, in general, the best quotes are filled first. Moreover it would directly give the last observed price traded, which may be important in contingent claim contracts, especially in barrier type contracts (see e.g. [Tal]). Speaking for the latter interpretation is the fact that because we resort to the mid-price as the reference point of the controlled process it may be more consistent to return to it. In this case α may be positive. In any case, it will only make a difference in the empirical estimation of the parameters. Figure 2 shows an example of an order-book and figure 3 the exact average transaction price derived from it as well as the new last slippage price as a function of trade size.

Our model set-up, albeit structured similarly, is different from those of [A&C1], [A&C2], [H&S1] and [H&S2], who resort to arithmetic Brownian motion

$$dS_t = \mu dt + \sigma dW_t, \tag{10}$$

as the process for the underlying. In their respective papers, it is an acceptable und computationally convenient model regarding portfolio trading applications. But it may cause serious concerns when it is applied to the pricing of derivatives. Mainly this is due to the fact that the spot of the underlying may become negative with positive probability, whereas with geometric Brownian motion this is impossible. Another reason for their choice may have been the symmetry of up and down movements of the spot, but as long as λ or the quantity traded are small, our reaction function (5) will also be locally linear. Moreover, as derived in the next subsection, the exponential form of the reaction function will make the resulting tree Markovian, whereas it would be path-dependent for a linear model. Lastly, our model is also free of arbitrage opportunities, as we now demonstrate.

Proposition 2.1 (Non-existence of arbitrage opportunities)

If the risky and riskless assets $(S_t(\omega), B_t)_{\forall t,\omega}$ follow the processes (6)-(9) and (2) respectively, there does not exist a particular holding strategy $(H'_t(\omega), \hat{H}'_t(\omega))_{\forall t,\omega}$, with $H'_{t_0} = H'_T(\omega)$, $\forall\omega$, which is self-financing and value-conserving, i.e.

$$\left(\hat{H}'_{t_i}(\omega) - \hat{H}'_{t_{i-1}}(\omega)\right) B_{t_{i-1}} + \left(H'_{t_i}(\omega) - H'_{t_{i-1}}(\omega)\right) \bar{S}_{t_{i-1}}(\omega) = 0, \qquad \forall t, \omega \tag{11}$$

and results in a positive expected gain

$$E[V_T - V_{t_0} | \mathcal{F}_{t_0}] > 0, \tag{12}$$

where $V_{t_i}(\omega) = H_{t_i}(\omega) S_{t_i}(\omega) + \hat{H}_{t_i}(\omega) B_{t_i}$, $\forall t, \omega$ is the mark to market value of the portfolio.

Proof: For simplicity and without loss of much generality we assume that $r = 0$, $B_{t_0} = 1$, $\alpha = 0$ and $\hat{H}_{t_0} = H_{t_0} = H_T = 0$. Thus, since $V_{t_0} = 0$, (12) reduces to

$$E[V_T | \mathcal{F}_{t_0}] = E[\hat{H}_T | \mathcal{F}_{t_0}],$$

which after repeated substitution of (11) gives

$$\begin{aligned} E\left[\hat{H}_{t_{n-1}} - (H_T - H_{t_{n-1}})\bar{S}_{t_{n-1}} \middle| \mathcal{F}_{t_0}\right] &= E\left[-\sum_{i=1}^{n} (H_{t_i} - H_{t_{i-1}}) \bar{S}_{t_{i-1}} \middle| \mathcal{F}_{t_0}\right] \\ &= E\left[-\sum_{i=1}^{n} \delta H_i S_{t_{i-1}} (1 + \mathrm{sign}(\delta H_i)\gamma) e^{\lambda \delta H_i} \middle| \mathcal{F}_{t_0}\right], \end{aligned} \tag{13}$$

where we define the operator $\delta H_i = H_{t_i} - H_{t_{i-1}}$, $i = 1 \ldots n$. Now, from (3), it is apparent that in the absence of trading S_{t_i} has to have the martingale property, thus

$$E_Q[S_{t_i} | \mathcal{F}_{t_0}] = S_{t_0}, \qquad \forall t_i. \tag{14}$$

Substituting (14) into (13) and separating the buy orders from the sell orders we obtain

$$-S_{t_0} \left(\sum_j \delta H_j (1 + \gamma) e^{\lambda \delta H_j} + \sum_l \delta H_l (1 - \gamma) e^{\lambda \delta H_l} \right) \leq 0, \tag{15}$$

where $\delta H_j \geq 0$, $\forall j$ and $\delta H_l < 0$, $\forall l$. Inequality (15) follows from the fact that $H_{t_0} = H_T$, therefore $\sum_j \delta H_j = -\sum_l \delta H_l$, but $\exp(\lambda \delta H_j) \geq 1$, $\forall j$ and $\exp(\lambda \delta H_l) < 1$, $\forall l$. This means that due to the increasing convexity of the exponential function a buy order will move an asset price up more, in relative terms, than a sell order will move it down. Thus transaction costs dominate the market manipulation effect and there are no arbitrage opportunities in trading the underlying.□

[Jar1] provides a proof of non-arbitrage for a more general class of reaction functions and also incorporates traded options into the market. In our model we ignore other traded options and the market impact on and due to them. The reason is, firstly, that we do not intend to employ them as a hedging instruments for other derivatives; secondly, that we assume that the large trader is supervised by a market authority and hence requires a valid economic reason to place vast orders in the market; and thirdly, that traded options themselves have finite liquidity, usually much lower than that of the underlying. The primary application of the model as presented in this paper is the hedging of over-the-counter (OTC) derivatives positions as well as trading strategies in the underlying, but not the manipulation of the listed derivatives market.

2.1 The hedging and pricing of contingent claims in discrete time

A contingent claim $C_t(\omega)$ is a time-dependent generic function of the values of the underlying assets in the economy. Depending on the structure of C over its lifetime, it requires the writer to exchange certain amounts $(H_t(\omega), \hat{H}_t(\omega))_{\forall t,\omega}$ with the holder at particular times. In general, contingent claims are valued in reference to the setup cost $V_{t_0}^*$ of a self-financing portfolio strategy in the underlying risky and riskless assets. This hedging strategy $(H_t^*(\omega), \hat{H}_t^*(\omega))_{\forall t,\omega}$, subject to an initial holding $H_{t_0}^* = H_0$, will exactly replicate or super-replicate any payoffs of the claim $C_t(\omega), \forall t, \omega$. Moreover, under the optimal hedging strategy $V_{t_0}^*$ is at its minimum.

As a special class of contingent claims European vanilla options have a single expiry time T. Upon the latter there exist different methods of how a European contract is settled. The settlement can either be at the writer's discretion, i.e. that the holder has to accept exchanging any combination of assets

$$H_T(\omega)S_T(\omega) + \hat{H}_T(\omega)B_T = C_T(\omega), \qquad \forall\omega, \tag{16}$$

or another possibility is physical delivery, so that at expiry $(H_T(\omega), \hat{H}_T(\omega))_T$ is fixed for every ω. Finally, under cash settlement at expiry we have $\hat{H}_T(\omega)B_T = C_T(\omega)$, $\forall\omega$. It becomes apparent that physical and cash deliveries are subsets of discretionary settlement, hence the latter is the least restrictive and may thus lead to a lower initial setup cost. Provided that delivery is at the writer's discretion, in a discrete-time economy the valuation of European vanilla type contingent claims under finite liquidity can be formulated as the following nonlinear program:

Proposition 2.2 (Replication ask price of a contingent claim)
The ask price of a contingent claim $C_t(\omega)$ at time t_0 is

$$C_{t_0} = \max(V_{t_0}^*, 0), \tag{17}$$

where

$$V_{t_0}^* = \min_{(H_t(\omega),\hat{H}_t(\omega))_{\forall t,\omega}} V_{t_0} = (H_{t_1}^* - H_0)\bar{S}_{t_0} + \hat{H}_{t_1}^* B_{t_0} + H_0 S_{t_0}, \tag{18}$$

and the optimal controls $(H_t^*(\omega), \hat{H}_t^*(\omega))_{\forall t,\omega}$ satisfy the initial holding

$$H_{t_0} = H_0, \tag{19}$$

the self-financing condition

$$(\hat{H}_{t_i}(\omega) - \hat{H}_{t_{i-1}}(\omega))B_{t_{i-1}} + (H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\bar{S}_{t_{i-1}}(\omega) = 0, \tag{20}$$

the payoff replication constraint

$$V_T(\omega) = H_T(\omega)S_T(\omega) + \hat{H}_T(\omega)B_T = C_T(\omega), \tag{21}$$

and the processes $(S_t(\omega), B_t)_{\forall t,\omega}$ are given by (6)-(9) and (2), respectively.

The solution $V_{t_0}^*$ of (18) represents the minimum amount of funds required for the writer to engage in a strategy

$$(H_t^*(\omega), \hat{H}_t^*(\omega))_{\forall t,\omega} = \arg \min_{(H_t(\omega),\hat{H}_t(\omega))_{\forall t,\omega}} V_{t_0}$$

that will allow her to meet $C_T(\omega), \forall\omega \in \Omega$. Thereby the last term of (18) represents the mark-to-market value of the initial quantity of the risky asset that is employed for hedging the claim.

Moreover (17) adds a natural lower bound for the claim as the writer would not pay for a position that doesn't offer him the chance of positive returns. Conversely, the bid price $\tilde{C}_{t_0}$, represents the amount of funds the trader would be willing to pay in order to be the receiver of the payoff. This is equivalent to the amount she could borrow against the contract as a collateral.

Corollary 2.1 (Replication bid price of a contingent claim)
The bid price of the contract is the solution of the program

$$\tilde{C}_{t_0} = \max(\tilde{V}^*_{t_0}, 0),$$

where

$$\tilde{V}^*_{t_0} = - \max_{(H_t(\omega), \hat{H}_t(\omega))_{\forall t,\omega}} V_{t_0} = (H^*_{t_1} - H_0)\bar{S}_{t_0} + \hat{H}^*_{t_1} B_{t_0} + H_0 S_{t_0},$$

subject to the initial holding (19) and self-financing condition (20), along with the replication constraint

$$V_T(\omega) = H_T(\omega) S_T(\omega) + \hat{H}_T(\omega) B_T = -C_T(\omega),$$

with the asset dynamics as in proposition (2.2).

We can show that $C_{t_0} \geq \tilde{C}_{t_0}$ i.e. that the ask price for a particular position must always be at least as big as the bid price, with equality when the market is perfectly liquid. For the proof we require the following result:

Lemma 2.1 (Non-homogeneity of prices)
Denote by $C^{(i)}_{t_0}$ and $(H^{(i)}_t(\omega), \hat{H}^{(i)}_t(\omega))_{\forall t,\omega}$, $i = 1, \ldots, 3$ the optimal values and holding strategies for the program given in proposition 2.2 with the right hand side of (21) replaced by $C^{(1)}_T(\omega) = \eta C_T(\omega)$, $C^{(2)}_T(\omega) = \zeta C_T(\omega)$ and $C^{(3)}_T(\omega) = (\eta + \zeta) C_T(\omega)$, where η, ζ are constants. Then,

1. *if $\eta, \zeta \geq 0$ or $\eta, \zeta \leq 0$, we have $|C^{(1)}_{t_0} + C^{(2)}_{t_0}| \leq |C^{(3)}_{t_0}|$,*
2. *if $\eta \geq 0 \geq \zeta$ or $\zeta \geq 0 \geq \eta$, we have $|C^{(1)}_{t_0} + C^{(2)}_{t_0}| \geq |C^{(3)}_{t_0}|$.*

Proof: The lemma follows directly from the facts that for the respective η and ζ given the form of the processes S_t and $\bar{S}_t$, the following hold:

1. At expiry

$$\left| H^{(3)}_T(\omega) \right| \geq \left| H^{(1)}_T(\omega) + H^{(2)}_T(\omega) \right|, \qquad \forall \omega;$$

thus

$$\left| C^{(1)}_T(\omega) + C^{(2)}_T(\omega) \right| \leq \left| C^{(3)}_T(\omega) \right|, \qquad \forall \omega,$$

and so, given the self-financing condition,

$$\begin{aligned} & \left(H^{(1)}_{t_i}(\omega) - H^{(1)}_{t_{i-1}}(\omega) \right) e^{\lambda \left(H^{(1)}_{t_i}(\omega) - H^{(1)}_{t_{i-1}}(\omega) \right)} + \left(H^{(2)}_{t_i}(\omega) - H^{(2)}_{t_{i-1}}(\omega) \right) e^{\lambda \left(H^{(2)}_{t_i}(\omega) - H^{(2)}_{t_{i-1}}(\omega) \right)} \\ \leq \ & \left(H^{(3)}_{t_i}(\omega) - H^{(3)}_{t_{i-1}}(\omega) \right) e^{\lambda \left(H^{(3)}_{t_i}(\omega) - H^{(3)}_{t_{i-1}}(\omega) \right)}, \qquad \forall t, \omega. \end{aligned}$$

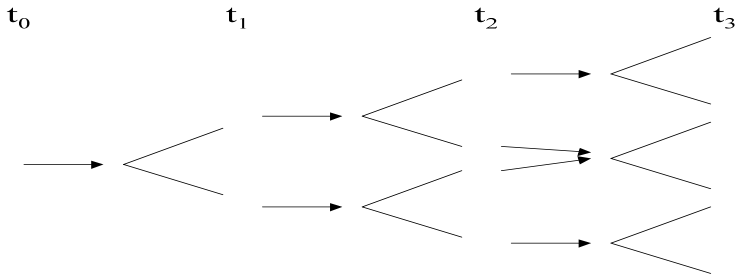


Figure 4: Asset tree, where lines represent random change and arrows slippage.

2. The converse of 1.□

By choosing $\eta = 1$ and $\zeta = -1$ we observe that $C_{t_0}^{(1)} = C_{t_0}$, $C_{t_0}^{(2)} = -\tilde{C}_{t_0}$ and $C_{t_0}^{(3)} = 0$, so that $C_{t_0} \geq \tilde{C}_{t_0}$. This is due to the fact that the trader will have to buy high and sell low, every time she rehedges. Hence the liquidity of the market gives the natural bid-ask spreads in the underlying. Also, the larger the absolute payoff $|C_T|$, the wider the relative spreads, since the reaction function is super-linear. Thus the price of x options will be greater than x times the price of one option. Again, intuitively a trader would like to put small orders into the market to avoid transaction costs.

Because, in general, $S_t(\omega)$ and $\bar{S}_t(\omega)$ are functions of $(H_t(\omega), \hat{H}_t(\omega))_{\forall t,\omega}$, i.e. the present and past stock-holdings, the problem in proposition 2.2 is path-dependent and the number of variables as well as constraints is growing exponentially as the number of time steps increases. The following example demonstrates the growth of distinct points in state space.

