

A Non-Arbitrage Liquidity Model with Observable Parameters 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 via a combination of a trader's individual transaction cost and a price slippage impact, which is felt by all market participants. 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; the nonlinear nature of liquidity leads to 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, including stochastic liquidity.

Keywords: liquidity, option pricing, transaction costs

*Work supported by the EPSRC, Charterhouse, Socrates and ESF

†Contact: howison@maths.ox.ac.uk, tel +44(0)1865 270500, fax +44(0)1865 270515; the authors are grateful for comments and assistance from Michael Mitton, George Papanicolaou, Hyun Shin, Jeff Dewynne, Simon Acomb and Conor Campbell.

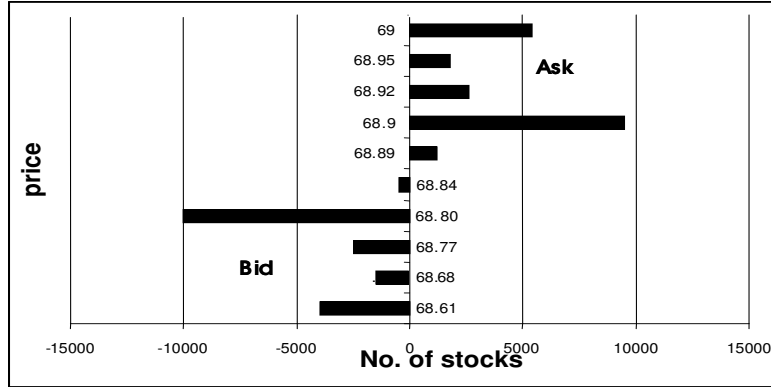


Figure 1: An example of an order book of Deutsche Telekom stock on 13 Jan 2000 at 12:14:41pm.

1 Introduction

Two of the underlying assumptions of, among others, the basic Black-Scholes or CAPM economies are firstly that markets are frictionless, and secondly that agents are price-takers. The former assumption implies the absence of transaction costs and the latter that no single participant's trading strategies can affect asset prices in the market. In practice, however, markets substantially deviate from these assumptions because for virtually all traded assets there exist both bid-ask spreads and a limited market depth, i.e. limited availability of the asset. In general, the former represent a revenue source as well as a risk insurance buffer for market makers, who attempt to buy low and sell high, and the latter is the volume of an asset available to buy or sell at a particular price. Together they represent the order book of an asset, which serves as the market inventory with respect to which the execution of the order is carried out. Figure 1 is a snapshot of an order book¹ in an electronic equity market for a particular stock.

In theory and in practice, if there are numerous market makers and other participants willing 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.² Obviously market participants prefer markets or assets with high liquidity, because they can get in and out of their positions quickly and cheaply, however large a fraction of the market they constitute. Hence intuitively, say, the equity of a large company that has a high free float and trades on a well-regulated exchange will, ceteris paribus, have

¹Even if there does not exist a centralised exchange for an asset, unless there is only one specialist dealer, an order book can always be constructed by layering the quotes of all the participants.

²Other possible factors that might affect liquidity are the availability of information about the traded asset, its free float, the legislation of the market itself, etc.

narrower spreads, compared to a closely held small-cap in an over-the-counter (OTC) market.³ Also the more liquid a market the more approximately linearly scalable will trading strategies in it be. So, for example the cost of trading one million stocks will amount to one million times the cost of transacting one single stock. The same considerations will also apply to contingent claims written on that asset. Because the value of a derivative, both originally in the Black-Scholes theory (see [B&S]) and at least partly in practice, 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 (see e.g. [Fig] or [Lon]). Generally, however, there is no consensus approach how to calculate liquidity premia of derivatives, nor how to parameterise, measure or observe the liquidity of a market or an asset.

Qualitatively liquidity, or the lack of it, causes two effects. Firstly, it has an impact on the average transaction price per unit of asset traded. Whereas it may be possible to trade small quantities of an asset at the best possible price on offer at a particular time, which ought to be 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 the 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,⁴ large trades of one agent may remove entire price layers from the order book. Moreover, it is usually the norm that the price taken from the last layer of market depth tapped will be published, i.e. appears as the official last price traded, albeit possibly only for a short time. Subsequently this may lead to market participants adjusting their prices or entering new quotes. Thus in effect the market has been moved by a single trader. Figure 2 shows the average transaction prices and next last price traded as functions of the traded quantity derived from a snapshot of an order book.

In reality, it is common that asset prices are pushed, in some cases deliberately, in a certain direction by comparatively large trades⁵ (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, for instance,

³In this case the more liquid stock should also trade at a fundamental premium. But in this paper equity valuation will be considered an exogenous factor.

⁴As is for instance the case on many European stock exchanges and also on the American NASDAQ, it is assumed that all participants either deal in a centralised transparent order-driven or electronically-linked broker market, instead of a specialists' market as is for example the NYSE.

⁵One case where this seems to happen regularly are so-called 'triple witching days' when traded futures and options expiry almost simultaneously. One such incident on the London Stock Exchange in September 2002 will be described in more detail later on in the paper.

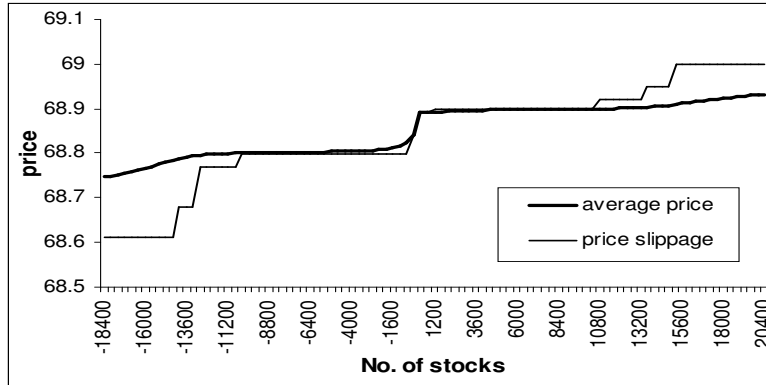


Figure 2: The exact average transaction price and resulting price slippage (last price traded) as a function of no. of stocks traded for the order book of Figure 1. The flat parts of the price slippage curve represent the price layers of the order book.

execute a large portfolio trade or stop-loss, or to set up or unwind a derivatives hedge. In the latter case if, following the Black-Scholes theory, they attempt to Delta-hedge, the Delta and Gamma, i.e. the amount of the underlying they have to hold and approximately add or remove respectively, may become large close to expiries and/or payoff discontinuities of contracts in their portfolios. Since 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 from the risk-free amount that can be earned on a replicating portfolio. Beyond this, a trader who has a good intuition of the liquidity of the market may, instead of hedging a position, be inclined to liquidate the accumulated hedge quantity and thus push the market in a desired direction. Normally, traders are not supposed to know the positions of other participants in the market; however, if they 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 it would be possible to exploit this 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 analytically capture these various effects of finite liquidity on asset markets, in particular in relation to contingent claims, the papers of [Lel], [HWW] model continuous-time and [B&V], [ENU], [BLPS] discrete-time transaction costs. For the latter they consider proportional, fixed and sign-dependent models and calculate optimal hedging strategies and valuations of derivatives. Extending

these models to the price impact effect in discrete time [G&L], [Jar1], [Jar2], and in continuous time [S&W], [Frey], [F&S], [S&P], [Krak], propose a number of similar approaches. Firstly, they introduce a reaction function that models the immediate impact of a large trade on the price of an asset as a function of both a liquidity scaling parameter and the trade size. They thereby derive non-arbitrage conditions in the market and also show that large trades lead to an increase in realised volatility of an asset. Secondly, the papers by [A&C1], [A&C2], [H&S1], [H&S2] and [B&L] in both a continuous and discrete-time framework further distinguish between a temporary price impact and a permanent slippage effect on the asset, by making its new equilibrium price a function of both the pre- and post-trade transaction prices. However, none of these papers provide ways of observing liquidity in practice, nor do they provide empirical evidence in support of their models.

In this paper a combination of existing and new transaction cost and slippage effect models is employed and constructed in such a way that the parameters involved are *observable* (that is, directly estimatable from order book data, rather than needing to be backed out from derivatives prices) given a particular order book of an asset. Commencing in Section 2, the discrete-time version of the model is derived, which is based on binomial trees of [CRR] extended by a nonlinear controlled process that represents the transaction cost and slippage effects, which is subsequently applied to the valuation of options. In Section 3 a number of nonlinear and quasilinear partial differential equations (PDEs) for both the value and the hedge quantities are derived in the continuous-time limit, under particular choices for the parameter values. In Section 4 an empirical analysis of an order book for various stocks traded on the German Xetra system is performed. This then allows for a consistent definition of liquidity and a calibration of the model to the data. In Section 5 various extensions to the model like multiple assets and stochastic liquidity are presented. Finally, Section 6 summarises the key findings and suggests further areas of research stemming from the model.

2 The discrete-time model

The main building block of a valuation framework for derivatives or portfolio trades is a suitable model of the underlying asset prices or, more generally, the state variables. Thus the initial setting of this analysis is a discrete-time finite-horizon economy where trading in assets takes place at times $\{t_0, t_1, \dots, t_n = T\}$. The state of the economy is given by the finite set of trajectories $\Omega = \{\omega_1, \dots, \omega_m\}$ and the step-wise revelation of the true state by the filtration, i.e. increasing sequence of σ -algebras, $(\mathcal{F}_t)_{t \in \{t_0, \dots, T\}}$. 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 $(\mathcal{F}_t)_{t \in \{t_0, \dots, T\}}$.

Resorting to the widely used binomial trees of [CRR], randomness, which represents other participants’ trading in the stock, is modelled 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 $t_i - t_{i-1}$, $i = 1, \dots, n$. Therefore

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

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

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

Two notable properties of the model are, firstly, the absence of arbitrage in the market provided that $0 < d < 1 + r < u$. By the fundamental theorem of asset pricing⁶ (see, e.g., [M&R]) this implies that expectations of the discounted risky asset have to be taken with respect to the risk-neutral probability measure $Q \equiv \{q, 1 - q\}$ and form a martingale when taking the bond as numeraire asset:

$$E_Q[S_T | \mathcal{F}_{t_0}] = S_{t_0} B_T / B_{t_0}, \quad T > t_0. \quad (3)$$

Secondly, u , d and q can be calibrated⁷ such that S_t follows (risk-neutral) geometric Brownian motion

$$dS_t = rS_t dt + \sigma S_t dW_t, \quad (4)$$

in the continuous-time limit: $t_i - t_{i-1} \rightarrow dt$. Here σ is the stock’s exogenous volatility and dW_t a Wiener process, i.e. the increments of standard Brownian motion.

On top of the random process for the underlying stock, we construct a controlled process that represents the effect that an individual, possibly large or influential, trader has on the market. This trader’s holding process in the risky asset is denoted 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 $(\mathcal{F}_t)_{t \in \{t_0, \dots, T\}}$ and are one-step-ahead predictable with respect to it. The latter point entails that the trader’s portfolio can be rebalanced in between the random changes to the underlying asset. Now, when assuming 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

⁶It states the equivalence of the absence of arbitrage opportunities in a stochastic model and the existence of an equivalent, under market completeness unique, martingale probability measure.