Example 2.1 We consider the three period economy with the trading times $\{t_0, t_1, t_2, t_3\}$, the set of states $\Omega = \{\omega_{uuu}, \omega_{uud}, \ldots, \omega_{ddd}\}$ and the information revelation[6] $\mathcal{F}_{t_0} = \{\Omega\}$, $\mathcal{F}_{t_1} = \{\omega_u = \{\omega_{uuu}, \ldots, \omega_{udd}\}, \omega_d\}$, $\mathcal{F}_{t_2} = \{\omega_{uu}, \ldots, \omega_{dd}\}$ and $\mathcal{F}_{t_3} = \{\{\omega_{uuu}\}, \ldots, \{\omega_{ddd}\}\}$. Then the asset's dynamics are

$$
\begin{aligned}
t_0 :&\quad S_{t_0} \to S_{t_0}(1+\mathrm{sign}(H_{t_1} - H_0)\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_1} - H_0)} \\
t_1 :&\quad \begin{cases} uS_{t_0}(1+\mathrm{sign}(H_{t_1} - H_0)\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_1} - H_0)} \\ dS_{t_0}(1+\mathrm{sign}(H_{t_1} - H_0)\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_1} - H_0)} \end{cases} \\
&\quad \to uS_{t_0}\Pi_{i=1}^{2}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}})\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_2}(\omega_u) - H_0)} \\
&\quad \to dS_{t_0}\Pi_{i=1}^{2}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}})\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_2}(\omega_d) - H_0)} \\
t_2 :&\quad \begin{cases} u^2S_{t_0}\Pi_{i=1}^{2}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}})\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_2}(\omega_u) - H_0)} \\ udS_{t_0}\Pi_{i=1}^{2}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}})\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_2}(\omega_u) - H_0)} \\ duS_{t_0}\Pi_{i=1}^{2}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}})\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_2}(\omega_d) - H_0)} \\ d^2S_{t_0}\Pi_{i=1}^{2}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}})\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_1}(\omega_d) - H_0)} \end{cases} \\
&\quad \to u^2S_{t_0}\Pi_{i=1}^{3}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{uu}) - H_0)} \\
&\quad \to udS_{t_0}\Pi_{i=1}^{3}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{ud}) - H_0)} \\
&\quad \to duS_{t_0}\Pi_{i=1}^{3}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{du}) - H_0)} \\
&\quad \to d^2S_{t_0}\Pi_{i=1}^{3}(1+\mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{uu}) - H_0)}
\end{aligned}
$$

[6] Strictly speaking $\mathcal{F}_t$ is the symbol for the σ-algebra of a given partition at every t_i, i.e. all the unions and complements of its elements. Nonetheless we abuse the notation.

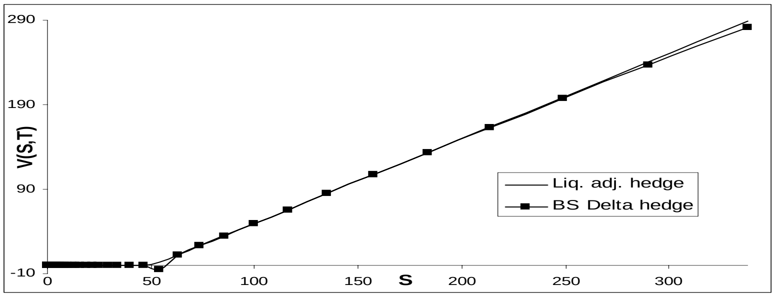


Figure 5: Payoff replication of a Call option when using the Black-Scholes Delta in comparison to the liquidity adjusted hedge. Here $T = 1$, $S_{t_0} = K = 50$, $\sigma = 0.2$, $r = 0.05$, $\gamma = 0.01$, $\alpha = 1$, $\lambda = 0.01$, $H_0 = 0$ and 50 timesteps.

$$t_3 : \left\{ \begin{array}{l} u^3 S_{t_0} \Pi_{i=1}^3 (1 + \mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{uu}) - H_0)} \\ u^2 d S_{t_0} \Pi_{i=1}^3 (1 + \mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{uu}) - H_0)} \\ u^2 d S_{t_0} \Pi_{i=1}^3 (1 + \mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{ud}) - H_0)} \\ u d^2 S_{t_0} \Pi_{i=1}^3 (1 + \mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{ud}) - H_0)} \\ u^2 d S_{t_0} \Pi_{i=1}^3 (1 + \mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{du}) - H_0)} \\ u d^2 S_{t_0} \Pi_{i=1}^3 (1 + \mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{du}) - H_0)} \\ u d^2 S_{t_0} \Pi_{i=1}^3 (1 + \mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{dd}) - H_0)} \\ d^3 S_{t_0} \Pi_{i=1}^3 (1 + \mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_3}(\omega_{dd}) - H_0)} \end{array} \right.$$

It becomes apparent that when solving the problem in proposition 2.2 a one-period model has two possible states, a two period model four and an n-period model 2^n. The controlled process makes the asset tree 'bushy' and thus causes exponential growth of the number of variables/constraints. But for sufficiently simple[7] contingent claims $C_T(\omega)$, with single-signed Delta[8] two distinct trajectories with identical number of up and down moves at a time t_i will result in identical holdings in stock and bond, e.g. $H_{t_3}(\omega_{ud}) = H_{t_3}(\omega_{du})$. This can be demonstrated backwards step by step. Firstly, at nodes of number of upsteps n_u and downsteps n_d

$$\prod_{i=1}^{n_u+n_d} (1 + \mathrm{sign}(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\gamma)^{(1-\alpha)} = ((1+\gamma)^{n_u}(1-\gamma)^{n_d})^{(1-\alpha)}$$

for positive Delta, and with n_u and n_d swapped for negative Delta. Secondly, if in example 2.1 $(H'_{t_3}(\omega_{ud}), \hat{H}'_{t_3}(\omega_{ud}))$ is a solution to the replication constraint (21) at its point in state space, then so is $(H_{t_3}(\omega_{du}), \hat{H}_{t_3}(\omega_{du})) = (H'_{t_3}(\omega_{ud}), \hat{H}'_{t_3}(\omega_{ud}))$, due to uniqueness. This can be generalised for n time-steps. Then the tree becomes recombining and thus feasible to implement as the number of variables/constraints will be of $O(n^2)$. Still, it represents a possibly large-scale nonlinear optimisation problem (see appendix A). The asset's dynamics are visualised in figure 4.

Table 1 represents a numerical example of proposition 2.2 and corollary 2.1 for a call option, i.e. where $C_T(\omega) = \max(S_T(\omega) - K, 0)$, $\forall\omega$. It becomes apparent that, as γ and λ become bigger, i.e. as the liquidity of the market for the underlying decreases, the option's own bid-ask spreads

[7] Claims for which there exists a unique solution holding strategy $(H'_t(\omega), \hat{H}'_t(\omega))_{\forall t,\omega}$. In general, due to the highly nonlinear setup of the model, it is impossible to show analytically that the claim is unique. We will later present a dynamical programming algorithm that drops this requirement.

[8] This requirement can be dropped when $\gamma = 0$.

European Calls
risk-free rate 5% volatility 20% expiry 12 months initial holding 0 strike 50

			ASK PRICES					BID PRICES			
40 time steps											
alpha 1	moneyness	lambda	0.01	0.001	0.0001	0	exact B-S	0	0.0001	0.001	0.01
gamma 0	0.9		1.073	1.007	1.001	1.000	2.546	1.000	0.999	0.993	0.923
	1		1.059	1.006	1.001	1.000	5.225	1.000	0.999	0.994	0.938
	1.1		1.047	1.005	1.000	1.000	8.831	1.000	1.000	0.995	0.952
gamma 0.001	0.9		1.124	1.060	1.053	1.052		0.947	0.946	0.939	0.866
	1		1.086	1.034	1.029	1.028		0.972	0.971	0.966	0.909
	1.1		1.061	1.020	1.015	1.015		0.985	0.985	0.980	0.936
gamma 0.01	0.9		1.542	1.486	1.480	1.480		0.399	0.398	0.388	0.282
	1		1.309	1.263	1.259	1.258		0.690	0.689	0.681	0.604
	1.1		1.184	1.147	1.143	1.143		0.847	0.846	0.841	0.789
alpha 0	moneyness	lambda	0.01	0.001	0.0001	0		0	0.0001	0.001	0.01
gamma 0	0.9		1.281	1.029	1.003	1.000		1.000	0.997	0.971	0.688
	1		1.180	1.019	1.002	1.000		1.000	0.998	0.982	0.797
	1.1		1.109	1.012	1.001	1.000		1.000	0.999	0.988	0.874
gamma 0.001	0.9		1.374	1.126	1.100	1.097		0.901	0.898	0.871	0.580
	1		1.228	1.070	1.053	1.051		0.949	0.947	0.930	0.737
	1.1		1.135	1.038	1.027	1.026		0.974	0.973	0.962	0.846
gamma 0.01	0.9		2.191	1.962	1.938	1.936		0.019	0.018	0.005	0.000
	1		1.651	1.509	1.495	1.493		0.467	0.464	0.440	0.278
	1.1		1.372	1.280	1.270	1.269		0.784	0.783	0.776	0.717
alpha -1	moneyness	lambda	0.01	0.001	0.0001	0		0	0.0001	0.001	0.01
gamma 0	0.9		1.472	1.051	1.005	1.000		1.000	0.995	0.948	0.434
	1		1.289	1.032	1.003	1.000		1.000	0.997	0.969	0.610
	1.1		1.165	1.018	1.002	1.000		1.000	0.998	0.981	0.797
gamma 0.001	0.9		1.608	1.192	1.147	1.142		0.856	0.850	0.803	0.277
	1		1.359	1.105	1.077	1.073		0.927	0.924	0.895	0.509
	1.1		1.205	1.056	1.040	1.038		0.963	0.961	0.945	0.777
gamma 0.01	0.9		2.845	2.422	2.380	2.375		0.000	0.000	0.000	0.000
	1		1.995	1.754	1.730	1.727		0.373	0.372	0.245	0.000
	1.1		1.567	1.419	1.404	1.402		0.778	0.778	0.761	0.538
60 time steps											
alpha 1	moneyness	lambda	0.01	0.001	0.0001	0		0	0.0001	0.001	0.01
gamma 0	0.9		1.066	0.998	0.991	0.990		0.990	0.989	0.982	0.911
	1		1.061	1.008	1.002	1.002		1.002	1.001	0.996	0.939
	1.1		1.045	1.003	0.998	0.998		0.998	0.998	0.993	0.949
gamma 0.001	0.9		1.126	1.060	1.053	1.052		0.926	0.925	0.918	0.843
	1		1.093	1.041	1.035	1.035		0.968	0.967	0.962	0.903
	1.1		1.062	1.020	1.016	1.015		0.981	0.980	0.976	0.930
gamma 0.01	0.9		1.614	1.557	1.552	1.551		0.242	0.241	0.230	0.111
	1		1.350	1.306	1.301	1.301		0.618	0.617	0.609	0.521
	1.1		1.204	1.167	1.163	1.163		0.821	0.820	0.815	0.760

Table 1: Call option premia under finite liquidity relative to their perfect liquidity Black-Scholes equivalents.

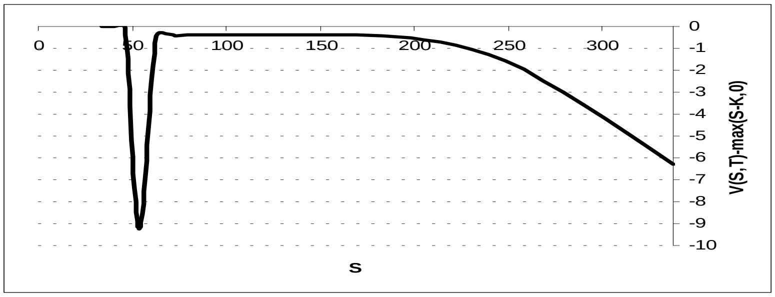


Figure 6: Hedging shortfall at expiry of figure 5 under Black-Scholes Delta hedging.

become wider; γ has a significantly larger effect in this process, especially as the number of time-steps increases. Also, as α decreases the option prices increase, in particular when γ is positive. This is because the slippage of the asset price causes an increase in realised volatility of the asset beyond it's exogenous σ. On the bid side, when both γ and absolute α are large, the lower bound for the option comes into effect. This means that when going long the option the cost of hedging it would always exceed its value, so that a market maker, who cannot accept the possibility of a shortfall, would only take it for free. Furthermore, figures 5 and 6 show the hedging error for a Call option payoff, when the Black-Scholes Delta is used instead of the liquidity-adjusted one. It becomes apparent that this leads to a significant hedging shortfall risk at-the-money and deep in-the-money. For the former the reason is that asset price changes may be magnified by hedging activity, which is especially crucial at expiry as the Delta often becomes discontinuous. The latter is due to the fact that hedging activity in an illiquid underlying leads to an overall increase in realised volatility.

Certain assumptions of proposition 2.2 and corollary 2.1 can be relaxed, for both programs, to make the optimisation problems less restrictive.

Observation 2.1 (Independence of initial holding)
If the liquidity at the initiation of the portfolio is perfect, thus $\lambda = \gamma = 0$, then the objective function (18) collapses to

$$\min_{(H_t, \hat{H}_t)_{\forall t,\omega}} \quad H_{t_1} S_{t_0} + \hat{H}_{t_1} B_{t_0}.$$

Then the initial holding condition (19) is non-binding and we denote the optimal Delta of the option by $H^*_{t_1}$. The solutions of the two programs are identical, if $H_0 = H^*_{t_1}$.

Hence, if if there are no frictions in the setting up of the initial position $H^*_{t_1}$ (the Delta) in the stock, then C_{t_0} is independent of the initial endowment in the risky asset H_0. However, if it is not, then the value of C_{t_0} is crucially dependent upon H_0. If the latter is a large positive or negative amount, then C, theoretically, could be manipulated arbitrarily. But, due to the non-existence of arbitrage, by selling or buying a large quantity of stock, the loss on transaction costs will overweigh any potential market manipulation benefits. Therefore the last term in (18) represents an implicit mark-to-market value of the allocated initial holding in the risky asset. It is easy to observe that the optimal initial holding H_0 would be the Delta $H^*_{t_1}$ under the conditions of observation 2.1. For the trader it would be preferable to scale the payoff C_T of the written option, so that $H_0 = H^*_{t_1}$, rather than setting it up by trading the underlying in the market. This demonstrates that a trader with diversified holdings in both underlying and contingent claims, enjoys economies of scale when allocating capital and hedging a book. Figure 7 shows the value of C_{t_0} under various H_0 and denotes $H^*_{t_1}$.

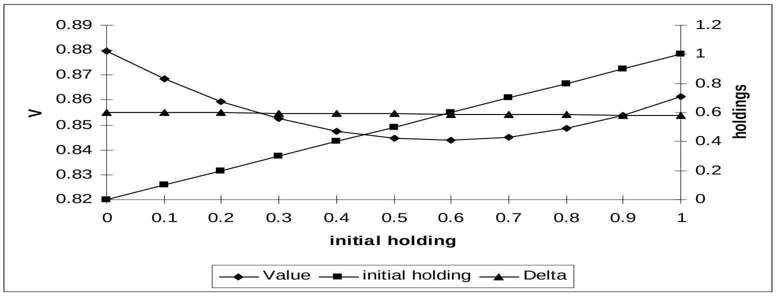


Figure 7: The left axis refers to the option value under different contract Delta and initial holding combinations, as given by the right axis. The minimum contract value is given when $H_0 = H^*_{t_1}$, i.e. where the initial holding and Delta lines cross. Here $T = 0.5$, $S_{t_0} = K = 10$, $\sigma = 0.25$, $r = 0.05$, $\gamma = \alpha = 0$, $\lambda = 0.01$ and 30 timesteps.

Observation 2.2 *The optimal solution C_{t_0} of the program in proposition 2.2 stays unchanged if (20) is relaxed into the inequality*

$$(\hat{H}_{t_i}(\omega) - \hat{H}_{t_{i-1}}(\omega))B_{t_{i-1}} + (H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\bar{S}_{t_{i-1}}(\omega) \leq 0. \tag{22}$$

Proof: Denoting the solution of the program under condition (22) by $\bar{C}_{t_0}$, because the solution space under (22) contains the one under (20) we have that $C_{t_0} \geq \bar{C}_{t_0}$. Also suppose that $(H_t(\omega), \hat{H}_t(\omega))_{\forall t,\omega}$ is optimal with

$$(\hat{H}_{t_i}(\omega) - \hat{H}_{t_{i-1}}(\omega))B_{t_{i-1}} + (H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\bar{S}_{t_{i-1}}(\omega) = -d < 0,$$

for at least one (t_i, ω_j). Then defining a new strategy $(H^*_t(\omega), \hat{H}^*_t(\omega))_{\forall t,\omega}$ with $H^*_t(\omega) = H_t(\omega), \forall t, \omega$ and $\hat{H}^*_t(\omega) = \hat{H}_t(\omega), \forall t, \omega$ except for (t_i, ω_j), where $H^*_{t_i}(\omega) = H^*_{t_i}(\omega) + d/B_{t_{i-1}}$. This will turn condition (22) into (20), while still satisfying condition (21). Thus since d is positive but arbitrary $C_{t_0} \leq \bar{C}_{t_0}$. $\square$

Observation 2.2 states the intuitive assumption that it is not optimal to withdraw funds, while engaging in a replication strategy. But as the next proposition states, the same intuition does not necessarily hold for the tightening of the terminal condition.

Proposition 2.3 (Existence of optimal super-replication strategies)
If the replication condition (21) is turned into a super-replication constraint

$$V_T(\omega) = H_T(\omega)S_T(\omega) + \hat{H}_T(\omega)B_T \geq C_T(\omega), \qquad \forall\omega, \tag{23}$$

then the optimal solution may change.

Proof: A numerical example similar to that of [BLPS] would demonstrate the validity of the proposition. But it is easy to see that if $\alpha = 1$, $\lambda = 0$ and $H_0 = H^*_{t_1}$ then our model collapses to a standard proportional transaction cost model for which the conditions of existence of super-replication strategies are derived by [Rut].$\square$

Therefore, if the replication condition (21) is turned into an inequality constraint, the optimal solution may change and the contract may actually become cheaper. This is due to the fact that as rehedging is costly, it may be a cheaper strategy to keep a certain hedge quantity constant and eventually super-replicate a payoff, instead of liquidating it and thus incurring transaction costs. The property that it may be sub-optimal to rebalance portfolios with proportional transaction costs has been discovered by [Con] for consumption-investment problems. For the hedging of contingent claims [W&W] categorise transaction cost models that allow super-replication like e.g. [BLPS] or [ENU] as "global-in-time", whereas exact replication models like those of [Lel], [B&V] or [HWW] are referred to as "local-in-time", also referring to the respective solution methods.

[Rut] terms the solution of the super-replication formulation of proposition 2.3 as "perfect hedging" and notes that while the cheapest strategy that super-replicates the payoff is preference-free, it does not represent an arbitrage price as such, because the writer would be able to make a riskless profit. Nonetheless due to the nonlinearity of short and long positions, it would not be arbitrage, but competition that would force a lower price. The latter would depend on the respective market participants' risk appetite, because for lower selling prices there would exist a positive probability of a hedging shortfall.

Moreover, as [BLPS], [Rut] and [W&W] found, super-replication strategies exist if, firstly, transaction costs are large enough and, secondly, if claims can be settled in cash or at the discretion of the writer, or equivalently, when it is costless to perform the final portfolio liquidation. In each case, there will exist so-called no-transaction bands around the hedge quantity that determine a region where the marginal costs of rehedging are greater than the marginal benefit of exactly meeting a future claim. [ENU] provide a solution method consisting of a two-stage dynamical program that determines both the solution and the hedging strategy. Instead of a unique strategy that satisfies the constraints of the program, there may now exist a compact set. To solve for the cheapest of the super-replicating strategies, they discretise the space of trading strategies, which is otherwise continuous, and calculate the expected hedging shortfall under an arbitrary but positive probability measure. Solving backwards they eventually choose the cheapest strategy that results in a zero shortfall. This method also automatically takes care of limited divisibility of traded assets and lot sizes of the underlying.

Proposition 2.4 (Super-replication ask price)
If the replication condition (21) is turned into a super-replication constraint

$$H_T(\omega)S_T(\omega) + \hat{H}_T(\omega)B_T \geq C_T(\omega), \qquad \forall\omega, \tag{24}$$

then the solution C_{t_0} of the program of proposition 2.2 is equivalent to

$$D_{t_0} = \min c$$

subject to

$$W_{t_0}(c) = 0,$$

where

$$W_{t_0}(c) = \min_{(H_t(\omega),\hat{H}_t(\omega))_{\forall t,\omega}} E_P\left[\max\left(C_T - H_T S_T - \hat{H} B_T, 0\right)\right] \tag{25}$$

subject to

$$H_{t_1}\bar{S}_{t_0} + \hat{H}_{t_1}B_{t_0} = c$$

as well as (19), (21) and the super-replication (24). The expectation is taken with respect to an arbitrary probability measure $P(\omega) > 0$.

Proof: (See [ENU]) It is easy to see that the solution C_{t_0} of proposition 2.2 satisfies the constraints of the program of proposition 2.4, thus

$$C_{t_0} \geq D_{t_0}.$$

But at the same time D_{t_0} satisfies the constraints in proposition 2.2 as well, hence

$$C_{t_0} \leq D_{t_0}$$

and they are equal.□

The implementation and backward solution of the dynamical programming algorithm is shown in the appendix B.

3 The continuous-time limit

As the frequency of rebalancings of a portfolio is increased, the trader has to transact ever-smaller amounts more frequently. In the limit small amounts are traded continuously. In general, as is the case in the standard transaction cost models, if transaction costs as fraction of value traded are of $O(1)$, as in the models of e.g. [B&V], [Lel] or [HWW], then V_{t_0} approaches its zero-diffusion solution as the rehedging interval $\delta t \to 0$. In fact as [SSC] found, for all proportional transaction cost models, in the continuous-time limit, the sole optimal solution for the hedging of contingent cash flows that guarantees that $V_T(\omega) \geq C_T(\omega), \forall\omega$ is to take a 100% Delta position ($H_{t_1} = 1$), i.e. a static hedge up-front. Nonetheless as [Lel], if the rehedging interval δt is small, but non-infinitesimal, and the Delta is appropriately adjusted, the probability of a large hedging error becomes small. In our model liquidity has an impact which is a combination of orders of magnitude $O(\delta t)$, $O(\sqrt{\delta t})$ and $O(1)$, depending on the parameters α, γ, λ, which themselves are all assumed to be $O(1)$. We commence by considering $\gamma > 0$. Also, for notational simplicity we will drop the time subscript.

Theorem 3.1 *Under the price dynamics and the self-financing conditions of proposition 2.2, as rehedging takes place after intervals δt, if $\alpha < 1$ the cost of a generic replicating portfolio $V(S,t)$, to leading order of its Gamma and provided that its Delta is single-signed, is governed by the PDE*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\hat{\sigma}^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0, \tag{26}$$

where the variance is given by

$$\hat{\sigma}^2 = \left(\frac{1 + \frac{1}{2}\left((1+\gamma)^{2(1-\alpha)} + (1-\gamma)^{2(1-\alpha)}\right) - (1+\gamma)^{(1-\alpha)} - (1-\gamma)^{(1-\alpha)}}{\delta t} \right), \tag{27}$$

subject to contract-specific boundary conditions.

Proof: See appendix C.□

We see that the Black-Scholes formula still applies, but with a modified volatility. Moreover, to leading order of the Gamma term,[9] the asset's exogenous volatility has vanished, because the large trader's rehedging activity entirely dominates the asset price's diffusion, the latter representing the asset dynamics due to other market participants. Also the bid-ask spread effect dominates the price elasticity effect, as λ doesn't appear either. However, as mentioned before, once $\alpha = 1$, i.e. the absence of asset price slippage, the dominance of the market manipulation effect vanishes and the Gamma coefficient becomes smaller. Now if we allow a term of second Gamma order, then we obtain the transaction costs model of [HWW].

[9] Here we assume that the rehedging intervals δt are sufficiently small so that $\hat{\sigma} \gg \sigma$.

Corollary 3.1 *In theorem 3.1, if $\gamma > 0$ and $\alpha = 1$, then, to second Gamma order, $V(S,t)$ satisfies*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{1}{2}\tilde{\sigma}^2 S^2 \left|\frac{\partial^2 V}{\partial S^2}\right| + rS\frac{\partial V}{\partial S} - rV = 0, \tag{28}$$

with the variance

$$\tilde{\sigma}^2 = \gamma\sigma\sqrt{\frac{2}{\pi\delta t}}. \tag{29}$$

Proof: See appendix C.□

Furthermore if Gamma is positive throughout, as is the case for long calls and puts, then (28) becomes the transaction costs model of [Lel],[10] which has a Black-Scholes solution with modified variance

$$\bar{\sigma}^2 = \sigma^2 + \gamma\sigma\sqrt{\frac{2}{\pi\delta t}}.$$

So if $\alpha = 1$, to leading order, there is no increase in effective volatility due to market manipulation, but only due to transaction costs; still as $\delta t \to 0$ the Gamma term vanishes and the portfolio value approaches the cost of the static hedge solution.

But under the circumstances where the bid-ask spreads are tight enough so that we can set $\gamma = 0$, there exists a true continuous-time hedging limit. Then, because transaction costs only appear implicitly through an increasing price impact, in the limit the quantity that needs to be traded is small enough[11], so that transaction costs will stay finite for appropriate structures of C. The Gamma terms remain O(1) and, moreover, there are no longer restrictions on the single-signedness of the Delta.

Theorem 3.2 *Under the price dynamics and the self-financing conditions of proposition 2.2, if $\gamma = 0$, in the continuous-time limit, the value of a generic replicating portfolio $V(S,t)$ is governed by the PDE*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \lambda\sigma^2 S^3 \left(\frac{\partial^2 V}{\partial S^2}\right)^2 + \frac{1}{2}\lambda^2(1-\alpha)^2\sigma^2 S^4 \left(\frac{\partial^2 V}{\partial S^2}\right)^3 + rS\frac{\partial V}{\partial S} - rV = 0, \tag{30}$$

subject to contract-specific boundary conditions.

Proof: See appendix C.□

Corollary 3.2 *Under the assumptions of theorem 3.2 the hedge quantity $\Delta = \partial V/\partial S$ is governed by the PDE*

$$\begin{aligned} \frac{\partial \Delta}{\partial t} + \frac{1}{2}\sigma^2 S^2 \left(1 + 4\lambda S\frac{\partial \Delta}{\partial S} + 3\lambda^2(1-\alpha)^2 S^2 \left(\frac{\partial \Delta}{\partial S}\right)^2\right)\frac{\partial^2 \Delta}{\partial S^2} & \\ +(r+\sigma^2)S\frac{\partial \Delta}{\partial S} + 3\lambda\sigma^2 S^2 \left(\frac{\partial \Delta}{\partial S}\right)^2 + 2\lambda^2(1-\alpha)^2\sigma^2 S^3 \left(\frac{\partial \Delta}{\partial S}\right)^3 &= 0. \end{aligned} \tag{31}$$

[10]For reason's of consistency we also have to note that, allowing a second-order Gamma term we implicitly assumed that the drift of the process of the underlying was 0, as otherwise it would be of the same order and thus non-negligible. In the original derivation of [Lel] this point was not addressed clearly.

[11]It turns out to be of $O(\sqrt{dt})$

Proof: Taking the partial derivative of (30) with respect to S completes the proof.$\square$

Equation (30) is a fully nonlinear PDE and (31) is a quasi-linear PDE, i.e. the coefficient of its highest order partial derivative is a function of a lower order one. The former PDE thus structurally resembles the ones derived by [Schö], [F&S], [S&P] and the latter the one of [Frey]. The fact that both PDEs collapse to their Black-Scholes equivalents

$$\mathcal{L}_{BS}V = \frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV = 0, \tag{32}$$

and

$$\frac{\partial \mathcal{L}_{BS}V}{\partial S} = \frac{\partial \Delta}{\partial t} + (r+\sigma^2)S\frac{\partial \Delta}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 \Delta}{\partial S^2} = 0$$

when $\lambda = 0$, i.e. when the market is perfectly liquid, can be exploited to find approximations of V and Δ for small λ. Denoting by V_{BS} and Δ_{BS} the Black-Scholes values, we expand V in a regular perturbation series

$$V \approx V_{BS} + \lambda V_1 + \lambda^2 V_2 \tag{33}$$

and substitute it into (30). By matching the corresponding orders of magnitude we obtain equations for the first- and second-order terms of the value:

$$\begin{aligned}
\mathcal{L}_{BS}V_1 &= -\sigma^2 S^3 \left(\frac{\partial^2 V_{BS}}{\partial S^2}\right)^2, \\
\mathcal{L}_{BS}V_2 &= -\sigma^2 S^3 \left(\frac{\partial V_1}{\partial S}\right)^2 - \frac{1}{2}(1-\alpha)^2\sigma^2 S^4 \left(\frac{\partial^2 V_{BS}}{\partial S^2}\right)^3,
\end{aligned}$$

and by differentiating Δ with respect to S also

$$\begin{aligned}
\frac{\partial \mathcal{L}_{BS}V_1}{\partial S} &= -3\sigma^2 S^2 \left(\frac{\partial \Delta_{BS}}{\partial S}\right)^2 - 2\sigma^2 S^3 \frac{\partial \Delta_{BS}}{\partial S}\frac{\partial^2 \Delta_{BS}}{\partial S_t^2}, \\
\frac{\partial \mathcal{L}_{BS}V_2}{\partial S} &= -3\sigma^2 S^2 \left(\frac{\partial \Delta_1}{\partial S}\right)^2 - 2\sigma^2 S^3 \frac{\partial \Delta_1}{\partial S}\frac{\partial^2 \Delta_1}{\partial S} - 2(1-\alpha)^2\sigma^2 S^3 \left(\frac{\partial \Delta_{BS}}{\partial S}\right)^3 \\
&\quad -\frac{3}{2}(1-\alpha)^2\sigma^2 S^4 \left(\frac{\partial \Delta_{BS}}{\partial S}\right)^2 \frac{\partial^2 \Delta_{BS}}{\partial S^2},
\end{aligned}$$

whose solutions depend on the boundary conditions and can be written in integral form,[12] but, in general, direct numerical solutions of the PDEs are easier to compute.

3.1 The pricing and hedging of contingent claims in continuous time

As the models in theorem 3.1 and corollary 3.1 are similar or even identical to existing ones on transaction costs, we will henceforth focus on the genuinely liquidity-oriented model in theorem 3.2. The latter also has the advantage that a true continues-time limit exists and that there are no restrictions on the Delta of the payoff. Nonetheless, for most type of payoffs $C(S,T)$ care has to be taken because at expiry they are usually non-smooth or even discontinuous. For instance, the payoff of a call option has a discontinuous gradient at $S = K$ at expiry, hence its Gamma is a Dirac delta-function. But the square of this delta-function is not well-defined in the classical distributional sense and therefore (30) at expiry would be ill-posed in the absence of further constraints. Another

[12] [S&P] derive an integral expression of the first-order term for the case when V is a vanilla call option.

condition that has to be imposed is that $V(S,t) \geq 0$. Because the square of the Gamma term is always positive, therefore e.g. short positions with negative Gamma may lead to negative prices.[13] The latter would, economically, not be realistic since the buyer of an option would never face any obligations. Finally, the initial holding of the asset has to be taken into account, because the set-up costs of the initial Delta is significant.

Proposition 3.1 (Ask price of a contingent claim)
The ask price of a contingent claim $C(S,t)$ at time is

$$C(S,t) = \min_{(V(S,T))\forall S} \quad V\left((1+\text{sign}(\Delta^* - H)\gamma)^{(1-\alpha)} S e^{\lambda(1-\alpha)(\Delta^*-H)}, t\right) + (\Delta^* - H)\left((1+\text{sign}(\Delta^*-H)\gamma)e^{\lambda(\Delta^*-H)} - 1\right) S \tag{34}$$

subject to

$$V(S,T) \geq C(S,T), \qquad \forall S, \tag{35}$$

where H is the initial holdings and Δ^* is such that

$$\Delta^* = \left.\frac{\partial V}{\partial S}\right|_{\left((1+\text{sign}(\Delta^*-H)\gamma)^{(1-\alpha)} S e^{\lambda(1-\alpha)(\Delta^*-H)}, t\right)}. \tag{36}$$

Moreover $V(S,T)$ follows (30), with the free boundary $V(S,t) \geq 0$ everywhere.