⁷In fact [CRR] choose $u = e^{\sigma\sqrt{dt}}$, $d = u^{-1}$ and $q = (e^{r dt} - d)/(u - d)$.

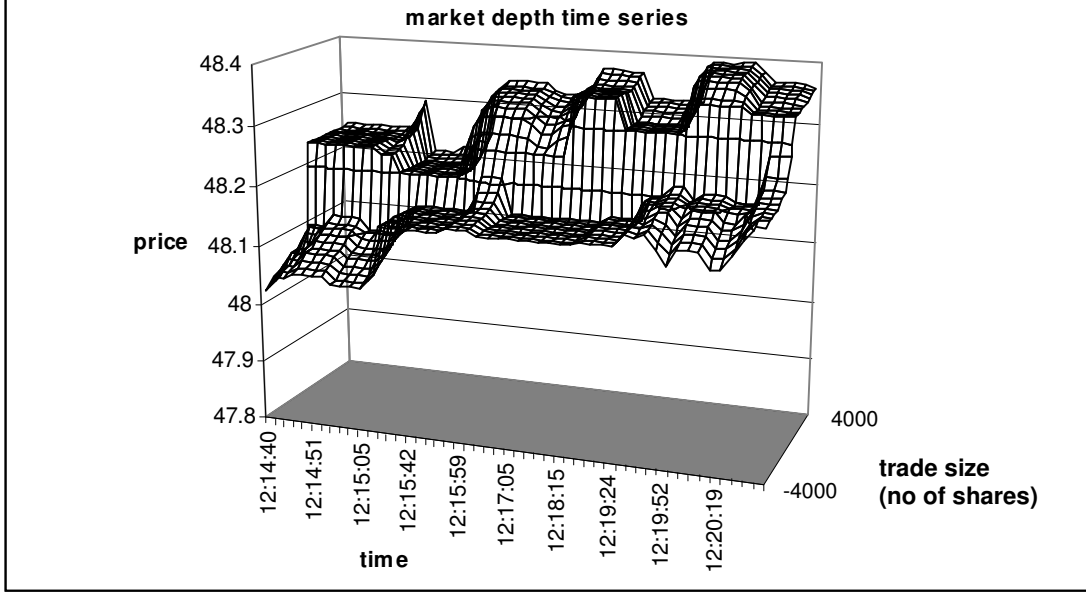


Figure 3: Average transaction price time series for BASF stock in a particular trading period derived from its electronic order book on the Xetra exchange. At the beginning of the time period shown liquidity is high on the ask side (the surface is nearly flat, so that 1 to 4000 stocks can be bought at 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’. In our model γ measures half the average height of the ‘cliff’ and λ is related to the average slope below and above it.

the ask, will be below and above it, respectively. Also, if the quantity to be traded is large, then, possibly, 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. Its process $(\bar{S}_t(\omega))_{\forall t, \omega}$ is defined as a function

$$\bar{S}_{t_{i-1}}(\omega) \equiv f(S_{t_{i-1}}(\omega), H_{t_i}(\omega) - H_{t_{i-1}}(\omega), \gamma, \lambda) \quad (5)$$

of the observable spot $(S_t(\omega))_{\forall t, \omega}$, trade size $(H_{t_i}(\omega) - H_{t_{i-1}}(\omega))_{\forall t, \omega}$ and liquidity parameters $\gamma, \lambda \geq 0$, where, as will be demonstrated, the former is a proxy for the relative width of the bid-ask spread and the latter for the market depth. Intuitively, the reaction or price-impact function f in (5), in addition to being increasing with respect to the trade size, should have the asymptotic properties that

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

and, for small $\epsilon > 0$,

$$f(S_{t_{i-1}}, H_{t_i} - H_{t_{i-1}} = \pm\epsilon, \gamma, \lambda) > S_{t_{i-1}}(1 \pm O(\epsilon))$$

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 time series is employed in most financial applications such as for example performance ratios or technical trading rules. But, as Figure 3 shows, when including a price impact function the true average transaction price time series has a dependence on the traded quantity and is thus higher dimensional. This reflects the fact no agent can transact at the mid-market price. Instead, the average transaction price can be represented as the integrated and averaged order book:

$$\bar{S}_t \equiv \frac{\int_{H_t}^{H_t+\delta H_t} \psi(S_t, H - H_t) dH}{\delta H_t}, \quad (6)$$

where δH_t is defined as the change in holdings and ψ is the order book density.

One form of f that captures the properties above,⁸ is

$$\bar{S}_{t_{i-1}} \equiv S_{t_{i-1}} (1 + \text{sign}(H_{t_i} - H_{t_{i-1}})\gamma) e^{\lambda(H_{t_i} - H_{t_{i-1}})}, \quad (7)$$

where the explicit dependence on the trajectory ω has been suppressed and will henceforth only be shown when not obvious. In (7) the $\text{sign}(\cdot)$ function along with the parameter γ models the relative bid-ask spreads and the $\exp(\cdot)$ term with λ models the market depth, which represents the elasticity of the stock 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. Moreover, this form of the impact function implies that the order book density ψ , as an approximation of the observable true values, is given by

$$\psi(S, \delta H_t) = S e^{\lambda \delta H_t} (1 + \gamma \text{sign}(\delta H_t) + \delta H_t (\lambda (1 + \gamma \text{sign}(\delta H_t)) + 2\gamma \delta(\delta H_t)))$$

where $\delta(\cdot)$ is the Dirac-delta function.

In addition to the pure transaction cost effect, as already mentioned, in practice, there is a market impact effect that is felt by all participants. A large trade removes the best quotes from the order-book, which, in a centralised order-driven exchange, effectively represents the market.⁹ Thus some layers are no longer available to any of the other market participants and the latter will possibly adjust their new quotes accordingly. Hence, in effect, the market has been moved.

⁸Parts of equation (7) have already been proposed by [Jar1] and [Frey]. But they only included the exponential term

⁹Usually there is also a broker market on top of an electronic order-book, but most brokers will have a dual presence and, still, any trade even if pre-arranged has to go through the order book.

Depending on whether the transaction is a buy or a sell order, the average transaction price is below or above the last price traded, unless only one level of market depth has been tapped. 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 linear combination of the two previous equilibrium and average transaction log-prices, namely taking their geometric average.¹⁰ Adding this permanent effect to the temporary reaction (7) along with the random change (1) gives the full discrete-time price dynamics

$$S_{t_{i-1}} \tag{8}$$

$$\rightarrow \bar{S}_{t_{i-1}} \equiv S_{t_{i-1}} (1 + \gamma \text{sign}(H_{t_i} - H_{t_{i-1}})) e^{\lambda(H_{t_i} - H_{t_{i-1}})} \tag{9}$$

$$\rightarrow S_{t_{i-1}}^\alpha \bar{S}_{t_{i-1}}^{1-\alpha} = S_{t_{i-1}} (1 + \gamma \text{sign}(H_{t_i} - H_{t_{i-1}}))^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_i} - H_{t_{i-1}})} \tag{10}$$

$$\rightarrow \begin{cases} u S_{t_{i-1}} (1 + \gamma \text{sign}(H_{t_i} - H_{t_{i-1}}))^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_i} - H_{t_{i-1}})} \equiv S_{t_i}(\omega_u) \\ d S_{t_{i-1}} (1 + \gamma \text{sign}(H_{t_i} - H_{t_{i-1}}))^{(1-\alpha)} e^{\lambda(1-\alpha)(H_{t_i} - H_{t_{i-1}})} \equiv S_{t_i}(\omega_d). \end{cases} \tag{11}$$

Thus, facing a mid-market price (8), a market order of a quantity $H_{t_i} - H_{t_{i-1}}$ in the stock induces the trader to transact at the average transaction price (9) and transforms the original price into the next last price traded (10), which is subject to the subsequent random change (11).

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. $0 \leq \alpha \leq 1$. Their explanation of α is that large trades may not contain fundamental new information and may 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 based on intuition than 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, there appears a pure transaction costs model similar to those of [B&V], [BLPS] and [ENU]. This would imply that other market participants did not believe that the trade bore new fundamental information.

Given the interpretation of α as an observable slippage parameter that transforms the average transaction price into the next published price, there is a choice whether (10) should represent the exact next last-price-traded, or the new mid-market price derived therefrom. In the former case α would be non-positive because it is reasonable to assume that, in general, the best quotes

¹⁰In any model of this kind there is a tradeoff between parsimony and tractability as compared with accuracy of fit; as we see later, our choices appear to be a good compromise.

are filled first. Moreover it would directly produce the last observed price traded, which may be important in contingent claim contracts, especially of barrier type (see e.g. [Tal]). Speaking for the latter interpretation is the fact that because when using 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. But this distinction will only make a difference in the empirical estimation of the parameters and not in the model itself, so that this question is treated in Section 4.2.

The model set-up with respect to the slippage effect, albeit structured similarly, is different from those of [A&C1], [A&C2], [H&S1] and [H&S2], who postulate arithmetic Brownian motion

$$dS_t = \mu dt + \sigma dW_t, \quad (12)$$

as the process for the underlying. In their respective papers, this is an acceptable and computationally convenient model regarding portfolio trading applications. But it may cause serious concerns when it is applied to the pricing of derivatives. This is mainly 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 (7) will also be locally linear. Moreover, as derived in the next section, the exponential part of the reaction function makes the resulting tree Markovian, whereas it would be path-dependent for a linear model. Lastly, our model is also free of arbitrage opportunities, as long as $-\alpha$ is not too large, as we now show.