The second term of the objective function represents an initial set-up cost, which is incurred if the current asset holding H are different from the required delta. As in the discrete-time case it becomes apparent that it is cheapest initially to hold an amount as close to the position's Delta as possible in order to save on transaction costs. The super-replication condition (35) ensures that payoff discontinuities will be smoothed out. In general, finding the optimal solution, which can for instance be done through the discrete time model, would be computationally expensive. Instead, in order to find a close approximation [S&P] suggest that e.g. for European call and put options (30) is valid only up to a small time before expiry: $T - \epsilon$. Another possibility that [Frey] proposes is to approximate the payoff by a smooth function, as for example by

$$C(S,T) = \frac{1}{2}\left(S - K + \sqrt{(S-K)^2 + \epsilon}\right)$$

for a Call option.

Figure 8 demonstrates the liquidity effects on a Call option. It becomes apparent that our model produces a volatility frown, instead of the regularly observed skew and smile pattern. But as for instance [Con2] noted, in general, transaction costs and liquidity effects cannot explain the smile. This is because both are directly positively related to the Gamma of an option, which is largest at-the-money for vanillas. So if, as most of the time in equity markets, a skew persisted, that would make out-of-the-money Puts more expensive, our model would explain the shape of the skew in-the-money. But, clearly, there must be other reasons for the out-of-the-money part of the skew or smile, as the replication argument states that transaction cost will be lowest for deeply in- and out-of-the-money options.

[13] This fully corresponds to the discrete-time results of table 1.

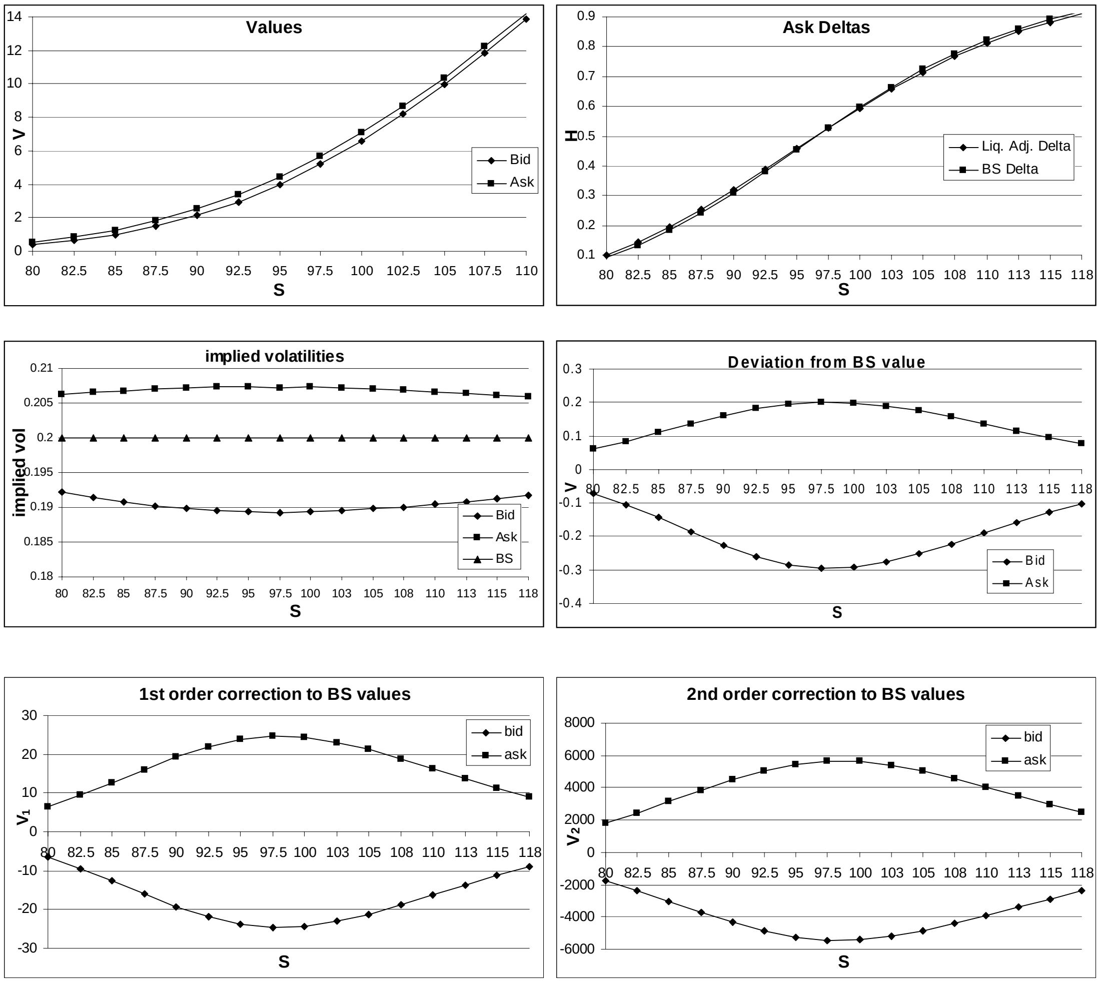


Figure 8: Long and short Call options with $T = 0.5$, $\sigma = 0.2$, $K = 100$, $r = 0.05$, $\lambda = 0.01$, $\alpha = 0$, $H = \Delta^*$, $\epsilon = 1/800$, the latter applied to the modification of T.

4 Parameterisation and calibration of the model

4.1 A proxy measure for liquidity

The definition of liquidity has, so far, not converged to a level of standardisation as that of, say, volatility. For instance economists may have a different notion of liquidity than financial mathematicians. In an empirical analysis under the former concept [Per] suggests asset flows as a proxy. In a trading and investment context [CRS] provide measures of liquidity based on bid-ask spreads of assets. Finally, from an option pricing perspective, a market depth approach is taken by [Krak], who explicitly defines liquidity as the reciprocal of $\Delta H/\Delta S$, i.e. the sensitivity of the stock price to the quantity traded. However in this form the parameter is not dimensionless and depends on the absolute size of both the quantity and nominal stock price. Our model is a combination of the last two. The parameter γ is a direct measure of the bid-ask spreads between the best layers in an order book, whereas λ scales the slope of the average transaction price i.e. measures the market depth. Finally α transforms the average transaction price into either the new last price traded, which will appear on the screens of all market participants or, if applicable, the new mid-market price derived therefrom.

Ideally one would like to have an intuitive and observable measure of liquidity, that is both dimensionless and comparable across assets or markets. For this purpose the product λdH and $(1+\text{sign}(dH)\gamma)$ are dimensionless variables, where λ is approximately the marginal degree of price change per unit number of assets traded dH that the trader faces in a transaction beyond the initial relative bid-ask spread $1 \mp \gamma$, so that

$$\begin{aligned}\%\text{ change in average transaction price} &= \frac{(1+\text{sign}(dH)\gamma)S_{t_0}e^{\lambda dH} - S_{t_0}}{S_{t_0}} \\ &\approx \text{sign}(dH)\gamma + \lambda dH + \text{sign}(dH)\gamma\lambda dH.\end{aligned}$$

Across assets this can further be made comparable by defining

$$\hat{\lambda} = \lambda \times \text{Total no. of assets outstanding},$$

so that $\hat{\lambda}$ could be a universal definition of liquidity for e.g. equity as

$$\%\text{ acquisition (disposal) premium (discount)} \approx \pm\gamma + \hat{\lambda}(1 \pm \gamma) \times \%\text{ market cap traded.} \tag{37}$$

A second definition of liquidity could be given in terms of degree of market slippage, i.e. movement of asset price due to the trade itself, in which case

$$\%\text{ asset price slippage} \approx (1 \pm \gamma)^{(1-\alpha)} \times (1 + \hat{\lambda}(1-\alpha) \times \%\text{ market cap traded}) - 1,$$

which simplifies to

$$\%\text{ asset price slippage} \approx \hat{\lambda}(1-\alpha) \times \%\text{ market cap traded}$$

when γ is set to zero. All three parameters can be directly observed through the order book of various stocks, which then under their respective combinations give a proxy for liquidity. A similar definition to (37) with $\gamma = 0$ was proposed by [Kyle] and tested for by [BHK] on stocks traded on the NYSE, which however is a specialist market so that the full order book is only available to one market maker. Therefore the average transaction price change and slippage are not directly determined and have elements of randomness. More recently, [DFIS] simulated the price dynamics

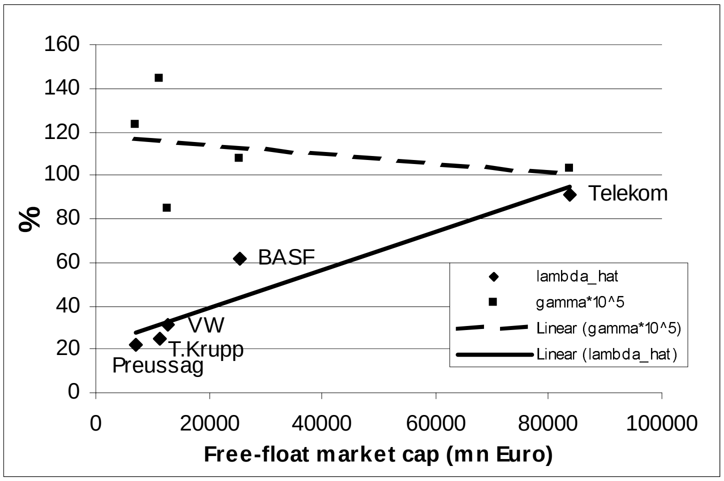


Figure 9: Average acquisition premia $\hat{\lambda}$ and bid-ask spreads γ for various large stocks and their trend lines.

of an order book in a non-specialist market and came up with a similar parameterisation for liquidity in terms of bid-ask spreads and slope.

Under our parameterisation figure 9 shows the ranking of average γ and $\hat{\lambda}$, as defined by (37) for various stocks against free-float market capitalisation. It becomes apparent, that bid-ask spreads seem to be approximately negatively related to the market value of a company but, at first glance counter-intuitively, the marginal change in average transaction price is clearly positively related. This may be due to the fact that, at least for blue chip stocks, at any one moment the market depth of the smaller ones is relatively larger than for the bigger ones. This means, when extrapolating, if at one point in time a market order for a 1% free-float stake of a large stock like Telekom is placed, the acquisition premium will amount to about 90%. On the other a smaller cap like Preussag would only commmand a 20% premium. Another reason may be that the larger stocks are more volatile and therefore the opportunity cost of holding inventory in an order book is higher.

4.2 Estimation of the liquidity parameters

Our data set consisted of about four and a half hours of five layers of market depth data for five large cap stocks traded on the Xetra system of the German stock exchange. The latter is a fully electronic market that matches quotes automatically. Even though this set seems relatively small, still each stock had between 300 and 1000 ticks per hour, where at least one of the layers was updated in price or quantity offered. We assumed that a potential hedger would require the immediate execution of an order and would thus fill one of the limit orders in the book. This assumption would not necessarily be invalidated if on top of the order book another broker market existed, as it may be reasonable to assume that the liquidity supplied by brokers is related to the observable part of the order book.

To estimate the parameters we employed ordinary least squares minimisation because our model setup fits well into a linear regression framework, which further provides the R^2 measure of goodness-of-fit. We also let the fitted parameters be time-dependent, i.e. λ_t and γ_t to subsequently test

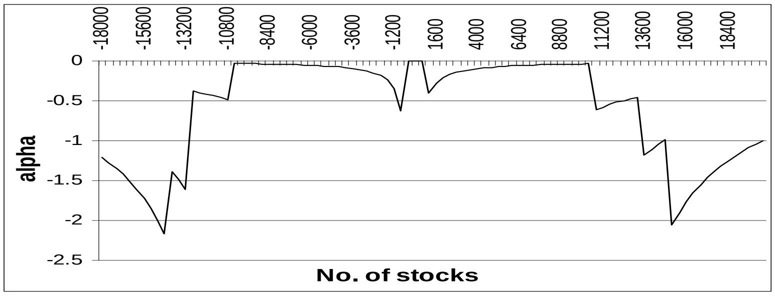


Figure 10: α for the time series of figure 3.

stability. Depending on the a priori assumptions of which parameters to include, we added them step-by-step thus increasing complexity and observed how the fit of the model developed. Initially we commenced by setting $\gamma_t = 0$ and ignored the market slippage, then λ_t and S_t could be directly estimated through the simple linear regression with log-transform of the response:

$$\ln(\bar{S}_t) = \ln(S_t) + \lambda dH_t + \epsilon_t. \tag{38}$$

Here $\bar{S}_t$ represents the exact average transaction price under a specific traded quantity dH_t and S_t the estimated mid-market price. The choice of intervals of dH_t proved to have relatively little influence on the results, so that we took the same equi-distant quantities for each stock, i.e. batches of 400 shares, which typically provided 40-100 data points for each regression. The results of the analysis are given in table 2. The time-weighted arithmetic average R^2 statistic shows that the functional form of the controlled process seems to be able to provide a good fit at most time points. But as can be observed from the corresponding graph there is a significant discontinuity due to bid-ask spreads, so to further improve the fit the next extension was to include γ_t in the estimation. The extended regression model, again under log-transform of the response is

$$\ln(\bar{S}_t) = \ln(S_t) + \ln(1 + \text{sign}(dH_t)\gamma_t) + \lambda_t dH_t + \epsilon_t. \tag{39}$$

Then a good approximation of γ_t could be obtained by conjecturing it is small, so that

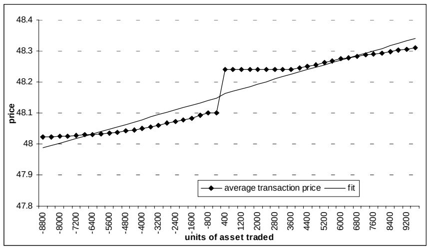
$$\ln(1 + \text{sign}(dH_t)\gamma_t) \approx \text{sign}(dH_t)\gamma_t,$$

which fits into a multiple linear regression framework. As the result table shows, the inclusion of γ_t further improved the R^2 value. But, as the section on stochastic liquidity will demonstrate, the values of λ_t and γ_t across time were far from constant. One problem that estimation method (39) causes is that as the slope of the market depth becomes steeper, the intercepts come closer together and it is even possible, albeit rare, that the estimated best fitting γ may become slightly negative. This could be prevented by restricting γ to represent the exact bid-ask spread at any time, thus requiring one fewer parameter to be estimated through least-squares. But this additional restriction is at the expense of estimation accuracy as the results demonstrate. The least-squares linear regression model is

$$\ln(\bar{S}_t) - I_{[x>0]}(dH_t)\ln(S_t^+) - I_{[x<0]}(dH_t)\ln(S_t^-) = \lambda_t dH_t + \epsilon_t, \tag{40}$$

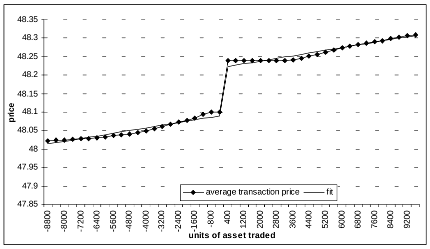
where $I_{[A]}(x)$ is the indicator function and S_t^-, S_t^+ the best bid and ask prices at any time respectively, so that $\gamma = (S_t^+ + S_t^-)/2$. This model could then again be improved in terms of accuracy by allowing for two different different slopes, i.e. a separate λ for bid and ask depth:

$$\ln(\bar{S}_t) - I_{[x>0]}(dH_t)\ln(S_t^+) - I_{[x<0]}(dH_t)\ln(S_t^-) = \lambda_t^+ dH_t I_{[x>0]}(dH_t) + \lambda_t^- dH_t I_{[x<0]}(dH_t) + \epsilon_t, \tag{41}$$



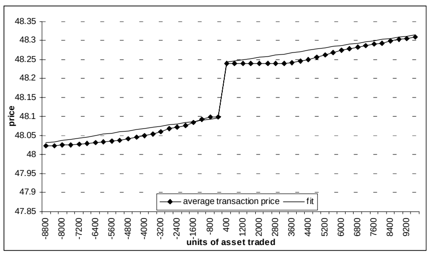
gamma=0

	lambda	R^2
BASF	2.53E-07	87.31%
Telekom	1.64E-07	84.09%
Thy.Krupp	1.56E-07	82.63%
VW	2.47E-07	88.31%
Preussag	3.36E-07	87.82%



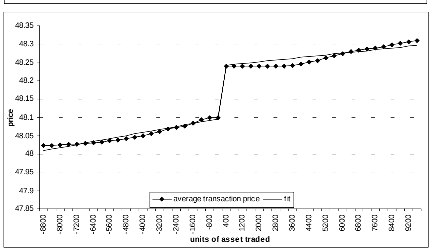
gamma>0

	lambda	gamma	R^2
BASF	1.16E-07	0.00108	98.33%
Telekom	7.48E-08	0.00103	98.09%
Thy.Krupp	6.90E-08	0.00145	98.33%
VW	1.30E-07	0.00085	98.07%
Preussag	1.65E-07	0.00123	98.53%



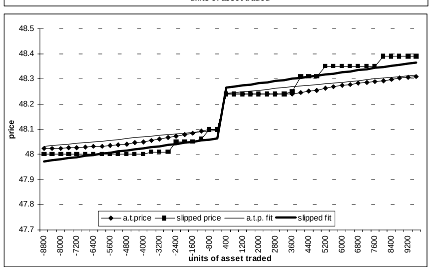
exact gamma

	lambda	gamma	R^2
BASF	1.18E-07	0.00099	94.29%
Telekom	7.41E-08	0.00095	89.04%
Thy.Krupp	7.54E-08	0.00120	81.80%
VW	1.23E-07	0.00081	89.49%
Preussag	1.67E-07	0.00115	92.78%



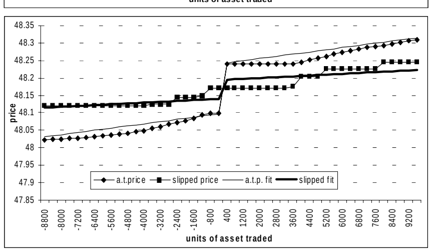
exact gamma, 2 lambda

	lambda_bid	lambda_ask	gamma	R^2
BASF	1.29E-07	1.50E-07	0.00099	96.65%
Telekom	8.80E-08	9.49E-08	0.00095	95.15%
Thy.Krupp	1.23E-07	9.92E-08	0.00120	97.65%
VW	1.34E-07	1.65E-07	0.00081	94.99%
Preussag	2.37E-07	1.55E-07	0.00115	96.49%



slippage to exact next last price traded

	alpha	lambda	gamma	R^2 (2nd)
BASF	-0.53	1.18E-07	0.00099	93.95%
Telekom	-0.54	7.41E-08	0.00095	91.99%
Thy.Krupp	-0.59	7.54E-08	0.00120	93.76%
VW	-0.66	1.23E-07	0.00081	92.11%
Preussag	-0.54	1.67E-07	0.00115	93.86%



slippage to next last mid-price

	alpha	lambda	gamma	R^2 (2nd)
BASF	0.51	1.18E-07	0.00099	81.53%
Telekom	0.52	7.41E-08	0.00095	76.36%
Thy.Krupp	0.48	7.54E-08	0.00120	82.39%
VW	0.42	1.23E-07	0.00081	80.31%
Preussag	0.50	1.67E-07	0.00115	81.07%

Table 2: Time weighted arithmetic average estimates of λ, γ and α for various stocks and their goodness-of-fit. Top to bottom the regression models refer to (38), (39), (40), (41), (42), (43). Graphs represent the various fits for a particular snapshot of the order book of BASF at 12:14:41pm on 13 Jan 2000.

The modelling implications of this will be discussed in a subsequent section. When additionally fitting the slippage parameter α, because it can be calculated exactly, a second equation is given for the slipped price $\hat{S}_t$, namely

$$\ln(\hat{S}_t) = \ln(S_t) + (1 - \alpha_t)(\ln(1 + \mathrm{sign}(dH_t)\gamma_t) + \lambda_t dH_t) + \epsilon_t, \tag{42}$$

where λ and γ are taken from the initial fitting of the average transaction price. But as mentioned in section 2, for reasons of consistency it may be more appropriate to calibrate the slippage parameter to the slipped mid-price instead of the next last traded price $\hat{S}_t$. The former could be estimated by the geometric average

$$\sqrt{\hat{S}_t \left(S_t^- I_{[x>0]}(dH_t) + S_t^+ I_{[x<0]}(dH_t)\right)},$$

so that the left-hand side of (42) is replaced by

$$\frac{1}{2}\left(\ln(\hat{S}_t) + \ln\left(S_t^- I_{[x>0]}(dH_t) + S_t^+ I_{[x<0]}(dH_t)\right)\right). \tag{43}$$

But as figure 10 shows, in general, care has to be taken with α because it is usually a function of dH_t. Nonetheless, the dynamical programming formulation of appendix B is able to cope with both time- and trade quantity-dependent α, so that a generic funtion $\alpha(S, t, dH)$ could be introduced.

5 Extensions and applications

5.1 Multiple underlying assets

The basic model with $\gamma = 0$ easily generalises if the economy has m risky assets $(S_t^{(l)}(\omega))_{\forall t,\omega}$, $l = 1 \ldots m$. In continuous time they follow respective geometric Brownian motions

$$dS_t^{(l)} = \mu_l S_t^{(l)} dt + \sigma_l S_t^{(l)} dX_t^{(l)},$$

which in discrete time gives the price diffusion and impact dynamics $\forall l$ as

$$\begin{aligned} S_{t_{i-1}}^{(l)} \quad &\rightarrow \quad \bar{S}_{t_{i-1}}^{(l)} = S_{t_{i-1}}^{(l)} e^{\lambda_l\left(H_{t_i}^{(l)} - H_{t_{i-1}}^{(l)}\right)} \\ &\rightarrow \quad \begin{cases} u_l \left(S_{t_{i-1}}^{(l)}\right)^{\alpha_l} \left(\bar{S}_{t_{i-1}}^{(l)}\right)^{1-\alpha_l} = u_l S_{t_{i-1}}^{(l)} e^{\lambda_l(1-\alpha_l)\left(H_{t_i}^{(l)} - H_{t_{i-1}}^{(l)}\right)} \\ d_l \left(S_{t_{i-1}}^{(l)}\right)^{\alpha_l} \left(\bar{S}_{t_{i-1}}^{(l)}\right)^{1-\alpha_l} = d_l S_{t_{i-1}}^{(l)} e^{\lambda_l(1-\alpha_l)\left(H_{t_i}^{(l)} - H_{t_{i-1}}^{(l)}\right)} \end{cases}. \end{aligned} \tag{44}$$

Firstly, the economy still doesn't allow arbitrage.

Proposition 5.1 (Non-existence of arbitrage opportunities)
If the risky and riskless assets $(S_t^{(l)}(\omega), B_t)_{\forall t,\omega,l}$ follow the processes (44) and (2) respectively, there does not exist a particular holding strategy $(H_t^{(l)}{}'(\omega), \hat{H}_t'(\omega))_{\forall t,\omega,l}$, with $H_{t_0}^{(l)}{}' = H_T'(\omega)$, $\forall \omega, l$, which is self-financing and value-conserving i.e.

$$\left(\hat{H}_{t_i}'(\omega) - \hat{H}_{t_{i-1}}'(\omega)\right) B_{t_{i-1}} + \sum_{l=1}^{m} \left(H_{t_i}^{(l)}{}'(\omega) - H_{t_{i-1}}^{(l)}{}'(\omega)\right) \bar{S}_{t_{i-1}}^{(l)}(\omega) = 0, \qquad \forall t, \omega, l,$$

which results in a positive expected gain

$$E[V_T - V_{t_0} | \mathcal{F}_{t_0}] > 0,$$

where $V_{t_i}(\omega) = \sum_{l=1}^m H_{t_i}^{(l)}(\omega) S_{t_i}^{(l)}(\omega) + \hat{H}_{t_i}^{(l)}(\omega) B_{t_i}$, $\forall t, \omega$ is the mark to market value of the portfolio.

Proof: Analogous to proposition 2.1.□

Secondly, the nonlinear optimisation program for a contingent claim $C_t(S_t^{(1)}, \ldots, S_t^m)$ is given by:

Proposition 5.2 (Replication ask price of a contingent claim)
The ask price of a contingent claim $C_t(\omega)$ at time t_0 is

$$C_{t_0} = \min_{(H_t(\omega)^{(l)}, \hat{H}_t^{(l)}(\omega))_{\forall t,\omega,l}} V_{t_0} = \sum_{l=1}^{m} \left(H_{t_1}^{(l)*} - H_0^{(l)} \right) \bar{S}_{t_0}^{(l)} + \hat{H}_{t_1}^{*} B_{t_0} + \sum_{l=1}^{m} H_0^{(l)} S_{t_0}^{(l)}, \tag{45}$$

where the optimal controls $(H_t^{(l)*}(\omega), \hat{H}_t^{(l)*}(\omega))_{\forall t,\omega,l}$ satisfy the initial holdings (46)

$$H_{t_0}^{(l)} = H_0^{(l)}, \tag{46}$$

the self-financing conditions (47)

$$(\hat{H}_{t_i}(\omega) - \hat{H}_{t_{i-1}}(\omega))B_{t_{i-1}} + \sum_{l=1}^{m} \left(H_{t_i}^{(l)}(\omega) - H_{t_{i-1}}^{(l)}(\omega) \right) \bar{S}_{t_{i-1}}^{(l)}(\omega) = 0, \tag{47}$$

the payoff replication constraint (48)

$$V_T(\omega) = \sum_{l=1}^{m} H_T^{(l)}(\omega) S_T^{(l)}(\omega) + \hat{H}_T(\omega) B_T = C_T(\omega), \tag{48}$$

and the processes $(B_t(\omega), S_t^{(l)}(\omega), \bar{S}_t^{(l)}(\omega))_{\forall t,\omega,l}$ are given by (2) and (44).

If, as in the univariate case, C_T has a sufficiently simple structure, then there exists a unique optimal solution, but now since a permutation of m binomial trees has a number of nodes that grows like $(n+1)^m$ with timesteps n, the order of magnitude of the number of constraints variables is $O(n^{m+1})$ and it becomes apparent that the discrete-time method will be impractical. Again we can derive a PDE as its continuous-time limit, which can be used by finite difference methods and may make it more feasible to compute solutions.

Theorem 5.1 *Under the price dynamics and the replication conditions of proposition 5.2, in the continuous-time limit the value of a self-financing replicating portfolio $V(S^{(1)}, \ldots, S^{(m)}, t)$ is governed by the PDE*

$$\frac{\partial V}{\partial t} + r \sum_{l=1}^{m} S^{(l)} \frac{\partial V}{\partial S^{(l)}} + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_i \sigma_j \rho_{ij} S^{(i)} S^{(j)} \frac{\partial^2 V}{\partial S^{(i)} \partial S^{(j)}} + \sum_{l=1}^{m} \lambda_l \sigma_l^2 \left(S^{(l)} \right)^3 \left(\frac{\partial^2 V}{\partial \left(S^{(l)} \right)^2} \right)^2$$
$$+ \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j (1-\alpha_i)(1-\alpha_j) \sigma_i \sigma_j \rho_{ij} \left(S^{(i)} \right)^2 \left(S^{(j)} \right)^2 \frac{\partial^2 V}{\partial \left(S^{(i)} \right)^2} \frac{\partial^2 V}{\partial \left(S^{(j)} \right)^2} \frac{\partial^2 V}{\partial S^{(i)} \partial S^{(j)}} - rV = 0,$$

subject to contract-specific boundary conditions.

Proof: A straightforward multivariate extension of theorem 3.2, noting that $dX_t^{(i)} dX_t^{(j)} = \rho_{ij} dt$, $\rho_{ii} = 1$ and $dX_t^{(i)} dX_{t-dt}^{(j)} = 0$ are the respective correlation coefficients between the Brownian motions.□

5.2 Distinct bid and ask liquidity

Typically, as figure 2 demonstrates, the market depth on the bid and ask side is not symmetric . If there are large imbalances, intuitively, this leads to price movements and to an increase in volatility. But by buying when everybody wants to sell and vice versa, the liquidity for the transaction will be good. The converse will probably hold if one follows all other market participants. The reaction function (5) offers only one scaling parameter for the slope of the average transaction price on the bid and ask side. It is a linear, thus symmetric, approximation for small λ or quantity traded. As the empirical results in table 2 have shown, additional flexibility can be added by using two parameters for the slope of the average transaction price function λ^+, λ^- to account for distinct bid and ask liquidity. Then process (6) can be extended to

$$\begin{aligned}
& S_{t_{i-1}} \\
\to \quad & \bar{S}_{t_{i-1}} = S_{t_{i-1}} \left(e^{\lambda^+(H_{t_i}-H_{t_{i-1}})} I_{\mathbb{R}_+}(H_{t_i}-H_{t_{i-1}}) + e^{\lambda^-(H_{t_i}-H_{t_{i-1}})} I_{\mathbb{R}_-}(H_{t_i}-H_{t_{i-1}}) \right) \\
\to \quad & \begin{cases} uS_{t_{i-1}} \left(e^{\lambda^+(1-\alpha^+)(H_{t_i}-H_{t_{i-1}})} I_{\mathbb{R}_+}(H_{t_i}-H_{t_{i-1}}) + e^{\lambda^-(1-\alpha^-)(H_{t_i}-H_{t_{i-1}})} I_{\mathbb{R}_-}(H_{t_i}-H_{t_{i-1}}) \right) \\ dS_{t_{i-1}} \left(e^{\lambda^+(1-\alpha^+)(H_{t_i}-H_{t_{i-1}})} I_{\mathbb{R}_+}(H_{t_i}-H_{t_{i-1}}) + e^{\lambda^-(1-\alpha^-)(H_{t_i}-H_{t_{i-1}})} I_{\mathbb{R}_-}(H_{t_i}-H_{t_{i-1}}) \right) . \end{cases}
\end{aligned}$$

where λ^+, λ^-, α^+ and α^- are the bid and ask liquidity parameters, respectively, and for simplicity we set $\gamma = 0$.