Proposition 2.1 (Non-existence of arbitrage opportunities)

If, for $\alpha \in [0, 1]$, the risky and riskless assets $(S_t(\omega), B_t)_{\forall t, \omega}$ follow the processes (8)-(11) 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.

$$\left(\hat{H}'_{t_i}(\omega) - \hat{H}'_{t_{i-1}}(\omega) \right) B_{t_{i-1}} + \left(H'_{t_i}(\omega) - H'_{t_{i-1}}(\omega) \right) \bar{S}_{t_{i-1}}(\omega) = 0, \quad \forall t, \omega \quad (13)$$

and results in a positive expected gain

$$E[V_T - V_{t_0} | \mathcal{F}_{t_0}] > 0, \quad (14)$$

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

Proof: For simplicity and without loss of much generality it is assumed that $r = 0$, $B_{t_0} = 1$ and $\hat{H}_{t_0} = H_{t_0} = H_T = 0$. Also it suffices to show it for $\alpha = 0$, as this is the most extreme form of

slippage. Thus, since $V_{t_0} = 0$ and $B_T = 1$, (14) reduces to

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

which after repeated substitution of (13) gives

$$\begin{aligned} E \left[\hat{H}_{t_{n-1}} - (H_T - H_{t_{n-1}}) \bar{S}_{t_{n-1}} \middle| \mathcal{F}_{t_0} \right] &= E \left[- \sum_{i=1}^n (H_{t_i} - H_{t_{i-1}}) \bar{S}_{t_{i-1}} \middle| \mathcal{F}_{t_0} \right] \\ &= E \left[- \sum_{i=1}^n \delta H_i S_{t_{i-1}} (1 + \gamma \text{sign}(\delta H_i)) e^{\lambda \delta H_i} \middle| \mathcal{F}_{t_0} \right], \end{aligned} \quad (15)$$

where the operator $\delta H_i = H_{t_i} - H_{t_{i-1}}$, $i = 1 \dots n$. Now, from (3), it is apparent that in the absence of trading S_{t_i} has to have the martingale property if the underlying market itself is free of arbitrage. Thus for each t_i it is possible to write the expectation as

$$E_Q[S_{t_i} | \mathcal{F}_{t_0}] = S_{t_0}. \quad (16)$$

Substituting (16) into (15) and separating the buy orders from the sell orders yields

$$-S_{t_0} \left(\sum_j \delta H_j (1 + \gamma) e^{\lambda \delta H_j} + \sum_l \delta H_l (1 - \gamma) e^{\lambda \delta H_l} \right) \leq 0, \quad (17)$$

where, for all j, l , we have $\delta H_j \geq 0$ and $\delta H_l < 0$. Inequality (17) follows from the fact that $H_{t_0} = H_T$, therefore $\sum_j \delta H_j = -\sum_l \delta H_l$, but $\exp(\lambda \delta H_j) \geq 1$ and $\exp(\lambda \delta H_l) < 1$. The increasing convexity of the exponential function means that a buy order will move an asset price up more, in relative terms, than a sell order will move it down. Thus transaction costs are either in the same direction or, if not, dominate the market manipulation effect and there are no arbitrage opportunities in trading the underlying. \square

Proposition 2.1 implicitly defines arbitrage in the sense of riskless cash profits, rather than riskless paper profits. Whereas it may be possible to, say, buy an amount of stock and immediately attain a riskless mark-to-market profit in the portfolio due to slippage, the subsequent liquidation of the amount will always result in a non-positive net proceed. In the same sense [Jar1] provides a proof of non-arbitrage for a more general class of reaction functions and also incorporates traded options into the market in [Jar2]. In fact the real no-arbitrage interval for α in Proposition 2.1 is wider, namely α may be negative without admitting riskless profits from trading strategies. The exact non-arbitrage bands are dependent on the number of timesteps, λ , γ and impractical to show explicitly for even small numbers of timesteps. But usually numerically it becomes immediately obvious if arbitrage is possible, as for instance the ask price of a long vanilla option would turn out to be negative or zero.

In the model developed in this paper, other traded options and the market impact on and due to them are ignored. The reason is, firstly, that in the basic model set-up it is not intended to employ them as a hedging instruments for other derivatives as it would immediately raise the complexity and economic intuition would be sacrificed. Secondly, traded options themselves have finite liquidity, usually much lower than that of the underlying, so that they usually cannot be used for dynamic arbitrage strategies. The primary application of the model as presented in this paper is the hedging of OTC derivatives positions as well as trading strategies in the underlying, but nonetheless in the summary an extension to the inclusion of traded options will be proposed.

2.1 The valuation and hedging of contingent claims in discrete time

A contingent claim $C_t(\omega)$ is a time- and, possibly, path-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 stopping 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 an optimal hedging strategy $V_{t_0}^*$ is at its minimum or maximum.

As a special class of contingent claims European vanilla options have a single expiry time T at which a cash or an asset exchange may occur. Upon expiry there are usually different ways 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

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

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 the terminal condition is $\hat{H}_T(\omega)B_T = C_T(\omega), \forall \omega$. It is 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 at time t_0 of a European contingent claim $C_t(\omega) > 0, \forall t, \omega$ is

$$C_{t_0} = \max(V_{t_0}^*, 0), \quad (18)$$

where

$$V_{t_0}^* \equiv \min_{(H_t(\omega), \hat{H}_t(\omega))_{\forall t, \omega}} V_{t_0} = (H_{t_1}^* - H_0) \bar{S}_{t_0} + \hat{H}_{t_1}^* B_{t_0} + H_0 S_{t_0}, \quad (19)$$

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

$$H_{t_0}^* = H_0, \quad (20)$$

the self-financing conditions

$$(\hat{H}_{t_i}(\omega) - \hat{H}_{t_{i-1}}(\omega)) B_{t_{i-1}} + (H_{t_i}(\omega) - H_{t_{i-1}}(\omega)) \bar{S}_{t_{i-1}}(\omega) = 0, \quad \forall t, \omega, \quad (21)$$

the payoff replication constraints

$$V_T(\omega) \equiv H_T(\omega) S_T(\omega) + \hat{H}_T(\omega) B_T = C_T(\omega), \quad \forall \omega, \quad (22)$$

and the processes $(S_t(\omega), B_t)_{\forall t, \omega}$ are given by (8)-(11) and (2), respectively.

The solution $V_{t_0}^*$ of (19) represents the minimum quantity 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 is consistent with the required terminal value $C_T(\omega), \forall \omega \in \Omega$. The right-most term of (19) represents the mark-to-market value of the initial quantity of the risky asset that is employed for hedging the claim. Moreover (18) adds a natural lower bound for the claim as the writer would not pay for a position that offers no positive returns. In fact, however, when the market is free of arbitrage opportunities, then the $\max(\cdot)$ function in (18) is superfluous. 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 that could be borrowed against the contract as collateral.

Corollary 2.1 (Replication bid price of a contingent claim)

The bid price of the contract of 2.2 is the solution of the program

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

where

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

subject to the initial holding (20) and self-financing condition (21), along with the replication constraint

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

with the asset dynamics as in Proposition (2.2).

Unlike for the ask price, the $\max(\cdot)$ function is required for the bid price, as we demonstrate below. It can be shown 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 high as the bid price, with equality when the market is perfectly liquid. Following a similar path for the proof as [BLPS], the following result is required:

Lemma 2.1 (Non-homogeneity of prices)

Denote by $C_{t_0}^{(i)}$ and $(H_t^{(i)}(\omega), \hat{H}_t^{(i)}(\omega))_{\forall t, \omega}$, $i = 1, \dots, 3$ the optimal values and holding strategies for the program given in Proposition 2.2 with the right-hand side of (22) replaced by $C_T^{(1)}(\omega) = \eta C_T(\omega)$, $C_T^{(2)}(\omega) = \zeta C_T(\omega)$ and $C_T^{(3)}(\omega) = (\eta + \zeta) C_T(\omega)$, where η, ζ are constants. Then,

1. if $\eta, \zeta \geq 0$ or $\eta, \zeta \leq 0$, it follows that $|C_{t_0}^{(1)} + C_{t_0}^{(2)}| \leq |C_{t_0}^{(3)}|$,
2. if $\eta \geq 0 \geq \zeta$ or $\zeta \geq 0 \geq \eta$, it follows that $|C_{t_0}^{(1)} + C_{t_0}^{(2)}| \geq |C_{t_0}^{(3)}|$.

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

1. At expiry

$$\left| H_T^{(3)}(\omega) \right| \geq \left| H_T^{(1)}(\omega) + H_T^{(2)}(\omega) \right|, \quad \forall \omega;$$

thus

$$\left| C_T^{(1)}(\omega) + C_T^{(2)}(\omega) \right| \leq \left| C_T^{(3)}(\omega) \right|, \quad \forall \omega,$$

and so, without loss of generality and for simplicity assuming that $\gamma = 0$, given transaction cost impact on the self-financing condition

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

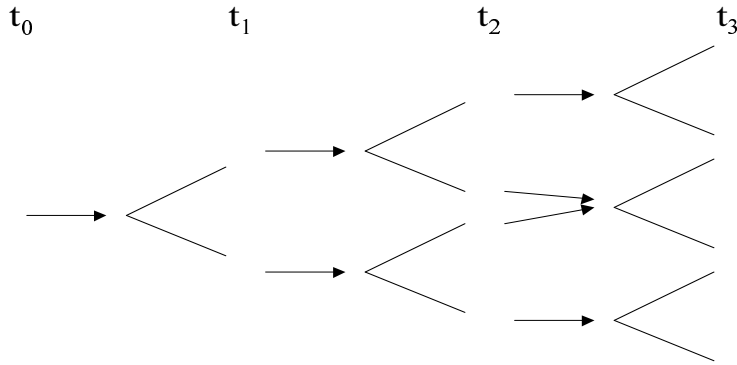


Figure 4: Asset tree, where lines represent random change and arrows slippage.

because the exponential part of the reaction function is super-linear, 1. follows.

2. The converse of 1. \square

Choosing $\eta = 1$ and $\zeta = -1$ gives $C_{t_0}^{(1)} = C_{t_0}$, $C_{t_0}^{(2)} = -\tilde{C}_{t_0}$ and $C_{t_0}^{(3)} = 0$, so that $C_{t_0} \geq \tilde{C}_{t_0}$. Intuitively this is obvious because the trader will have to buy high and sell low at every rehedg, thus paying bid-ask spreads. Hence the liquidity of the market for the underlying gives natural bid-ask spreads for a contingent claim on it, which represent non-arbitrage bands. Also, the larger the absolute payoff $|C_T|$, the wider the relative spreads. Thus the price of x options will be greater than x times the price of one option. So obviously 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 grows exponentially as the number of time steps increases. This also implies, that trees are contract-specific, as they depend on the respective Deltas. The following example demonstrates the growth of distinct points in state space.

Example 2.1 Consider a three period economy with the trading times $\{t_0, t_1, t_2, t_3\}$, the set of states $\Omega = \{\omega_{uuu}, \omega_{uud}, \dots, \omega_{ddd}\}$ and the information revelation $\mathcal{F}_{t_0} = \{\Omega\}$, $\mathcal{F}_{t_1} = \{\omega_u = \{\omega_{uuu}, \dots, \omega_{udd}\}, \omega_d\}$,

$\mathcal{F}_{t_2} = \{\omega_{uu}, \dots, \omega_{dd}\}$ and $\mathcal{F}_{t_3} = \{\{\omega_{uuu}\}, \dots, \{\omega_{ddd}\}\}$. Then the asset's dynamics are

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

It is 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 price tree ‘bushy’ and thus causes exponential growth of the number of variables and constraints. However, it is a useful property of our model that, for sufficiently simple¹¹ contingent claims $C_T(\omega)$ with single-signed Gamma,¹² 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

¹¹Claims for which there exists a unique solution holding strategy $(H'_t(\omega), \hat{H}'_t(\omega))_{\forall t, \omega}$. In general however, due to the highly nonlinear set-up of the model, it is usually difficult if not impossible to show analytically that a particular claim has a unique solution. But a subsequent Proposition will introduce a dynamical programming algorithm that drops this requirement.

¹²This requirement can be dropped when $\gamma = 0$.