However the increase in calibration flexibility is at the cost of computational requirements. Now, when valuing derivatives, even under exact replication, S_t becomes path-dependent and the resulting tree bushy. But again the two-stage dynamical program of appendix B can be applied and it will approximate the solution in finite time. For this purpose the self-financing condition (59) has to be replaced by

$$\begin{aligned}
& (H_{t_i}-H_{t_{i-1}}) \Big(e^{\lambda^+((H_{t_i}-H_0)-\alpha^+(H_{t_{i-1}}-H_0))} I_{\mathbb{R}_+}(H_{t_i}-H_{t_{i-1}}) \\
+ & e^{\lambda^-((H_{t_i}-H_0)-\alpha^-(H_{t_{i-1}}-H_0))} I_{\mathbb{R}_-}(H_{t_i}-H_{t_{i-1}}) \Big) u^{n-j} d^j \tilde{S}_{t_{i-1}} + (\hat{H}_{t_i}-\hat{H}_{t_{i-1}}) B_{t_{i-1}} \quad \le \quad 0, \qquad \forall i,j.
\end{aligned}$$

5.3 Stochastic liquidity

As is the case with Black-Scholes volatility, in practice, treating liquidity as a constant parameter is a convenient modelling assumption that, sometimes, admits tractable solutions. But in reality our empirical analysis has shown that both λ and γ are highly stochastic as figure 11 indicates. However, their dynamics seem to be reasonably stationary, at least over the short-term, so that a mean-reverting, non-negative diffusive process, as is standardly used for stochastic volatility modelling, could be appropriate. An easy model for, say, $\lambda_t(\omega)$ would be a mean-reverting Ornstein-Uhlenbeck process:

$$d\lambda_t = a(b-\lambda_t)dt + c\,dW_{\lambda_t}, \tag{49}$$

where a, b and c are constants. It is however well-known that following this process λ could become negative with positive probability. Another alternative that would keep λ positive would be a process similar to the one used in the [CIR] interest rate model:

$$d\lambda_t = a(b-\lambda_t)dt + c\sqrt{\lambda_t}\,dW_{\lambda_t}. \tag{50}$$

Again, to estimate the parameters a, b, c, representing the reversion speed, mean and volatility scalar, respectively, from our data set, we resort to ordinary least squares. For that purpose we took

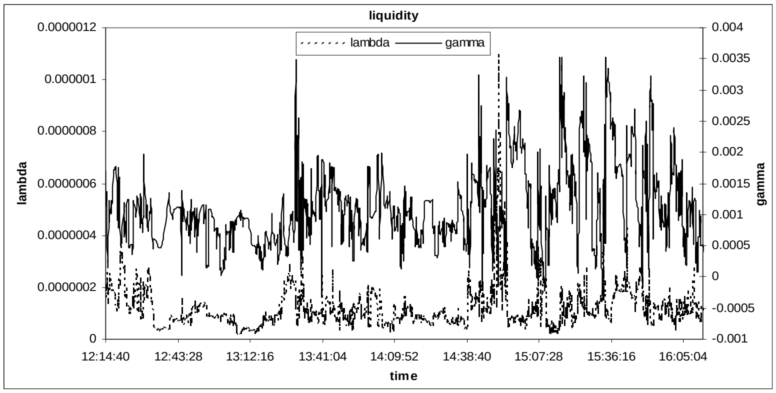


Figure 11: λ and γ tick time series for BASF over a specific time period.

minutely averages of the data and for (49) estimated the linear regression model

$$\begin{aligned}\lambda_{t+1} &= ab\delta t + (1 - a\delta t)\lambda_t + c\sqrt{\delta t}\epsilon_{t+1} \\ &= \alpha + \beta\lambda_t + \tilde{\epsilon}_{t+1},\end{aligned}$$

where $\epsilon_{t+1} \sim N(0,1)$ and $\tilde{\epsilon}_{t+1} \sim N(0, \sigma_\lambda^2)$, $\forall t$. Performing the least-squares fitting the best parameter estimates are then given by

$$a = \frac{1-\beta}{\delta t}, \qquad b = \frac{\alpha}{1-\beta}, \qquad c = \frac{\sigma_\lambda}{\sqrt{\delta t}}.$$

To estimate the parameters of (50) we first invoke Itô's formula and obtain

$$d\sqrt{\lambda_t} = \frac{1}{2\sqrt{\lambda_t}}\left(d\lambda_t - \frac{1}{4}c^2 dt\right) = \left(\frac{1}{2}ab - \frac{1}{8}c^2\right)\lambda_t^{-\frac{1}{2}}dt - \frac{1}{2}a\sqrt{\lambda_t}dt + cdW_{\lambda_t},$$

whose parameters we estimate through the linear regression model

$$\begin{aligned}\sqrt{\lambda_{t+1}} &= \left(1 - \frac{1}{2}a\delta t\right)\sqrt{\lambda_t} + \left(\frac{1}{2}ab - \frac{1}{2}c^2\right)\delta t\lambda_t^{-\frac{1}{2}} + c\sqrt{\delta t}\epsilon_{t+1} \\ &= \beta_1\sqrt{\lambda_t} + \beta_2\lambda_t^{-\frac{1}{2}} + \tilde{\epsilon}_{t+1},\end{aligned}$$

with the distribution of the errors as before, so that subsequently

$$a = \frac{2(1-\beta_1)}{\delta t}, \qquad b = \frac{\beta_2 + \frac{1}{8}\sigma_\lambda^2}{1-\beta_1}, \qquad c = \frac{\sigma_\lambda}{\sqrt{\delta t}}.$$

Table 3 gives the results of the least-squares fitting for both λ and γ as estimated through (39). The second model seems to fit marginally better for λ and vice versa for γ.

In general, when employing a diffusive process of type

$$d\theta_t = u(\theta_t, t)dt + v(\theta_t, t)dW_{\theta_t}$$

for a generic parameter θ and functions u, v a one-factor Black-Scholes type PDE would be extended by partial derivatives with respect to the new factor:

$$\mathcal{L}_{BS}V = -(u - mv)\frac{\partial V}{\partial \theta} - \frac{1}{2}v^2\frac{\partial^2 V}{\partial \theta^2} - \rho\sigma v S\frac{\partial^2 V}{\partial S\partial\theta} \tag{51}$$

lambda		BASF	Telekom	ThyKrupp	VW	Preussag
mr-OU	a	3489.13	5801.76	3729.64	2318.44	2851.14
	b	1.15E-07	7.45E-08	6.77E-08	1.33E-07	1.65E-07
	c	0.000636	0.000560	0.000292	0.000630	0.000634
	R^2	59.10%	37.56%	56.15%	71.53%	65.21%
CIR	a	3089.23	2758.61	2562.14	1941.02	2642.41
	b	1.03E-07	5.32E-08	6.08E-08	1.14E-07	1.53E-07
	c	0.81	1.02	0.55	0.84	0.75
	R^2	61.60%	37.80%	61.71%	72.77%	66.18%

gamma		BASF	Telekom	ThyKrupp	VW	Preussag
mr-OU	a	4498.49	8150.87	4257.65	4281.21	3002.14
	b	0.0011	0.0010	0.0014	0.0008	0.0012
	c	5.81	6.24	6.26	5.43	5.08
	R^2	49.26%	20.98%	51.29%	51.91%	63.75%
CIR	a	4833.58	7382.95	2712.53	2255.26	3438.60
	b	0.0010	0.0009	0.0013	0.0006	0.0011
	c	87.57	100.06	88.41	99.31	73.86
	R^2	47.21%	20.34%	46.47%	54.42%	61.23%

Table 3: Parameter estimates a, b, c for the respective processes of λ and γ, estimated on minutely data.

	BASF			Telekom			ThyKrupp			VW			Preussag		
	dS	*dlam*	*dgam*	*dS*	*dlam*	*dgam*	*dS*	*dlam*	*dgam*	*dS*	*dlam*	*dgam*	*dS*	*dlam*	*dgam*
dS(BASF)	1.00														
dlam(BASF)	-0.21	1.00													
dgam(BASF)	-0.19	-0.30	1.00												
dS(Telekom)	0.40	0.01	-0.18	1.00											
dlam(Telekom)	-0.03	-0.02	0.07	0.04	1.00										
dgam(Telekom)	0.12	-0.02	0.05	0.02	-0.16	1.00									
dS(ThyKrupp)	0.15	-0.04	-0.11	0.18	0.07	-0.03	1.00								
dlam(ThyKrupp)	-0.08	-0.08	-0.11	0.00	-0.05	-0.17	0.11	1.00							
dgam(ThyKrupp)	0.01	0.06	0.01	-0.02	-0.04	0.13	-0.21	-0.33	1.00						
dS(VW)	0.01	0.01	-0.02	0.01	0.00	0.03	0.05	0.01	-0.07	1.00					
dlam(VW)	-0.14	0.07	-0.01	-0.07	-0.03	0.08	-0.01	-0.12	0.03	0.25	1.00				
dgam(VW)	0.10	0.06	0.05	0.03	0.01	-0.02	0.08	-0.03	0.13	0.11	-0.31	1.00			
dS(Preussag)	0.01	0.02	-0.02	0.01	0.00	0.03	0.06	0.01	-0.07	1.00	0.25	0.12	1.00		
dlam(Preussag)	-0.01	0.02	-0.01	-0.05	-0.07	0.06	0.08	-0.09	-0.03	0.08	0.13	-0.01	0.09	1.00	
dgam(Preussag)	-0.05	-0.04	0.01	0.05	0.05	0.06	-0.04	0.17	0.11	0.33	0.09	0.08	0.34	-0.40	1.00

Table 4: Correlation matrix of dS, $d\lambda$ and $d\gamma$ for the stocks in our sample.

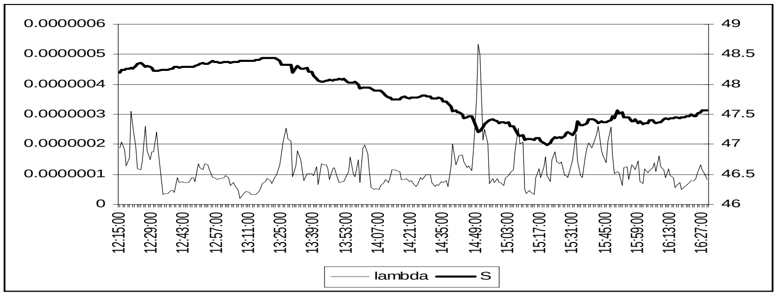


Figure 12: Minutely averaged λ (left axis) and spot mid-price of BASF over the reference time period.

where $\rho = E[dS_t d\theta_t]$ is the correlation between spot and parameter changes and $m(\theta_t, t)$ the market price of risk of the particular parameter. When including more than one additional stochastic parameter and/or more than one underlying, then (51) is extended by further cross-partial derivatives with their respective market prices of risk and correlation coefficients. Table 4 gives the estimated correlation matrix for our universe of underlyings and their respective processes of λ and γ. There seems to exist a significant negative correlation between a stock's γ and λ. This may be explained, that as orders are usually filled around the best quotes where, as already figure 3 indicated, the average transaction price function may be convex on the bid-side and concave on the ask-side, so that a widening of the bid-ask spreads leads to a flattening of the slope. The other correlations seem to be relatively low and with changing signs. A reason for it might be our relatively small dataset in terms of number of underlyings and time period. Overall this suggests that the cross-terms in the eventual PDE would be less significant. Additionally, by observing the magnitude of the a parameters, i.e. the time-scale of the reversion to the mean, it is obvious that it seems to be fast in comparison to either implied volatilities, standard option maturities or spot movements of the underlying, as figure 12 shows. For this type of process [FPS] show that this will allow asymptotic approximations when solving a PDE of type (51). However, the latter assumes that the liquidity parameter can be hedged. In theory, for this purpose, as is done by [Jar2], traded options can be introduced into the market, which also depend on the liquidity of the underlying and thus complete it. But the problematic part of this approach, which is often ignored, is that traded options themselves have finite liquidity and in practice it is usually far lower than that of the underlying. Therefore the transaction costs incurred when hedging liquidity will have a much larger effect than either leaving the position unhedged or, if possible, taking a static hedge up-front. Hence, realistically, the value of derivatives on an illiquid underlying may have to be given in terms of physical expectations, instead of risk-neutrality. It may thus prove useful if the employed stochastic model will be tractable in some form, as for instance our two chosen processes. In the next section we describe a class of derivatives that will isolate liquidity risk and propose a valuation approach under these assumptions.

5.4 A framework for liquidity derivatives

In general, liquidity derivatives isolate the risk exposure due to the unavailability of sufficient quantities of assets. [Scho] defines liquidity options as conferring the right to buy or sell a certain amount H of an asset at the quoted spot price S_t, exercisable within a prespecified time window $T - t_0$. Under perfect liquidity, this amounts to a Call or Put option with a strike price which is always exactly equal to the spot. Hence in a Black-Scholes world it would have no value. However when liquidity is finite this contract represents a direct insurance against or bet on the liquidity of an asset in the future. Under our liquidity model the payoff from the writer's point of view would be given by

$$C_\tau = -HS_\tau \left((1 + \mathrm{sign}(H - H_\tau)\gamma) e^{\lambda(H-H_\tau)} - 1 \right), \qquad t_0 \le \tau \le T, \tag{52}$$

where H_τ represents the accumulated amount of the asset up to time τ. On the other hand the mark to market value of what the holder would obtain upon exercise is

$$C_\tau = HS_\tau \left((1 + \mathrm{sign}(H - H_\tau)\gamma)^{(1-\alpha)} e^{\lambda(1-\alpha)(H-H_\tau)} - 1 \right), \qquad t_0 \le \tau \le T. \tag{53}$$

This asymmetry under physical delivery suggests that the holder would, most likely, exercise the option because of requirements other than selling the stock back to the market. This means that early exercise may be entirely random from the writer's point of view or might not occur at all and thus no dynamic hedging strategy may exist.

Instead, one possible static super-replication strategy would be for the writer to take a position of H in the asset immediately, in which case the upper bound for the contract premium would be given by the future value of the capital employed for the initial hedge, namely

$$C_{t_0} \le H(1 + \mathrm{sign}(H)\gamma) S_{t_0} e^{\lambda H} B_T. \tag{54}$$

This premium could then be further lowered by subtracting the physical expected value of the cash flow return, i.e.

$$\begin{aligned} C_{t_0} \quad = \quad & P(\text{exercise}) HE \left[(1 + \mathrm{sign}(H)\gamma) S_{t_0} e^{\lambda H} B_\tau - S_\tau e^{-r(\tau - t_0)} \right] \\ & -(1 - P(\text{exercise})) HE \left[(1 + \mathrm{sign}(-H)\gamma_T) S_T e^{-\lambda_T H} \right] e^{-r(T-t_0)} \end{aligned}$$

where, given no information about the holder's strategy, the exercise time τ and the probability of exercise $P(\cdot)$ would have to be guessed through for example a Bayesian prior. In this case uniform distributions for both would represent uninformative priors. Also the premium would be further lowered if the writer already held a surplus in the asset. A second hedging strategy would be to buy all of the position at once if exercised. Then, again under physical expectations, the value of this strategy is given by

$$C_{t_0} = P(\text{exercise}) HE \left[S_\tau \left((1 + \mathrm{sign}(H)\gamma_\tau) e^{\lambda_\tau H} - 1 \right) \right].$$

To calculate C_{t_0} under both hedging strategies requires solving an expectation of the form

$$E \left[S_t \gamma_t e^{\lambda_t} \right].$$

Under the assumption that S is log-normally distributed explicit solutions will only exist if λ is non-stochastic or normally distributed, e.g. of Ornstein-Uhlenbeck type, and γ non-stochastic or log-normal. But even if these conditions are not met it would be straightforward to calculate it by Monte-Carlo simulation. Instead, however, if the writer of this contract knew more about the holder's purpose in holding the asset, he could use a more-informative distribution for both exercise time and exercise probability and, possibly, there would be ways of dynamically replicating the contract. In general, this type of liquidity derivatives is closely related to portfolio or program trading and the next section discusses it as a further application.

Another class of liquidity derivatives has been proposed by [B&H], namely options on their own Greeks or on those of other options. In particular, options on the Delta or Gamma of another option represent liquidity derivatives in the sense that for the writer of the underlying option they effectively cap the required hedging quantity. When the latter, which is directly related to Delta and Gamma, becomes large then the writer transfers the exposure to the liquidity of the market to a third party. Another type of contract that was suggested were options with their own Gamma capped, i.e. they settle early when its theoretical Gamma Γ_V hits a boundary Γ_0 and hence protect the hedger against having to make large transactions in the market. In practice it would be difficult to agree on the true Γ_V of the contract, since the liquidity or transaction costs faced by various participants are different. But if instead the Black-Scholes Γ_{BS} and value V_{BS} were taken as the proxy for the barrier and early settlement value, respectively, then the cap is easier to verify. It also would make a contingent claim cheaper, because by (30) $\sigma_{BS} \leq \sigma_V(S_t, t; \lambda, \alpha)$.

5.5 Applications to portfolio trading

Portfolio or program trading is usually understood to be the liquidation or rebalancing of a large equity portfolio containing one or more stocks. In general the portfolio is assumed to be large enough to exceed the market depth at a particular time, so that trading all of it would substantially influence the stock price. It therefore, usually, has to be broken up into smaller chunks. This is done by an agent trader who will try to execute a trade (for the client) as advantageously as possible and may guarantee a price up-front. In this case, depending on the agreement, any surpluses or shortfalls are then either borne by the agent or by the client. If we resort to our liquidity model then the problem of finding an optimal guaranteed price is given by the following formulation.