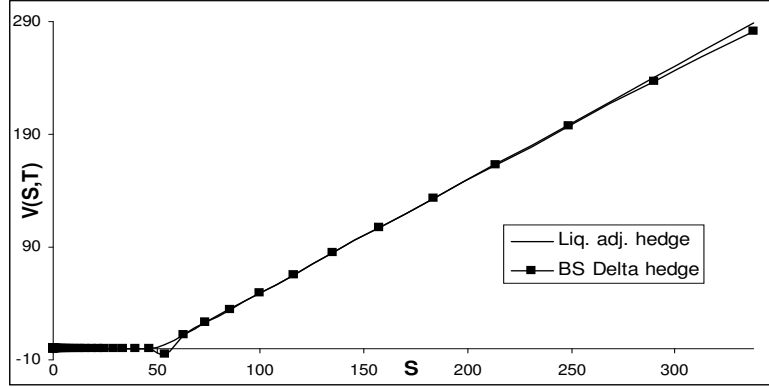


Figure 5: Payoff replication of a Call option when using the Black-Scholes Delta in comparison to the liquidity adjusted hedge. Here $T = 1$, $S_{t_0} = K = 50$, $\sigma = 0.2$, $r = 0.05$, $\gamma = 0.01$, $\alpha = 1$, $\lambda = 0.01$, $H_0 = 0$ and there are 50 timesteps.

be demonstrated backwards step by step. Firstly, at nodes of number of upsteps n_u and downsteps n_d ,

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

for positive Gamma,¹³ and with n_u and n_d swapped for negative Gamma. Secondly, if in Example 2.1 $(H'_{t_3}(\omega_{ud}), \hat{H}'_{t_3}(\omega_{ud}))$ is a solution to the replication constraint (22) 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 the uniqueness of the solution. This can be generalised for n timesteps. 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 presents 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 become wider; γ has a significantly larger effect in this process, especially as the number of timesteps increases. Also, as α decreases the absolute 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 its exogenous volatility σ . On the bid side, when both γ and $-\alpha$ are large, the lower bound for the option comes into effect. This means that when going long the option the cost of hedging

¹³Usually defined as $\Gamma \equiv \frac{\partial^2 V}{\partial S^2} > 0$, also called long Gamma, as for instance a long Call option. The main difference between long and short Gamma is, that when $S_{t_i} - S_{t_{i-1}} > 0$ it follows that $H_{t_{i+1}} - H_{t_i} < 0$ in the former case and vice versa in the latter.

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		1.077	1.008	1.001	1.000		1.000	0.999	0.992	0.920
	1		1.059	1.006	1.001	1.000		1.000	0.999	0.994	0.938
	1.1		1.047	1.005	1.000	1.000		1.000	1.000	0.995	0.951
gamma 0.001	0.9		1.138	1.071	1.064	1.063		0.936	0.935	0.928	0.852
	1		1.091	1.039	1.034	1.033		0.966	0.966	0.960	0.902
	1.1		1.064	1.022	1.018	1.017		0.982	0.982	0.978	0.932
gamma 0.01	0.9		1.631	1.573	1.568	1.567		0.245	0.244	0.232	0.112
	1		1.348	1.304	1.299	1.299		0.617	0.616	0.608	0.520
	1.1		1.206	1.169	1.166	1.165		0.823	0.822	0.817	0.761

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

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. It is also apparent that for $-\alpha$ sufficiently low the effect on the bid-ask spreads is asymmetric, namely the bid quote is relatively further away from the perfect liquidity value than the ask quote. All of these observations will become analytically obvious in the continuous-time limit model.

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. 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 scaling, i.e. that a hedging shortfall is dependent on the underlying value of the notional.

Certain assumptions of Proposition 2.2 and Corollary 2.1 can be relaxed, for both programs, making the optimisation problems either less restrictive or leading to the discovery of additional characteristics of the model:

Remark 2.1 (Independence of initial holding)

If the liquidity at the initiation of the hedging portfolio is perfect, i.e., $\lambda = \gamma = 0$, then the objective function (19) reduces to

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

The solution of the problem is insensitive to the initial holding condition (20) and if $H_0 = H_{t_1}^$, where the latter is the optimal Delta of the option, the solution is identical to the one of Proposition 2.2.*

Hence, 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 S and therefore also C , theoretically, could be manipulated arbitrarily. But, due to the non-existence of cash arbitrage, by selling or buying a large quantity of stock, the loss on transaction costs will outweigh any potential market manipulation benefits. In this respect the last term in (19) 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 Remark 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

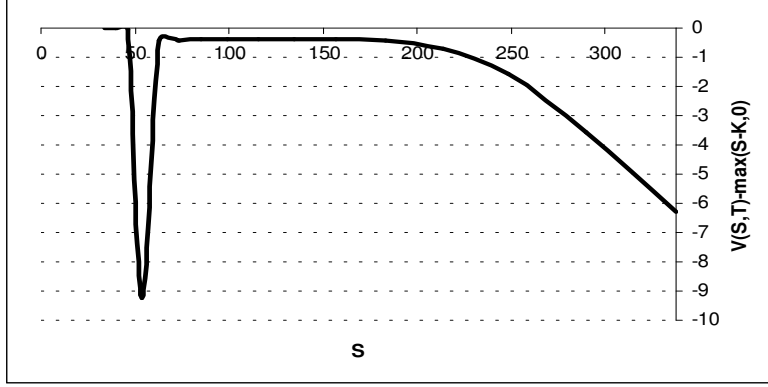


Figure 6: Hedging shortfall at expiry of Figure 5 under Black-Scholes Delta hedging.

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 indicates $H_{t_1}^*$.

Remark 2.2 (Conservation of funds)

The optimal solution C_{t_0} of the program in Proposition 2.2 stays unchanged if the self-financing condition (21) is relaxed to the inequality

$$(\hat{H}_{t_i}(\omega) - \hat{H}_{t_{i-1}}(\omega))B_{t_{i-1}} + (H_{t_i}(\omega) - H_{t_{i-1}}(\omega))\bar{S}_{t_{i-1}}(\omega) \leq 0. \quad (23)$$

Proof: Denoting the solution of the program under condition (23) by \bar{C}_{t_0} , because the solution space under (23) contains the one under (21), therefore $C_{t_0} \geq \bar{C}_{t_0}$. Also suppose that a particular strategy $(H_t''(\omega), \hat{H}_t''(\omega))_{\forall t, \omega}$ 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 $H_{t_i}^*(\omega) = H_{t_i}''(\omega) + d/B_{t_{i-1}}$, it will now satisfy the stronger condition (21), while still satisfying the original (22). Thus since d is positive but arbitrary $C_{t_0} \leq \bar{C}_{t_0}$. \square

Remark 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, a similar intuition does not necessarily hold when relaxing the terminal condition into an inequality, as in fact it may then be optimal to super-replicate a cash flow rather than meet it exactly.

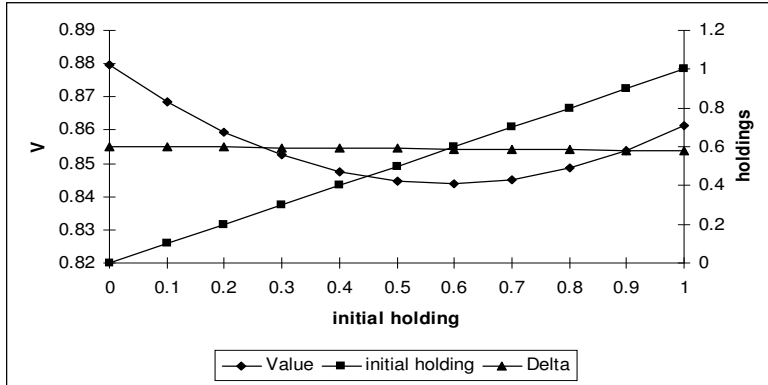


Figure 7: The left axis refers to the option value under different contract Delta and initial holding combinations, as given by the right axis. The minimum contract value is given when $H_0 = H_{t_1}^*$, i.e. where the initial holding and Delta lines cross. Here $T = 0.5$, $S_{t_0} = K = 10$, $\sigma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

Proposition 2.3 (Existence of optimal super-replication strategies)

If the replication condition (22) is turned into a super-replication constraint

$$V_T(\omega) = H_T(\omega)S_T(\omega) + \hat{H}_T(\omega)B_T \geq C_T(\omega), \quad \forall \omega, \quad (24)$$

then the optimal solution of Proposition 2.2 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 the 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 (22) 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 rehedgeing 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, 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 be 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, 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, which determine a region where the marginal costs of reheding 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 optimal solution and 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 (22) is turned into a super-replication constraint

$$H_T(\omega)S_T(\omega) + \hat{H}_T(\omega)B_T \geq C_T(\omega), \quad \forall \omega, \quad (25)$$

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

$$W_{t_0}(c) = \min_{(H_t(\omega), \hat{H}_t(\omega))_{\forall t, \omega}} E_P \left[\max \left(C_T - H_T S_T - \hat{H} B_T, 0 \right) \right] \quad (26)$$

subject to

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

as well as (20) and (22). The expectation is taken with respect to an arbitrary but strictly positive probability measure $P(\omega) > 0, \forall \omega$.

Proof: (See [ENU] for more details) 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. \square

The implementation and backward solution of the dynamical programming algorithm of Proposition 2.4 is described in Appendix B.

3 The continuous-time limit

As already mentioned, in the continuous-time limit the exogenous stock price process (1) converges to geometric Brownian motion (4). Moreover, a portfolio process is now defined as

$$V_t \equiv \int_{t_0}^t H_{t'} d\bar{S}_{t'} + \int_{t_0}^t \hat{H}_{t'} dB_{t'},$$

$d\bar{S}_t$ being the continuous-time equivalent of the combination of random and controlled processes (8)-(11) (where they exist). In order to analyse this continuous-time limit in more detail, we first consider the intermediate situation in which rehedges are carried out at discrete time intervals δt , between which the asset evolves by the standard geometric Brownian motion (4), so that its change is

$$\delta S = \mu S \delta t + \sigma S \delta W, \tag{27}$$

where δW is $N(0, \delta t)$. We can then compute the cash flows involved in a reheding before letting $\delta t \rightarrow 0$. (We could, alternatively, expand the binomial model of the previous section.)

As the frequency of rebalancing of a replicating portfolio for a contingent claim is increased, a hedger has to transact ever smaller amounts ever more frequently. In the limit infinitesimal amounts are traded continuously. In general, if transaction costs are proportional to the value traded, as in

the models of e.g. [B&V], [Lel] or [HWW], then any hedging strategy of this kind either leads to the infinite volatility solution, or becomes invalid, as $\delta t \rightarrow 0$. (If, in addition to the proportional, there are fixed cost components, then they push the initial cost $V_{t_0} \rightarrow \infty$ under continuous rebalancings.) 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 ($H_{t_0} = 1$), i.e. a static hedge, up-front. Nonetheless as [Lel] showed, if the rehedging interval δt is small, but non-infinitesimal, and the Delta is appropriately adjusted, a dynamic hedging strategy is still feasible and the probability of a large hedging error becomes small.