Proposition 5.3 (Risk-neutral portfolio trade quote)
Under risk-neutrality the portofolio trade quote is the solution of the program

$$\max_{(H_t(\omega))_{\forall t,\omega}} K \tag{55}$$

subject to the pre-and post-position constraints

$$H_{t_0}^{(l)} = H_0^{(l)}, \qquad H_T^{(l)} = H^{(l)}, \qquad \forall l,$$

the conservation of funds constraints (47) and risk-neutral super-replication condition

$$E_Q\left[\frac{W_T}{B_T}\right] \geq 0, \tag{56}$$

where

$$W_T(\omega) = \hat{H}_T(\omega)B_T - K, \qquad \forall\omega \tag{57}$$

is the terminal net wealth. The processes $(B_t, S_t^{(l)}(\omega), \bar{S}_t^{(l)}(\omega))_{\forall t,\omega,l}$ *follow (2) and (44), respectively.*

Here (57) implies that both surpluses and shortfalls are borne by the agent and K is set such that it is costless to enter the deal. Instead if surpluses are returned to the client, then the objective function changes to

$$C_{t_0} = \min_{(H_t(\omega))_{\forall t,\omega}} E_Q\left[\frac{W_T}{B_T}\right]$$

and the wealth function (57) is replaced by

$$W_T(\omega) = \max(K - \hat{H}_T(\omega), 0), \qquad \forall\omega.$$

Thus the agent is short an option in the liquidation value of the portfolio and requires a premium from the client up-front.

Also, the liquidation price is often guaranteed in advance in terms of a spread around the volume weighted average price (vwap) of the portfolio over a period of time. Then K in the wealth function (57) would be replaced by

$$K + \sum_l \left(H_0^{(l)} - H^{(l)} \right) vwap^{(l)},$$

where

$$vwap^{(l)} = \frac{\sum_i \left(\left(H_{t_i}^{(l)}(\omega) - H_{t_{i-1}}^{(l)}(\omega) \right) \bar{S}_{t_{i-1}}^{(l)}(\omega) + f\left(\lambda_{t_{i-1}}^{l}(\omega), S_{t_{i-1}}^{(l)}(\omega) \right) \right)}{H_0^{(l)} - H^{(l)} + \sum_i f\left(\lambda_{t_{i-1}}^{l}(\omega), S_{t_{i-1}}^{(l)}(\omega) \right)},$$

where f is a function giving the volume due to other market participants, that would need to be estimated and calibrated.

There exists a trivial solution for the the program in proposition 5.3, when λ is constant and $\gamma = 0$. For simplicity we will assume that $l = 1$ and $r = 0$. By lemma 2.1 it is cheaper in terms of transaction costs to trade many small quantities instead of a large bulk. We thus conjecture that the optimal solution must require the trader to break up the order into equal chunks, that are traded in equal time periods, so that

$$\arg \max_{(H_t(\omega))_{\forall t, \omega}} K = \frac{(H_0 - H)(n - i)}{n}, \tag{58}$$

where n is the number of discrete time periods. Expanding (56) and substitution of (47) and (58) gives

$$E_Q[W_T] = E_Q\left[\hat{H}_T\right] - K = \sum_{i=1}^{n} \frac{(H_0 - H)}{n} E_Q[S_{t_{i-1}}] e^{\lambda(H_0 - H)/n} = n \frac{(H_0 - H)}{n} S_{t_0} e^{\lambda(H_0 - H)/n},$$

which in the continuous time limit as $n \to \infty$ converges to the perfect liquidity forward price. Since the latter represents an upper bound for K the solution is indeed optimal. It becomes apparent that if trading is continuous the liquidity effect is of a lower order.

In general, portfolio trading problems under this liquidity model will need to be solved numerically with bushy trees and dynamical programming and would thus be computationally intensive. But, in practice, since the quotes were not derived from non-arbitrage relationships, they would rather serve as a point of orientation for a price that would eventually be determined competitively. In fact, instead of the risk-neutrality condition (56) the papers by amongst other [B&L], [A&C1], [A&C2], [H&S2] suggest a framework of utility functions, where the variance of $\hat{H}_T$ is incorporated. Moreover, as mentioned in an earlier section, they employ arithmetic Brownian motion as the asset price processes, which allows for analytic solutions.

6 Summary and proposals for further research

In this paper we have presented a parametric market model featuring an observable proxy measure of liquidity for assets. We derived a risk-neutral model for derivatives valuation based upon it, both in discrete time and for the continuous-time limit. The liquidity impact function contained three parameters: γ measuring the relative width of the bid-ask spread, λ as a proxy for the slope of the average price as a function of quantity, and α giving the eventual market price slippage. All of these parameters are directly observable in a non-specialists market order book of layered best bid and ask quotes. As an example of the latter we empirically analysed German equity market data estimating the order of magnitude of the parameters and observing their stability. The model proved to fit the real world data very accurately and it allowed us to systematically rank the liquidity of a stock, defined as the coefficient of the relative market capitalisation traded leading to relative price slippage. Somewhat surprisingly we concluded that market cap and the elasticity of liquidity under this proxy seem to be negatively related. Also, as the parameters proved to be non-constant we presented some extensions to the model in the form of stochastic liquidity, multiple underlying assets and distinct bid and ask liqudity. The liquidity parameters appeared to be fast mean-reverting and were well explained by some standard types of these processes.

As a direct application we incorporated this liquidity impact function in an option pricing framework, we appended the CRR binomial model by a controlled process and formulated the price as the solution of a constrained nonlinear optimisation program. Depending on the type of option and whether the position was long or short the model generated unique, hence risk-neutral, bid and ask prices. Furthermore, by allowing for super-replication strategies for many types of contracts, as is the case for most transaction cost models, there existed parameter ranges where the solution was superior to exact replication. For this case we formulated a dynamical program that solved for the cheapest super-replication strategy. In the continuous time limit we derived three PDEs, each valid for a certain order of magnitude of the parameterisation. If the bid-ask spread was assumed to be a significant factor, then it dominated the elasticity effect and hedging would need to be done in discrete time. But if in this case there existed no price slippage then it reduced to the Hoggard-Whalley-Wilmott transaction costs model. Most significantly, when bid-ask spreads were assumed to be negligible and only the slope of the impact function was considered, then the option hedging strategy could be implemented in continuous time. The only care that needed to be taken was how to deal with non-smooth option payoffs. Also at least numerically, it should be straightforward to extend the model by a stochastic volatility framework, which is part of future research.

Furthermore we proposed the model as a framework for the valuation of liquidity options and portfolio trades and gave some examples. Another application, which creates a link to market microstructure theory models as e.g. [Kyle] and [F&J] would be to employ the model for strike, barrier, position detection after observing sequences of large trades. This would represent the inverse problem of the hedging options. Essentially this would entail trying to decompose observed volatility into predictable parts, as is for instance done by [L&W] or [P&D], through techniques like maximum likelihood analysis to filter out the Delta and Gamma of the large market participant. Finally, in general, we believe that liquidity represents an additional dimension in most markets. As figure 1 has shown a price time series can be extended by an additional axis to account for limited availability of quantities at particular prices. This idea can also be applied to, say, a interest rate yield curve or an implied volatility smile/skew. We are confident that this paper is an initial step in to this direction.

A Solving nonlinear systems of equations

For a generic terminal condition we have to solve the system of implicit nonlinear functions

$$\begin{aligned} g_1(H,\hat{H},\omega_{2j-1},T) &= H_{T_j}\tilde{u}^{n-j+1}d^{j-1}\tilde{S}_{T_j} + \hat{H}_{T_j}B_T - C(\tilde{u}^{n-j}\tilde{d}^j\tilde{S}_{T_j}) = 0, \\ g_2(H,\hat{H},\omega_{2j},T) &= H_{T_j}\tilde{u}^{n-j}\tilde{d}^j\tilde{S}_{T_j} + \hat{H}_{T_j}B_T - C(\tilde{u}^{n-j}\tilde{d}^j\tilde{S}_{T_j}) = 0, \end{aligned}$$

$j = 1, \ldots, n$, where j and n are the number of down and time steps, respectively,

$$\begin{aligned}
\tilde{S}_{t_i} &= S_{t_0} e^{\lambda(1-\alpha)(H_{t_i}(\omega) - H_0)}, \\
\tilde{u} &= u(1 \pm \gamma)^{(1-\alpha)}, \\
\tilde{d} &= d(1 \mp \gamma)^{(1-\alpha)},
\end{aligned}$$

the sign chosen depending whether Delta is positive or negative, and $H_T(\omega_{2j}) = H_T(\omega_{2j-1}) = H_{T_j}, \forall j$. The intermediate self-financing conditions span the system:

$$\begin{aligned}
g_1(H, \hat{H}, \omega_{2j-1}, t_i) &= (H_{t_i} - H_{t_{i-1}}) e^{\lambda(H_{t_i} - H_{t_{i-1}})} \left(1 + \mathrm{sign}\left(H_{t_i} - H_{t_{i-1}}\right) \gamma\right) \tilde{u}^{i-j-1} \tilde{d}^j \tilde{S}_{t_{i-1}} \\
&\quad + (\hat{H}_{t_i} - \hat{H}_{t_{i+1}}) B_{t_{i-1}} = 0, \\
g_2(H, \hat{H}, \omega_{2j}, t_i) &= (H_{t_i} - H_{t_{i-1}}) e^{\lambda((H_{t_i} - H_{t_{i-1}})} \left(1 + \mathrm{sign}\left(H_{t_i} - H_{t_{i-1}}\right) \gamma\right) \tilde{u}^{i-j-2} \tilde{d}^{j+1} \tilde{S}_{t_{i-1}} \\
&\quad + (\hat{H}_{t_i} - \hat{H}_{t_{i-1}}) B_{t_{i-1}} = 0,
\end{aligned}$$

$i = 1, \ldots, n, j = 1, \ldots, i-1$. In both cases to solve for the holding process $(H_t(\omega), \hat{H}_t(\omega))_{\forall t, \omega}$ we need an algorithm that will converge to the roots. One standard possibility is the Newton method

$$J\left([H^{(l+1)} \ \hat{H}^{(l+1)}]^T - [H^{(l)} \ \hat{H}^{(l)}]^T\right) = -[g_1 \ g_2]^T,$$

where

$$J = \begin{bmatrix} \nabla_H g_1 & \nabla_{\hat{H}} g_1 \\ \nabla_H g_2 & \nabla_{\hat{H}} g_2 \end{bmatrix}$$

is the Jacobian matrix, which, after rearrangement, results in

$$\begin{aligned}
H^{(l+1)} &= \frac{a \nabla_{\hat{H}} g_2 - b \nabla_{\hat{H}} g_1}{\nabla_H g_1 \nabla_{\hat{H}} g_2 - \nabla_H g_2 \nabla_{\hat{H}} g_1}, \\
\hat{H}^{(l+1)} &= \frac{a \nabla_H g_2 - b \nabla_H g_1}{\nabla_{\hat{H}} g_1 \nabla_H g_2 - \nabla_{\hat{H}} g_2 \nabla_H g_1},
\end{aligned}$$

where

$$a = \hat{H}^{(l)} \nabla_{\hat{H}} g_1 + H^{(l)} \nabla_H g_1 - g_1, \qquad b = \hat{H}^{(l)} \nabla_{\hat{H}} g_2 + H^{(l)} \nabla_H g_2 - g_2,$$

and $l = 0, 1, \ldots$ is the number of iterations, which is chosen so that the difference between the parameter values is smaller than a arbitrary small constant ϵ, i.e.

$$\left| H^{(l)} - H^{(l-1)} \right| \leq \epsilon.$$

B Solving the dynamical program

Firstly, the space of feasible trading strategies $(H_{t_i}(\omega_j), \hat{H}_{t_i}(\omega_j))_{\forall i,j}$ needs to be discretised into a matrix of possible stock and bond holdings at every node of the tree. Because the fundamental asset price tree is recombining and the observed price tree Markovian when $H_{t_i}(\omega_j)$ is known, the number of matrices will only grow quadratically in time. Arbitrarily choosing the probability of up and downsteps as 0.5, gives the following backward recursion equation:

$$W_{t_i}(\omega_j) = \min_{(H_{t_i}(\omega_j), \hat{H}_{t_i}(\omega_j))} \frac{1}{2} \sum_j W_{t_{i+1}}(\omega_j),$$

subject to the terminal conditions

$$W_T(\omega_j) = \max\left(C_T(\omega_j) - H_T(\omega_j)S_T(\omega_j) - \hat{H}_T(\omega_j)B_T, 0\right), \qquad j = 1, \ldots, n,$$

and the self-financing constraints

$$\begin{aligned}(H_{t_i} - H_{t_{i-1}})&e^{\lambda(H_{t_i} - H_{t_{i-1}})}\left(1 + \operatorname{sign}\left(H_{t_i} - H_{t_{i-1}}\right)\gamma\right)\tilde{u}^{i-j-1}\tilde{d}^j\tilde{S}_{t_{i-1}} \\ &+(\hat{H}_{t_i} - \hat{H}_{t_{i-1}})B_{t_{i-1}} \leq 0, \qquad j = 0, \ldots, i-1; i = 1, \ldots, n. \end{aligned} \tag{59}$$

C Various proofs

Proof: (of theorems 3.1, 3.2 and corollary 3.1)
For (30) to hold, it must be invariant under the choice of starting point within a time interval. In turn we will commence with an already hedged portfolio that will need to be rehedged after observing the asset price diffusion and then derive the same result for a portfolio, which still needs to be rehedged before an asset price diffusion. As before, for notational convenience, we will drop the time subscript.

1. If, initially, we assume that H is already the correct hedge quantity, then we start with a position

$$\Pi = V(S, t; H) - HS = \hat{H}B, \tag{60}$$

where we observe that the right-hand side cash position equals exactly the value of the left-hand side contract and stock portfolio. Also we explicitly show the dependence of the contract value V on the quantity of stock held H to distinguish the change in V due to one's own trading, i.e. a change in H, from the exogenous diffusion of the asset, which is assumed to follow geometric Brownian Motion. Then, the left-hand side of (60), i.e. the mark-to-market value of the contract and the stock, over the small but not infinitesimal time interval δt, evolve as

$$\Pi + \delta\Pi = V(S + \delta S, t + \delta t; H) - H(S + \delta S)$$

due to the exogenous diffusion of the asset. Subsequently the portfolio is rehedged to

$$\begin{aligned} &\bar{\Pi} + \delta\bar{\Pi} \\ = \;& V(S + \delta S, t + \delta t; H + \delta H) - (H + \delta H)(S + \delta S)(1 + \operatorname{sign}(\delta H)\gamma)^{(1-\alpha)}e^{\lambda(1-\alpha)\delta H} \\ = \;& V\left((S + \delta S)(1 + \operatorname{sign}(\delta H)\gamma)^{(1-\alpha)}e^{\lambda(1-\alpha)\delta H}, t + \delta t; H\right) \\ & -(H + \delta H)(S + \delta S)(1 + \operatorname{sign}(\delta H)\gamma)^{(1-\alpha)}e^{\lambda(1-\alpha)\delta H}, \end{aligned} \tag{61}$$

where in the last line we use the relationship

$$V(S, t; H + \delta H) = V\left(S(1 + \operatorname{sign}(\delta H)\gamma)^{(1-\alpha)}e^{\lambda(1-\alpha)\delta H}, t; H\right), \tag{62}$$

incorporating the dependence of V on the current holding H of S under the chosen form of the price impact function. Now, we conjecture that

$$H = \frac{\partial V}{\partial S}, \tag{63}$$

and δH is small so that for a generic constant c we can expand the exponential term

$$e^{c\lambda(1-\alpha)\delta H} = 1 + c\lambda(1-\alpha)\delta H + \frac{c^2}{2}\left(\lambda(1-\alpha)\delta H\right)^2 + \ldots. \tag{64}$$

in a Taylor series. Substituting (63) into (61), and invoking both (64) and Itô's formula, gives

$$\begin{aligned}
&\bar{\Pi} + \delta\bar{\Pi} \\
= \;& V - S\frac{\partial V}{\partial S} - \delta H(S+\delta S)(1+\mathrm{sign}(\delta H)\gamma)^{(1-\alpha)}\left(1+\lambda(1-\alpha)\delta H + \frac{1}{2}\lambda^2(1-\alpha)^2\delta H^2\right) + \frac{\partial V}{\partial t}\delta t \\
& + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\left((S+\delta S)(1+\mathrm{sign}(\delta H)\gamma)^{(1-\alpha)}\left(1+\lambda(1-\alpha)\delta H + \frac{1}{2}\lambda^2(1-\alpha)^2\delta H^2\right) - S\right)^2. \quad (65)
\end{aligned}$$

The leading order term in the brackets in equation (65) is

$$S(1+\mathrm{sign}(\delta H)\gamma)^{(1-\alpha)} - S = O(1),$$

thus at least one order of magnitude larger than the other terms, which we thus initially choose to ignore. Also, the rehedging quantity to leading order is

$$\delta H = \delta\left(\frac{\partial V}{\partial S}\right) = \frac{\partial^2 V}{\partial S^2}\delta S. \tag{66}$$

If we conjecture that

$$\frac{\partial^2 V}{\partial S^2} \ll O(\delta t),$$

then δH as a coefficient becomes negligible and substitution into (65) results in

$$\bar{\Pi} + \delta\bar{\Pi} = V - S\frac{\partial V}{\partial S} + \frac{\partial V}{\partial t}\delta t + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\left(S(1+\mathrm{sign}(\delta H)\gamma)^{(1-\alpha)} - S\right)^2. \tag{67}$$

Now the right-hand side of (60), i.e. the cash portion, after asset price diffusion and rehedging changes to

$$\bar{\Pi} + \delta\bar{\Pi} = (1+r\delta t)(V-HS) - \delta H(S+\delta S)(1+\mathrm{sign}(\delta H)\gamma)e^{\lambda\delta H}, \tag{68}$$

where again we choose to ignore the term multiplied by δH, so that

$$\bar{\Pi} + \delta\bar{\Pi} = (1+r\delta t)(V-HS). \tag{69}$$

Now, because (67) still has a random term due to δS we need to take expectations. By noting that the drift terms of the Brownian Motion are of a lower order, we do not need to worry about the probability measure and thus to leading order

$$\begin{aligned}
E\left[\left(1+\mathrm{sign}\left(\frac{\partial^2 V}{\partial S^2}\delta S\right)\gamma\right)^{(1-\alpha)}\right] &= \int_0^\infty \left(1+\mathrm{sign}\left(\frac{\partial^2 V}{\partial S^2}\delta S\right)\gamma\right)^{(1-\alpha)}\phi(S)dS \quad (70)\\
&= \frac{1}{2}\left((1+\gamma)^{(1-\alpha)} + (1-\gamma)^{(1-\alpha)}\right), \quad (71)
\end{aligned}$$

where $\phi(S)$ is the log-normal density. Thus equating the expectations of (67) and (68) and substituting (71) yields the required result (26) of theorem 3.1. It can be checked that $\partial^2 V/\partial S^2$ turns out to be $O(\sqrt{\delta t}\exp(-1/\sqrt{\delta t}))$.

But now if $\alpha = 1$ the only term remaining in the brackets of equation (65) is δS, so that the new left-hand side to leading order Gamma is

$$\bar{\Pi} + d\bar{\Pi} = V - S\frac{\partial V}{\partial S} + \left(\frac{\partial V}{\partial t} - \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}\right)\delta t \tag{72}$$

On the right-hand side however, if we allow for a second-order Gamma term we obtain

$$\bar{\Pi} + \delta\bar{\Pi} = (1 + r\delta t)(V - HS) - \gamma \left| \frac{\partial^2 V}{\partial S^2} \delta S \right|, \tag{73}$$

by observing that

$$\operatorname{sign}\left(\frac{\partial^2 V}{\partial S^2} \delta S \right) \frac{\partial^2 V}{\partial S^2} \delta S = \left| \frac{\partial^2 V}{\partial S^2} \delta S \right|.$$

Taking expectations of (73) and equating it to (83), yields the required result (28) of corollary 3.1, noting that

$$E[|\delta S|] = \sqrt{\frac{2\delta t}{\pi}} \sigma S$$

and also now

$$\frac{\partial^2 V}{\partial S^2} = O\left(\delta t^{\frac{1}{4}} \exp\left(-\delta t^{-\frac{1}{4}} \right) \right).$$

Finally if $\gamma = 0$ then we can take the continuous time limit, so that $\delta t \to dt$. Then the left-hand side keeps all the terms with λ and we obtain

$$\bar{\Pi} + d\bar{\Pi} = V - S\frac{\partial V}{\partial S} - S\frac{\partial^2 V}{\partial S^2} dS + \left(\frac{\partial V}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{1}{2}\lambda^2 (1-\alpha)^2 \sigma^2 S^4 \left(\frac{\partial^2 V}{\partial S^2} \right)^3 \right) dt. \tag{74}$$

Now the right-hand side of (60), i.e. the cash portion, after asset price diffusion and rehedging changes to

$$\bar{\Pi} + d\bar{\Pi} = (1 + rdt)(V - HS) - dH(S + dS)e^{\lambda dH}.$$

Substituting (63), (66) and expanding the exponential term in a Taylor series yields

$$\bar{\Pi} + d\bar{\Pi} = (1 + rdt)\left(V - S\frac{\partial V}{\partial S} \right) - S\frac{\partial^2 V}{\partial S^2} dS - \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} dt - \lambda\sigma^2 S^3 \left(\frac{\partial^2 V}{\partial S^2} \right)^2 dt. \tag{75}$$

Finally, equating (74) and (75), after rearrangement, gives the desired result (30) of theorem 3.2.

2. If, instead, we assume that we hold an amount H which does not, yet, represent the correct hedge quantity, but is close, then we obtain the same result. We demonstrate it only for the case where $\gamma = 0$, but therefore in more detail. By the standard hedged portfolio argument, we hold

$$\Pi = V(S, t; H) - HS = \hat{H}B, \tag{76}$$

Since, as assumed, we still need to rehedge at time t, the newly balanced $\bar{\Pi}$ is composed of

$$\bar{\Pi} = V(S, t; H + dH) - (H + dH)Se^{\lambda(1-\alpha)dH}. \tag{77}$$

Subsequently, the fundamental risky asset price will change by dS. Hence (77) evolves into

$$\bar{\Pi} + d\bar{\Pi} = V(S + dS, t + dt; H + dH) - (H + dH)(S + dS)e^{\lambda(1-\alpha)dH}. \tag{78}$$

Therefore, applying (62) to (78) and invoking Itô's formula gives:

$$\begin{aligned} \bar{\Pi} + d\bar{\Pi} &= V\left((S + dS)e^{\lambda(1-\alpha)dH}, t + dt; H \right) - (H + dH)(S + dS)e^{\lambda(1-\alpha)dH} \\ &= V + \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S}\left((S + dS)e^{\lambda(1-\alpha)dH} - S \right) \\ &\quad + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\left((S + dS)e^{\lambda(1-\alpha)dH} - S \right)^2 - (H + dH)(S + dS)e^{\lambda(1-\alpha)dH}. \end{aligned} \tag{79}$$

It becomes apparent that by choosing the rebalancing quantity as

$$dH = \frac{\partial V}{\partial S} - H \tag{80}$$

results in

$$\begin{aligned}\bar{\Pi} + d\bar{\Pi} \quad &= \quad V - S\frac{\partial V}{\partial S} + \frac{\partial V}{\partial t}dt + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\Big((S^2 + 2SdS + \sigma^2 S^2 dt)e^{2\lambda(1-\alpha)\left(\frac{\partial V}{\partial S} - H\right)} \\ &\quad -2(S^2 + SdS)e^{\lambda(1-\alpha)\left(\frac{\partial V}{\partial S} - H_t\right)} + S^2\Big).\end{aligned} \tag{81}$$

If we now assume that the trading process started at a time $t_0 < t$, then

$$\frac{\partial V}{\partial S} - H \approx d\left(\frac{\partial V}{\partial S}\right) \approx \frac{\partial^2 V}{\partial S^2}dS^- = O(\sqrt{dt}), \tag{82}$$

where dS^- is the continuous-time equivalent of the diffusion at the preceding timestep. Substituting (82) as well as (64) for $c = 1$ and $c = 2$ into (81), noting that due to the independence of the increments of Brownian motion

$$dSdS^- = 0,$$

after rearrangement results in

$$\bar{\Pi} + d\bar{\Pi} = V - S\frac{\partial V}{\partial S} + \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{1}{2}\lambda^2(1-\alpha)^2\sigma^2 S^4\left(\frac{\partial^2 V}{\partial S^2}\right)^3\right)dt. \tag{83}$$

As the change in the contract and stock portfolio turns out to be entirely risk-free, it must be equal to the predictable change in the cash, i.e. the right-hand side of (76). Subtracting transaction costs from the latter the new cash position is

$$(\hat{H} + d\hat{H})B = V(S, t; H) - HS - dHSe^{\lambda dH},$$

which after the time-interval dt changes to

$$(\hat{H} + d\hat{H})(B + dB) = (1 + rdt)V(S, t; H) - HS - dHSe^{\lambda dH}.$$

Expanding the transaction cost term in a Taylor series, substituting (82) and rearranging, we arrive at

$$\begin{aligned}&(\hat{H} + d\hat{H})(B + dB) \\ = \quad &(1 + rdt)\left(V - HS - \left(\frac{\partial V}{\partial S} - H\right)S\left(1 + \lambda\left(\frac{\partial V}{\partial S} - H\right) + \frac{1}{2}\lambda^2\left(\frac{\partial V}{\partial S} - H\right)^2\right)\right) \\ = \quad &(1 + rdt)\left(V - S\frac{\partial V}{\partial S} - \lambda S\left(\frac{\partial V}{\partial S} - H\right)^2 + \frac{1}{2}\lambda^2 S\left(\frac{\partial V}{\partial S} - H\right)^3\right) \\ = \quad &V - S\frac{\partial V}{\partial S} - \lambda\sigma^2 S^3\left(\frac{\partial^2 V}{\partial S^2}\right)^2 dt + r\left(V - S\frac{\partial V}{\partial S}\right)dt.\end{aligned} \tag{84}$$

Thus equating (83) and (84), after rearrangement, leads to the governing PDE (30) for the contract value V and completes the proof of theorem 3.2.□

References

[A&C1] Almgren & Chriss, 1999, *Value under liquidation*, Risk, Dec.

[A&C2] Almgren & Chriss, 2001, *Optimal Execution of Portfolio Transactions*, Journal of Risk, 3, 5-39

[B&H] Bakstein & Howison, 2001, *Using Options on Greeks as Liquidity Protection*, University of Oxford, Working Paper

[BLPS] Bensaid, Lesne, Pagès & Scheinkman, 1992, *Derivative Asset Pricing with Transaction Costs*, Mathematical Finance, 2, 63-86

[B&L] Bertsimas & Lo, 1998, *Optimal Control of Execution Costs*, Journal of Financial Markets, 1, 1-50

[BHK] Breen, Hodrick & Korajczyk, 2000, *Predicting Equity Liquidity*, Kellog Graduate School of Management, Working Paper

[B&S] Black & Scholes, 1973, *The Pricing of Options and Corporate Liabilities*, Journal of Political Economy, 81, 637-654

[B&V] Boyle & Vorst, 1992, *Option Replication in Discrete Time with Transaction Costs*, The Journal of Finance, 47, 271-293

[CRS] Chordia, Roll & Subrahmanyam, 2000, *Order Imbalance, Liquidity and Market Returns*, UCLA, Working Paper

[Con] Constantinides, 1986, *Capital Market Equilibrium with Transactions Costs*, Journal of Political Economy, 94, 842-862

[Con2] Constantinides, 1997, *Transaction Costs and the Pricing of Financial Assets*, Multinational Finance Journal, 2, 93-99

[CIR] Cox, Ingersoll & Ross, 1985, *A Theory of the Term Structure of Interest Rates*, Econometrica, 53, 385-407

[CRR] Cox, Ross & Rubinstein, 1979, *Option Pricing: A Simplified Approach*, Journal of Financial Economics, 7, 229-263

[DFIS] Daniels, Farmer, Iori & Smith, 2002, *Demand Storage, Market Liquidity and Price Volatility*, Santa Fe Institute, Working Paper

[ENU] Edirisinghe, Naik & Uppal, 1993, *Optimal Replication of Options with Transactions Costs and Trading Restrictions*, Journal of Financial and Quantitative Analysis, 28, 117-138

[F&J] Farmer & Joshi, 2000, *The price dynamics of common trading strategies*, Santa Fe Institute, Working Paper

[Frey] Frey, 1996, *The Pricing and Hedging of Options in Finitely Elastic Markets*, University of Bonn, Discussion Paper

[F&S] Frey & Stremme, 1995, *Market Volatility and Feedback Effects from Dynamic Hedging*, University of Bonn, Discussion Paper

[FPS] Fouque, Papanicolaou & Sircar, 1999, *Financial Modeling in a Fast Mean-Reverting Stochastic Volatility Environment*, Asia-Pacific Financial Markets,6, 37-48

[HWW] Hoggard, Whalley & Wilmott, 1994, *Hedging Option Portfolios in the Presence of Transaction Costs*, Adv. Fut. Opt. Res., 7, 21-35

[H&S1] Huberman & Stanzl, 2000, *Arbitrage-free Price-update and Price-Impact Functions*, Yale, Working Paper

[H&S2] Huberman & Stanzl, 2000, *Optimal Liquidity Trading*, Yale, Working Paper

[Jar1] Jarrow, 1992, *Market Manipulation, Bubbles, Corners, and Short Squeezes*, Journal of Financial and Quantitative Analysis, 27, 311-336

[Jar2] Jarrow, 1994, *Derivative Security Markets, Market Manipulation and Option Pricing Theory*, Journal of Financial and Quantitative Analysis, 29, 241-261

[Krak] Krakovsky, 1999, *Pricing Liquidity into Derivatives*, Risk, Dec.

[Kyle] Kyle, 1985, *Continuous Auctions and Insider Trading*, Econometrica, 53, 1315-1336

[Lel] Leland, 1985, *Option Pricing and Replication with Transactions Costs*, The Journal of Finance, 40, 1283-1301

[L&W] Lo & Wang, 1995, *Implementing Option Pricing Models When Asset Returns Are Predictable*, The Journal of Finance, 50, 87-129

[Lon] Longstaff, 1995, *Option Pricing and the Martingale Restriction*, Review of Financial Studies, 8, 1091-1124

[Per] Persaud, 2000, *The Liquidity Puzzle*, Risk, June

[P&D] Putyatin & Dewynne, 1999, *Market Liquidity and its Effects on Option Valuation and Hedging*, Philosophical Transactions of the Royal Society, 357, 2093-2108

[Rut] Rutkowski, 1998, *Optimality of Replication in the CRR Model with Transaction Costs*, Applicationes Mathematicae, 25, 29-53

[Scho] Scholes, 1999, *Liquidity Options*, Risk, Nov.

[Schö] Schönbucher & Wilmott, 2000, *The Feedback Effect of Hedging in Illiquid Markets*, SIAM Journal of Applied Mathematics, 61, 232-272

[S&P] Sircar & Papanicolaou, 1998, *General Black-Scholes Models Accounting for Increased Market Volatility from Hedging Strategies*, Applied Mathematical Finance, 5, 45-82

[SSC] Soner, Shreve & Cvitanić, 1995, *There is No Nontrivial Hedging Portfolio for Option Pricing with Transaction Costs*, Annals of Applied Probability, 5, 327-355

[Tal] Taleb, 1997, *Dynamic Hedging*, Wiley

[W&W] Whalley & Wilmott, 1997, *An Asymptotic Analysis of an Optimal Hedging Model for Option Pricing with Transaction Costs*, Mathematical Finance, 7, 307-324