In the model proposed in the previous section, liquidity has an impact that is a combination of terms of varying orders of magnitude of the rehedging interval as $\delta t \rightarrow 0$. The leading order impact in fact depends on the value of the parameters α, γ, λ , which do not themselves directly scale with δt . As we shall see below, if it is assumed that $\gamma > 0$ and $\alpha < 1$, i.e. positive bid-ask spreads and a non-negligible slippage effect, then for very small rehedge intervals the cost of hedging a derivative indeed approaches the static hedge solution, because the hedging activity entirely dominates the asset price's diffusion, just as in a standard Leland-type model; furthermore, in this case bid-ask spreads dominate the price elasticity effect due to λ . We therefore begin with the case $\gamma = 0$. We also distinguish between cases with and without initial set-up costs.

Theorem 3.1 *Suppose that a contingent claim is hedged at fixed discrete time intervals δt , between which the asset follows (27), and that the impact of the hedging strategy is as in the previous section but with $\gamma = 0$. Suppose further that the replicating portfolio is initially perfectly hedged. Then as $\delta t \rightarrow 0$, the value of the derivative $V(S, t)$ satisfies the partial differential equation*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \lambda \sigma^2 S^3 \left(\frac{\partial^2 V}{\partial S^2} \right)^2 + \frac{1}{2}\lambda^2 (1 - \alpha)^2 \sigma^2 S^4 \left(\frac{\partial^2 V}{\partial S^2} \right)^3 + rS \frac{\partial V}{\partial S} - rV = 0, \quad (28)$$

with the constraint

$$V(S, t) \geq 0$$

and with appropriate contract-specific terminal and boundary conditions. Furthermore, the hedge ratio is $\Delta = \partial V / \partial S$, which satisfies

$$\begin{aligned} & \frac{\partial \Delta}{\partial t} + \frac{1}{2}\sigma^2 S^2 \left(1 + 4\lambda S \frac{\partial \Delta}{\partial S} + 3\lambda^2 (1 - \alpha)^2 S^2 \left(\frac{\partial \Delta}{\partial S} \right)^2 \right) \frac{\partial^2 \Delta}{\partial S^2} \\ & + (r + \sigma^2) S \frac{\partial \Delta}{\partial S} + 3\lambda \sigma^2 S^2 \left(\frac{\partial \Delta}{\partial S} \right)^2 + 2\lambda^2 (1 - \alpha)^2 \sigma^2 S^3 \left(\frac{\partial \Delta}{\partial S} \right)^3 = 0. \end{aligned} \quad (29)$$

Corollary 3.1 (Ask price of a contingent claim)

If the initial asset holding, at time t_0 of Theorem 3.1 is $H \neq \Delta(S_{t_0}, t_0)$, where Δ is as in Theorem 3.1, then the ask price of a contingent claim is

$$C(S_{t_0}, t_0) = V\left(S_{t_0} e^{\lambda(1-\alpha)(\Delta^* - H)}, t\right) + (\Delta^* - H) \left(e^{\lambda(\Delta^* - H)} - 1\right) S \quad (30)$$

provided that $C(S_{t_0}, t_0) > 0$ (otherwise it is zero) where Δ^* is such that

$$\Delta^* = \left. \frac{\partial V}{\partial S} \right|_{(S_{t_0} e^{\lambda(1-\alpha)(\Delta^* - H)}, t)} \quad (31)$$

where, for $t > t_0$, $V(S, t)$ is the solution of (28).

Proof. The derivation of the continuous-time limit is given in Appendix C. The ‘American’ constraint that $V(S, t) \geq 0$ ensures that the hedging strategy does not force the hedger into an irrational position. The equation for the Delta is obtained by differentiation of the equation for V . Lastly, the Corollary simply adjusts for set-up costs. \square

We now make some remarks on this model. Equations (37) and (28) are fully nonlinear PDEs and (29) quasilinear. Because the pricing equation is nonlinear, and specifically because of the gamma-squared term, long and short positions thus have different values. The pricing PDE thus structurally resembles the ones derived by [S&W], [F&S], [S&P] and the hedge PDE the one of [Frey]. The fact that the latter two collapse to their Black-Scholes equivalents

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

and

$$\mathcal{L}_{BS}\Delta \equiv \frac{\partial \Delta}{\partial t} + (r + \sigma^2)S \frac{\partial \Delta}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 \Delta}{\partial S^2} = 0 \quad (33)$$

when $\lambda = 0$, i.e. when the market is perfectly liquid, can be exploited to find approximations of V and Δ for small λ and analyse the nature of the nonlinearity. Most notably this will help to find economic interpretations of the squared and cubed Gamma terms in (28). Denoting by V_{BS} the Black-Scholes value, V is expanded in a regular perturbation series

$$V \sim V_{BS} + \lambda V_1 + \lambda^2 V_2 + \dots$$

and substituted into (28). Matching powers of λ yields equations for the first- and second-order terms of the value V :

$$\mathcal{L}_{BS}V_1 = -\sigma^2 S^3 \left(\frac{\partial^2 V_{BS}}{\partial S^2} \right)^2, \quad (34)$$

$$\mathcal{L}_{BS}V_2 = -2\sigma^2 S^3 \frac{\partial^2 V_{BS}}{\partial S^2} \frac{\partial^2 V_1}{\partial S^2} - \frac{1}{2}(1-\alpha)^2 \sigma^2 S^4 \left(\frac{\partial^2 V_{BS}}{\partial S^2} \right)^3. \quad (35)$$

The solutions of these equations depend on the boundary conditions and can be written in integral form, although in general direct numerical solutions of the PDEs are easier to compute.

It should be noted that for most contracts care has to be taken because the payoffs $V(S, T)$ 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, which appears in the source term of the first order correction equation (34), is not well-defined in the classical distributional sense, so the behaviour of (28) at expiry requires a more detailed analysis, which we do not present here (see [Mit]); in brief, a boundary layer analysis near the strike and near expiry allows us to exploit known results for nonlinear diffusion equations to construct local similarity-solution approximations which resolve the apparent discrepancy. Another way of dealing with payoff discontinuities, equivalent to the approach of [S&P], is to assume that (28) is valid only up to a small time before expiry: $T - \epsilon$, with perfect liquidity prevailing subsequently. Another possibility that [Frey] proposes is to approximate any kinked payoff by a smooth function, as for example through the regularisation

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

for a Call option. However, as [Mit] shows, neither of these is necessary.

To find reasonable interpretations of how the square and cube Gamma terms of (28) impact on V for a Call option equations (34) and (35) are examined more closely. The first correction term V_1 is driven by a Black-Scholes operator with a source term of Black-Scholes Gamma squared and, as V converges to its payoff at expiry, a terminal condition of $V_1(S, T) = 0$. An integral expression for the first-order term for the case when V is a vanilla Call option is given by [S&P], who show that it is always positive. Figure 8 below confirms this and also shows that it is most significant at-the-money for vanilla options, so that it can be interpreted as a transaction cost term.

The second order correction term V_2 has two separate source terms of which one vanishes when there is no slippage, i.e. when $\alpha = 1$. In the latter case it can be observed that it will cause an effect of opposite sign to the transaction cost term depending whether the hedger's position is long or short Gamma. In the former case the cost term is reduced and in the latter it is increased. This results in the total transaction costs being asymmetric, namely a long Gamma position being relatively more expensive than a short position, in the sense that a market maker would be prepared to pay relatively less for the option than the corresponding selling price. However, on the bottom left of Figure 8, when $\alpha < 1$ this effect is continuously reversed with decreasing magnitude of α due to the presence of the second source term. The latter can thus be interpreted as the feedback slippage effect.

Intuitively this can be understood by considering the difference between hedging short and long Gamma positions. For the former the slippage acts in the same direction as a diffusion, conversely for the latter. More specifically, as an example, when a hedger is short Gamma, having written, say, a Call option, and the stock price goes up, more of the stock has to be bought, pushing up the price further. Conversely a hedger who is long Gamma, say, long a Call has to sell more stock when the stock price goes up thus pushing down the price. This makes short Gamma positions relatively more expensive than long Gamma positions. The hedger would require a relatively higher premium for the former over compare with the latter.

Figure 8 demonstrates the liquidity effects on a Call option. The model produces volatility frowns, instead of the regularly observed skew and smile patterns. But as for instance [Con2] noted, in general, transaction costs and liquidity effects cannot explain the smile on their own. 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, making out-of-the-money Puts more expensive, the 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. Another point is that the equations are inherently nonlinear. The average premium per option is in fact superlinear with respect to the notional: as is demonstrated the premium for three times the notional is higher than three times the premium of one option notional, again corresponding to the discrete-time Lemma 2.1.

Finally, we briefly consider the case $\gamma > 0$, where the bid-ask spread is not negligible. When $\lambda = 0$, our model is the same as the extension of the Leland transaction costs model to non-single-signed Gamma described in [HWW]. That is, the value of a derivative satisfies the nonlinear equation

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = \frac{1}{2}\tilde{\sigma}^2 S^2 \left| \frac{\partial^2 V}{\partial S^2} \right|, \quad (36)$$

where

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

tends to infinity as $\delta t \rightarrow 0$. If the Gamma is always negative, then $|\partial^2 V/\partial S^2| = -\partial^2 V/\partial S^2$ and (36) is a standard Black–Scholes equation with effective variance $\sigma^2 + \tilde{\sigma}^2$, no matter how small δt . The continuous-time limit of the derivative price is then simply the 100%–Delta static hedge (the infinite volatility Black–Scholes solution). If, on the other hand, the Gamma is positive, then (28) becomes forward parabolic for $\tilde{\sigma}^2 > \sigma^2$ and hence the final value problem is ill-posed. The financial

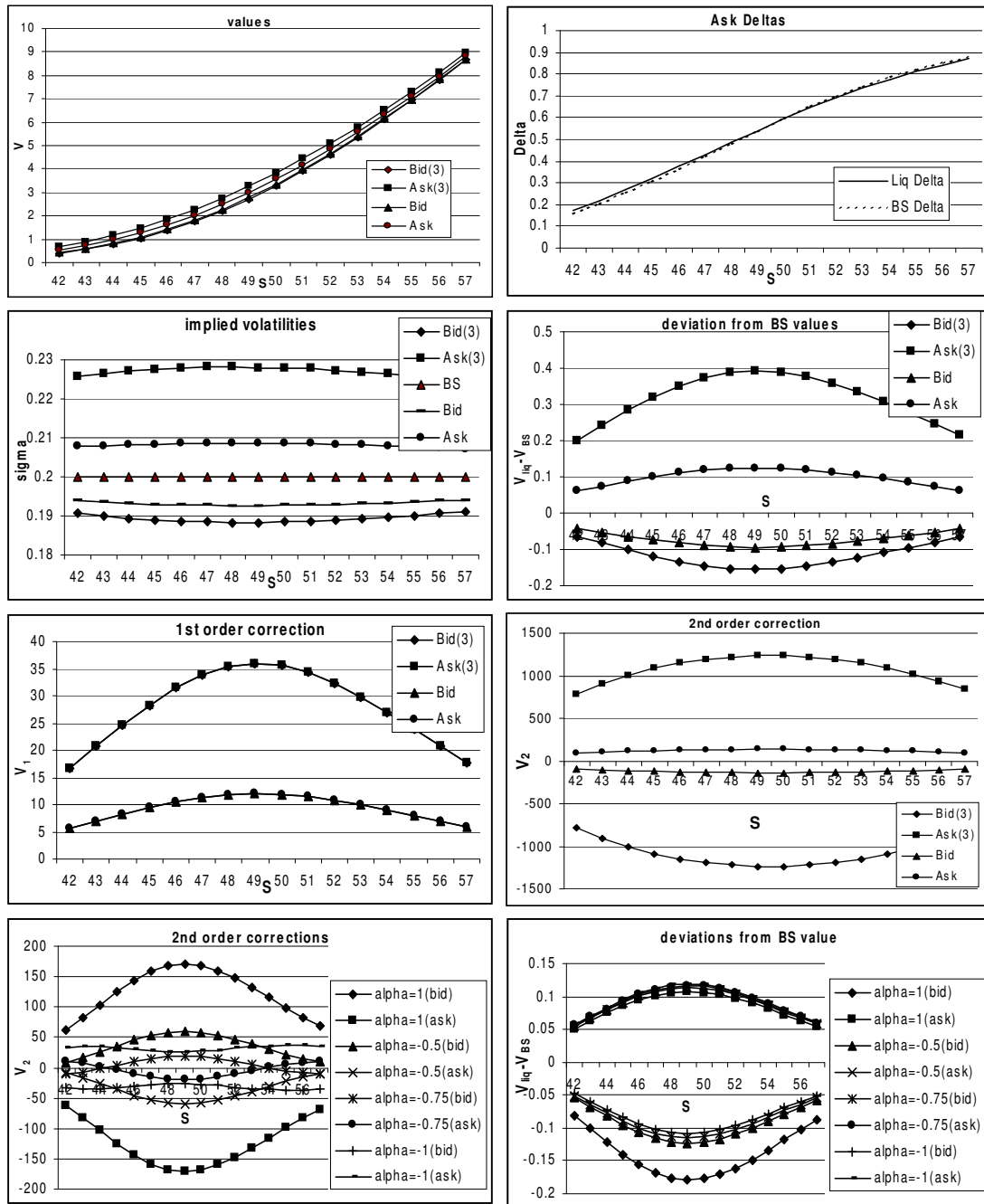


Figure 8: Long and short Call options with $T = 0.5$, $\sigma = 0.2$, $K = 50$, $r = 0.05$, $\lambda = 0.01$, $\alpha = -1.5$, $\gamma = 0$, $H = \Delta^*$, $\epsilon = 0.002$, the latter applied to the modification of the payoff time $T - \epsilon$ described in the text.

interpretation of this is that too frequent rehedging in this case is ruinously expensive and an alternative model must be sought. If the Gamma is not single-signed, the equation is therefore only valid for $\tilde{\sigma}^2 < \sigma^2$. In other words, for a given small level of transaction costs γ , there is a lower bound to the rehedging interval, below which the model is not valid. There is no true continuous-time limit, but nevertheless the approximate model (28) is a good approximation to the outcome of the hedging strategy provided that $\delta t > 2\gamma^2/(\pi\sigma^2)$.

With these remarks in mind, we state the equivalent of Theorem 3.1 when $\gamma > 0$.

Theorem 3.2 *Under the price dynamics and the self-financing conditions of Proposition 2.2, if $\sigma\sqrt{\delta t}/\gamma = O(1)$, the value of a generic replicating portfolio $V(S, t)$ is governed by the PDE*

$$\begin{aligned} \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2(1 + \bar{\gamma}^2(1 - \alpha)^2)S^2\frac{\partial^2 V}{\partial S^2} + \lambda\sigma^2S^3\left(\frac{\partial^2 V}{\partial S^2}\right)^2 + \frac{1}{2}\lambda^2(1 - \alpha)^2\sigma^2S^4\left(\frac{\partial^2 V}{\partial S^2}\right)^3 \\ + rS\frac{\partial V}{\partial S} + \sqrt{\frac{2}{\pi}}\bar{\gamma}\sigma^2S^2\left|\frac{\partial^2 V}{\partial S^2}\right| + \sqrt{\frac{2}{\pi}}\lambda(1 - \alpha)^2\bar{\gamma}\sigma^2S^3\left|\frac{\partial^2 V}{\partial S^2}\right|\frac{\partial^2 V}{\partial S^2} - rV = 0, \end{aligned} \quad (37)$$

with $\bar{\gamma}\sigma\sqrt{\delta t} = \gamma$, subject to contract-specific boundary conditions.

Proof: See Appendix C. \square

In order to obtain a feel for the orders of magnitude, as the subsequent section will demonstrate $\lambda \approx 10^{-8}$ per one unit of stock and $\gamma \approx 10^{-4}$. If, as is assumed, λ scales linearly with the number of shares δH , i.e. that they only appear as a product $\lambda\delta H$, then the corrections calculated in Figure 8 would be required when the underlying notional is equivalent to approximately 10^6 units of stock. Moreover, depending on the time scale used, δt in Theorem 3.2 is the order of magnitude of about 30-60 minutes.

4 Interpretation and calibration of the liquidity parameters

4.1 The interpretation and observation of liquidity

The definition of liquidity, both in the literature and in practice, has, so far, not converged to a similar level of standardisation to 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 price impacts calculated from cross-party and cross-border asset streams as a proxy.

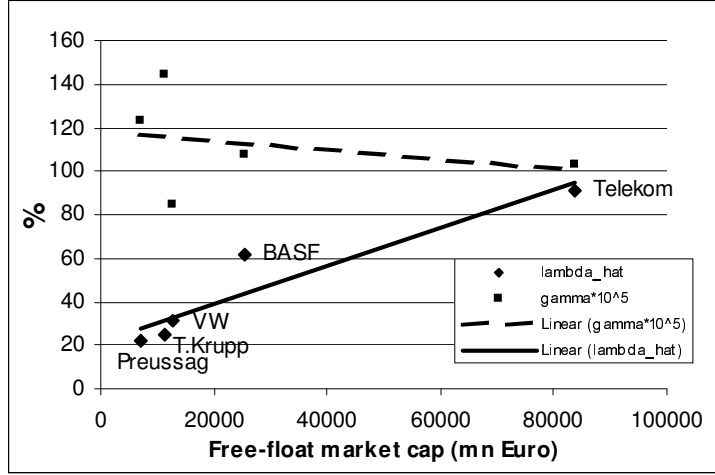


Figure 9: Average acquisition premia $\hat{\lambda}$, bid-ask spreads γ and their trend lines for various large stocks.

In a portfolio management and returns prediction context [CRS] provide measures of liquidity based on bid-ask spreads of assets along with an extensive empirical analysis. For the current paper most relevantly, from an option pricing perspective, a market-depth approach is taken by [Kraak] who explicitly defines liquidity as the reciprocal of $\delta N/\delta S$, i.e. the sensitivity of the stock price to the notional traded. In this form the parameter, having the dimensionality of the reciprocal of a number of stocks, effectively represents a relative change in price per share traded. This can further be normalised, as will be shown, by taking into account the total number of shares outstanding.

Our model uses a proxy measure that is a combination and an extension of the last two approaches. The parameter γ is a direct measure of (half) the relative bid-ask spreads between the best layers in an order book, whereas λ , dimensionally corresponding to the parameter of [Kraak] scales the slope of the average transaction price, i.e. measures the market depth. Finally α transforms the average transaction price into either the next last price traded, which will appear on the screens of all market participants or, if applicable, the new mid-market price derived therefrom. To further standardise this concept of liquidity ideally these parameters should be transformed into an intuitive and observable proxy measure that is both dimensionless and comparable across assets and/or markets. For this purpose $\lambda\delta H$ and $(1 + \gamma\text{sign}(\delta H))$ are both dimensionless variables, where λ represents approximately the marginal degree of price change per unit number of assets traded δH that the trader faces in a transaction beyond the initial relative bid-ask spread $1 \mp \gamma$, usually

termed ‘elasticity’ in economics. Hence

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

where the term involving $\gamma\lambda$ may be negligible. 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, say, equity as

$$\begin{aligned} &\text{\% acquisition (disposal) premium (discount)} \\ &\approx \pm\gamma + \hat{\lambda}(1 \pm \gamma) \times \text{\% market cap traded}. \end{aligned} \tag{38}$$

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

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

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 as will be demonstrated in the following section, and then in their respective combinations give a proxy for liquidity. A similar definition to (38) 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 in real time is only available to one single market maker (the specialist). Therefore the average transaction price change and slippage are not directly predictable for the other participants. More recently, [DFIS] simulated the price dynamics of a synthetic order book in a non-specialist market and arrived at a similar parameterisation for liquidity in terms of bid-ask spreads and elasticity.

Figure 9 shows the ranking of the time-weighted average values of γ and $\hat{\lambda}$, as defined by (38), for various German blue-chip stocks traded on the Xetra market against their free-float market capitalisation. Bid-ask spreads appear 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.

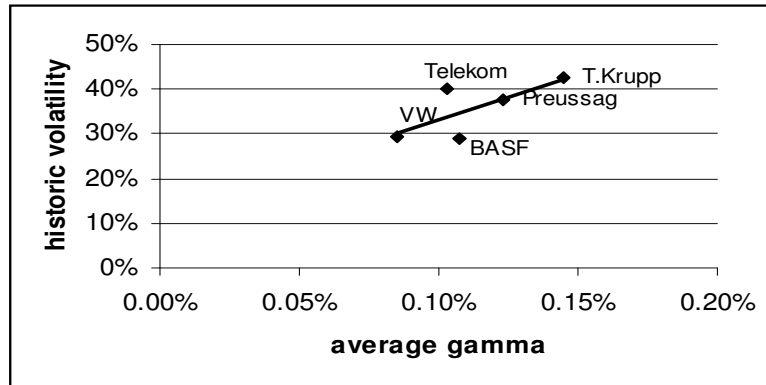


Figure 10: Annualised historical spot volatilities realised during the second half of 1999 plotted against γ .

This means, when extrapolating, if at one point in time a market order for a 1% free-float stake of a large stock like Deutsche Telekom is placed, the acquisition premium will amount to about 90%. On the other hand a smaller cap like Preussag would only command a 20% premium. However, in practice it is rare, albeit not impossible, that such a large amount would be traded at once, causing swings of this magnitude. (One particular such incident occurred on the London Stock Exchange with its blue-chip FTSE 100 index on Friday, 20 September 2002, which was a so-called ‘triple-witching day’, when options on both single equity and the index, as well as the index future expire and their settlement prices are determined by the average price within a 20 minute interval. On this particular day the swings in the index amounted to almost 8%, with some components swinging almost 100% within seconds.)

Another observable relationship may be that the opportunity cost of holding inventory in an order book is higher for more volatile stocks. In fact this is confirmed when plotting historic realised volatilities against γ as in Figure 10. This positive relationship is only observed for the width of the bid-ask spreads, but not for λ . This could be explained by the intuitive fact that most trades happen at the best bid and ask, whereas deeper layers are tapped less frequently.

4.2 Estimation of the liquidity parameters

The data set analysed consisted of about four and a half hours of five layers of market depth on both the bid and ask side for five large cap stocks traded on the Xetra system of the German stock exchange. This is a fully electronic trading platform that matches quotes automatically, and

virtually all of the daily volume of the blue chip stocks goes through it. Even though the data set appears relatively small, still each stock had between 300 and 1000 ticks per hour, i.e. instances where at least one of the layers was updated in price or in quantity offered. For the estimation of the parameters it was 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 is not necessarily invalidated by the fact that another broker market exists on top of the order book, as it is reasonable to assume that the liquidity supplied by brokers is related to the directly observable part of the order book.

To estimate the parameters an ordinary least squares minimisation was employed because the model set-up fits well into a linear regression framework, which further provides the R^2 measure of goodness-of-fit. Also, the fitted parameters were allowed to be time-dependent, i.e. of the form λ_t and γ_t , and subsequently their stability was surveyed. Depending on the a priori assumptions of which parameters should be included, they were added step-by-step, increasing the complexity of the calibration, and it was observed how the fit of the model developed. Initially $\gamma_t = 0$ and market slippage was ignored; then λ_t and S_t could be directly estimated by discretising the quantities H_t through the simple linear regression with log-transform of the response:

$$\ln(\bar{S}_t) = \ln(S_t) + \lambda\delta H_t + \epsilon_t. \quad (39)$$

Here \bar{S}_t represents the exact average transaction price under a specific traded quantity δH_t and S_t the estimated mid-market price. The choice of intervals of δH_t proved to have relatively little influence on the results, so the same equi-distant quantities for each stock were employed, namely 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. However, 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

$$\ln(\bar{S}_t) = \ln(S_t) + \ln(1 + \text{sign}(\delta H_t)\gamma_t) + \lambda_t\delta H_t + \epsilon_t. \quad (40)$$

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

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

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 we see in the section on stochastic liquidity below, the values

of λ_t and γ_t across time were far from constant. One problem that estimation method (40) 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 in this case is

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

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

$$\gamma = \frac{S_t^+ - S_t^-}{S_t^+ + S_t^-}.$$

This model could then again be improved in terms of accuracy by allowing for two different slopes, i.e. a separate λ for bid and ask depth:

$$\ln(\bar{S}_t) - I_{[x>0]}(\delta H_t) \ln(S_t^+) - I_{[x<0]}(\delta H_t) \ln(S_t^-) = \lambda_t^+ \delta H_t I_{[x>0]}(\delta H_t) + \lambda_t^- \delta H_t I_{[x<0]}(\delta H_t) + \epsilon_t. \quad (42)$$

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

$$\ln(\hat{S}_t) = \ln(S_t) + (1 - \alpha_t)(\ln(1 + \text{sign}(\delta H_t)\gamma_t) + \lambda_t \delta H_t) + \epsilon_t, \quad (43)$$

where λ and γ are taken from the initial fitting of the average transaction price. 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

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

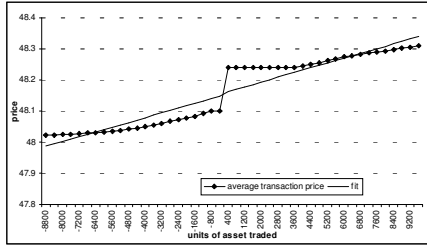
so that the left-hand side of (43) is replaced by

$$\frac{1}{2} \left(\ln(\hat{S}_t) + \ln(S_t^- I_{[x>0]}(\delta H_t) + S_t^+ I_{[x<0]}(\delta H_t)) \right); \quad (44)$$

an arithmetic average is equally possible to estimate the next last observed mid-price:

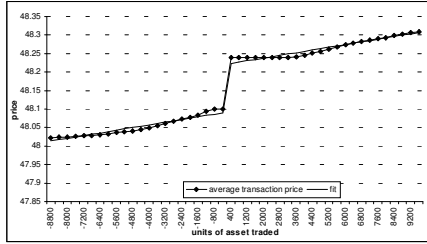
$$\frac{1}{2} \hat{S}_t + \frac{1}{2} (S_t^- I_{[x>0]}(\delta H_t) + S_t^+ I_{[x<0]}(\delta H_t))$$

so that the left-hand side of (4.6) can be replaced by its logarithm.



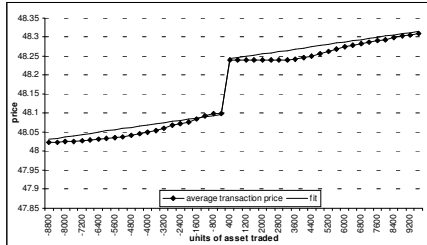
BASF
Telekom
Thy.Krupp
VW
Preussag

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%



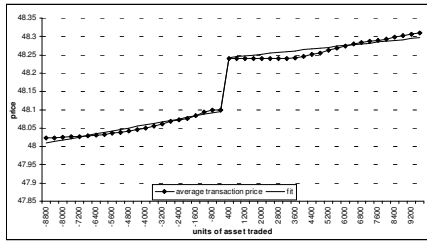
BASF
Telekom
Thy.Krupp
VW
Preussag

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%



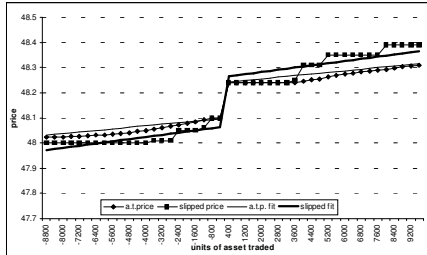
BASF
Telekom
Thy.Krupp
VW
Preussag

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%



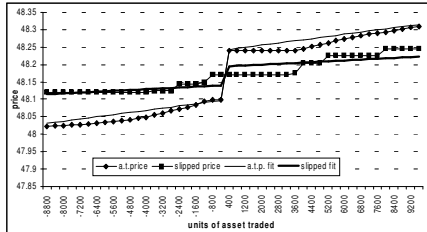
BASF
Telekom
Thy.Krupp
VW
Preussag

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%



BASF
Telekom
Thy.Krupp
VW
Preussag

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%



BASF
Telekom
Thy.Krupp
VW
Preussag

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 (39), (40), (41), (42), (43), (44). Graphs represent the various fits for a particular snapshot of the order book of BASF at 12:14:41pm on 13 Jan 2000.

5 Extensions to the basic model

5.1 Multiple underlying assets

The basic model with, for simplicity and without loss of much generality, $\gamma = 0$ easily generalises if the economy has m risky assets $(S_t^{(l)}(\omega))_{\forall t, \omega}$, $l = 1 \dots m$. In continuous time they follow the respective correlated 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 as

$$\begin{aligned} & S_{t_{i-1}}^{(l)} \\ \rightarrow & \bar{S}_{t_{i-1}}^{(l)} \equiv S_{t_{i-1}}^{(l)} e^{\lambda_l (H_{t_i}^{(l)} - H_{t_{i-1}}^{(l)})} \\ \rightarrow & \begin{cases} u_l \left(S_{t_{i-1}}^{(l)} \right)^{\alpha_l} \left(\bar{S}_{t_{i-1}}^{(l)} \right)^{1-\alpha_l} = u_l S_{t_{i-1}}^{(l)} e^{\lambda_l (1-\alpha_l) (H_{t_i}^{(l)} - H_{t_{i-1}}^{(l)})} \\ d_l \left(S_{t_{i-1}}^{(l)} \right)^{\alpha_l} \left(\bar{S}_{t_{i-1}}^{(l)} \right)^{1-\alpha_l} = d_l S_{t_{i-1}}^{(l)} e^{\lambda_l (1-\alpha_l) (H_{t_i}^{(l)} - H_{t_{i-1}}^{(l)})}. \end{cases} \end{aligned} \quad (45)$$

Firstly, the economy still does not allow arbitrage.

Proposition 5.1 (Non-existence of arbitrage opportunities)

If, for $\alpha_l \in [0, 1]$, $\forall l$, the risky and riskless assets $(S_t^{(l)}(\omega), B_t)_{\forall t, \omega, l}$ follow the processes (45) and (2) respectively, there does not exist a particular holding strategy $(H_t^{(l)' }(\omega), \hat{H}_t^{(l)}(\omega))_{\forall t, \omega, l}$, with $H_{t_0}^{(l)' } = H_T^{(l)}$, $\forall \omega, l$, which is self-financing and value-conserving i.e.

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

which results in a positive expected gain

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

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

Proof: Analogous to Proposition 2.1. \square

Again the true non-arbitrage bands for all α_l are in fact wider, although not explicit. Secondly, analogous to the single underlying case the nonlinear optimisation program for a generic contingent claim $C_t(S_t^{(1)}, \dots, S_t^{(m)})$ is given by the following:

Proposition 5.2 (Replication ask price of a contingent claim)

The ask price of a contingent claim $C_t(\omega) > 0$ at time t_0 is

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

with

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

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

$$H_{t_0}^{(l)*} = H_0^{(l)},$$

the self-financing conditions

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

the payoff replication constraint

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

and the processes $(B_t(\omega), S_t^{(l)}(\omega), \bar{S}_t^{(l)}(\omega))_{\forall t, \omega, l}$ are given by (2) and (45).

Again, as in the univariate case, the $\max(\cdot)$ function in the objective functions is not necessary when the market is free of arbitrage. If 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 time steps n , the order of magnitude of the number of constraints and variables is $O(n^{m+1})$ and it becomes apparent that the discrete-time method will be impractical. Again a PDE can be derived in the continuous-time limit, which is amenable to 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 with $\bar{\gamma}_l = 0, \forall l$ in the continuous-time limit the value of a self-financing replicating portfolio $V(S^{(1)}, \dots, S^{(m)}, t)$ is governed by the PDE

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

subject to contract-specific boundary conditions.

Proof: The multivariate case is a straightforward extension of Theorem 3.2, noting that $E[dW_t^{(i)}dW_t^{(j)}] = \rho_{ij}dt$, $\rho_{ii} = 1$ and $E[dW_t^{(i)}dW_{t-dt}^{(j)}] = 0$ are the respective correlation coefficients between the Wiener processes. \square

5.2 Distinct bid and ask liquidity

Typically, as shown in Figure 1, 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 buying at a time when everybody wants to sell and vice versa, the liquidity for the transactions will be good. The converse will probably hold if one follows all other market participants. Apart from γ , which is calculated with respect to the mid-market price, the reaction function (7) offers only one scaling parameter for the slope of the average transaction price on the bid and ask side. It is a linear, symmetric approximation for small λ times quantity traded. As the empirical results in Table 2 show, 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 (8) can be extended to

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

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

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 (49) has to be

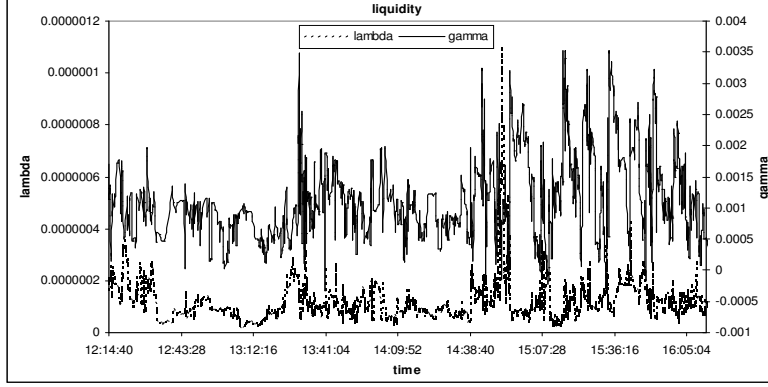


Figure 11: λ and γ tick time series for BASF over a specific time period.

replaced by

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

5.3 Stochastic liquidity

As is the case with employing Black-Scholes volatility, in practice, treating liquidity as a constant parameter is a convenient modelling assumption which, sometimes, admits tractable solutions. But in reality empirical analysis shows that both λ and γ are 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 standard for stochastic volatility modelling, could be appropriate. An easy model for, say, $\lambda_t(\omega)$ would be a mean-reverting Ornstein-Uhlenbeck process:

$$d\lambda_t = a(b - \lambda_t)dt + c dW_t^\lambda, \quad (46)$$

where a , b and c are constants. It is however well-known that following this process $\lambda_t(\omega)$ could become negative with positive probability. Another alternative that would keep $\lambda_t(\omega)$ positive would be a process similar to the one used in the [CIR] interest rate model or in various stochastic volatility formulations:

$$d\lambda_t = a(b - \lambda_t)dt + c\sqrt{\lambda_t} dW_t^\lambda. \quad (47)$$

To estimate the parameters a, b, c , representing the reversion speed, mean and volatility scaling parameters, respectively, the data was assumed to be a discrete realisation of the stochastic process

and again ordinary least squares were employed to obtain the maximum likelihood estimators. For that purpose minutely time-weighted arithmetic averages of the data were taken and for (46) the simple linear regression model was estimated through an Euler scheme (see e.g. [Jae]):

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

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

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

For (47) it is well known that the direct application of the Euler scheme leads to biased parameter estimates. But by invoking Itô's formula the response variable can be transformed as

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

whose parameters can again be estimated through the multiple linear regression model

$$\begin{aligned}\sqrt{\lambda_{t+1}} &= \left(1 - \frac{1}{2}a\delta t \right) \sqrt{\lambda_t} + \left(\frac{1}{2}ab - \frac{1}{8}c^2 \right) \delta t \lambda_t^{-\frac{1}{2}} + c\sqrt{\delta t}\epsilon_{t+1} \\ &= \beta_1 \sqrt{\lambda_t} + \beta_2 \lambda_t^{-\frac{1}{2}} + \tilde{\epsilon}_{t+1},\end{aligned}$$

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

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

Table 3 gives the results of the least-squares fitting for both $\lambda_t(\omega)$ and $\gamma_t(\omega)$ as estimated through (40). The second model seems to fit marginally better for $\lambda_t(\omega)$ and vice versa for $\gamma_t(\omega)$.

In general, when employing a diffusive process of type

$$d\theta_t = \mu_\theta(\theta_t, t)dt + \sigma_\theta(\theta_t, t)dW_t^\theta$$

for a generic stochastic parameter $\theta_t(\omega)$ and functions $\mu_\theta, \sigma_\theta$, a one-factor Black-Scholes type PDE would be extended by partial derivatives with respect to the new factor, so that it becomes

$$\mathcal{L}_{BSV} = -(\mu_\theta - m\sigma_\theta) \frac{\partial V}{\partial \theta} - \frac{1}{2}\sigma_\theta^2 \frac{\partial^2 V}{\partial \theta^2} - \rho\sigma\sigma_\theta S \frac{\partial^2 V}{\partial S \partial \theta} \quad (48)$$

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 (48) is extended by further partial derivatives and their respective market prices of risk and correlation coefficients.

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 $\lambda_t(\omega)$ and $\gamma_t(\omega)$, estimated on minutely data.

	BASF			Telekom			ThyKrupp			VW			Preussag		
	dS	$d\lambda$	$d\gamma$	dS	$d\lambda$	$d\gamma$	dS	$d\lambda$	$d\gamma$	dS	$d\lambda$	$d\gamma$	dS	$d\lambda$	$d\gamma$
$dS(\text{BASF})$	1.00														
$d\lambda(\text{BASF})$	-0.21	1.00													
$d\gamma(\text{BASF})$	-0.19	-0.30	1.00												
$dS(\text{Telekom})$	0.40	0.01	-0.18	1.00											
$d\lambda(\text{Telekom})$	-0.03	-0.02	0.07	0.04	1.00										
$d\gamma(\text{Telekom})$	0.12	-0.02	0.05	0.02	-0.16	1.00									
$dS(\text{ThyKrupp})$	0.15	-0.04	-0.11	0.18	0.07	-0.03	1.00								
$d\lambda(\text{ThyKrupp})$	-0.08	-0.08	-0.11	0.00	-0.05	-0.17	0.11	1.00							
$d\gamma(\text{ThyKrupp})$	0.01	0.06	0.01	-0.02	-0.04	0.13	-0.21	-0.33	1.00						
$dS(\text{VW})$	0.23	-0.16	-0.06	0.31	0.00	-0.04	0.04	0.06	0.01	1.00					
$d\lambda(\text{VW})$	-0.15	0.07	0.00	-0.07	-0.03	0.08	-0.02	-0.13	0.05	-0.22	1.00				
$d\gamma(\text{VW})$	0.10	0.06	0.05	0.03	0.01	-0.02	0.07	-0.03	0.14	-0.14	-0.35	1.00			
$dS(\text{Preussag})$	0.17	0.04	-0.03	0.15	-0.01	0.03	0.09	-0.02	0.13	0.06	-0.09	0.01	1.00		
$d\lambda(\text{Preussag})$	-0.01	0.02	-0.01	-0.05	-0.07	0.06	0.07	-0.09	-0.03	-0.06	0.12	-0.02	0.28	1.00	
$d\gamma(\text{Preussag})$	-0.05	-0.05	0.02	0.05	0.05	0.05	-0.06	0.17	0.14	-0.03	0.01	0.04	0.07	-0.45	1.00

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

Table 4 gives the estimated correlation matrix for the analysed universe of underlyings and their respective processes of $\lambda_t(\omega)$ and $\gamma_t(\omega)$. There seems to be a significant negative correlation between a stock's $\gamma_t(\omega)$ and $\lambda_t(\omega)$. This may be explained by noting that orders are usually filled around the best quotes where, as indicated in Figure 2, 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. Also there is, as expected, a positive correlation between spot across assets. The other correlations do not reveal a consistent picture, suggesting that their cross-terms in the eventual PDE would be less significant. Additionally, by observing the magnitude of the mean-reversion 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. Also $\gamma_t(\omega)$ seems to be moving faster than $\lambda_t(\omega)$, as suggested in Figure 11. This confirms that most trades happen at the best bid and ask. In general, for this type of process [FPS] show that this will allow asymptotic approximations when solving a PDE of type (48).

However, the derivation of equation (48) assumes that the liquidity parameter can somehow be hedged and thus a risk-neutral world exists. 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 for, say, stochastic volatility modelling, is that traded options themselves have finite liquidity and, in practice, this 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 in either cash or options. 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 chosen stochastic model is tractable in some form, as for instance the two processes proposed above.

6 Conclusion

We have presented a parametric market model along with an observable proxy measure of liquidity for particular traded assets. On the basis of this model derivatives can be both hedged and valued in a risk-neutral framework, thus under no-arbitrage, both in discrete time and in the continuous-time limit. In addition to the standard Black-Scholes model the employed liquidity proxy measure

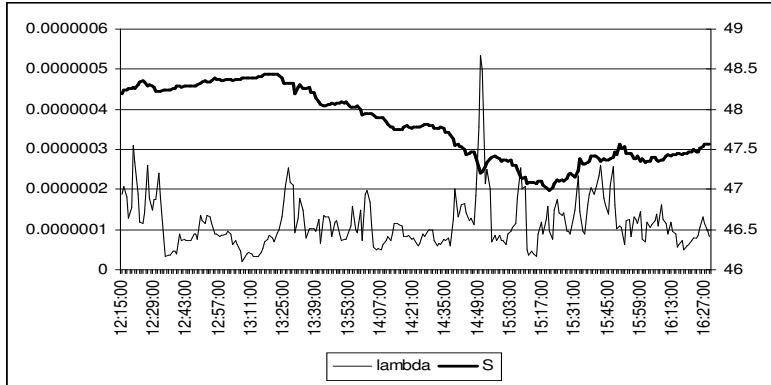


Figure 12: Minutely averaged λ (left axis) and spot mid-price of BASF over the reference time period. The separation of time-scales is clearly observable.

consists of a trade impact function with three parameters: γ measuring half the relative width of the bid-ask spread in reference to a mid-market price, λ as a proxy for the slope of the average transaction price as a function of quantity traded, namely the price elasticity of the asset, and α giving the eventual market price slippage due to the trade. Moreover, all of these parameters are directly observable in a non-specialists' market order book of layered best bid and ask quotes.

As a direct application we incorporated this liquidity impact function in an option pricing framework by appending a controlled process to the CRR binomial model. Then the value and the hedging strategy of the derivative were given by the solution and the optimal controls of a constrained non-linear optimisation program. Depending on the type of option and whether the position was long or short the model, due to its nonlinearity, generated unique and arbitrage free bid-ask spreads. Furthermore, by allowing for super-replication strategies, for many types of contracts, as is the case for most transaction cost models, there exist parameter ranges where the solution may be superior to exact replication. For this case we formulated a dynamical program that solved for the cheapest super-replication strategy.

The continuous-time limit of the model resulted in a number of PDEs, each valid for a certain order of magnitude of the parameters and the chosen rehedging period. If the bid-ask spread of the underlying asset was assumed to be a significant factor ($\gamma > 0$) and the rehedging interval small, then the bid-ask spreads of the underlying dominated its price elasticity effect and hedging of a contingent claim would need to be done in discrete time and a continuous-time model would be unsuitable. But if there existed no price slippage, i.e. $\alpha = 1$, then it reduced to the transaction

costs models of [HWW] and [Lel]. Most significantly, when the hedging period chosen was large enough both the bid-ask spreads and the elasticity added nonlinear terms to the PDE. When the former 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 then resulting PDE contained square and cube terms of Gamma, where the former represented the transaction costs adjustment and the latter the asymmetry of slippage under positive and negative Gamma. The only care that needed to be taken was how to deal with non-smooth option payoffs.

To find empirical evidence German equity market data was analysed for the purpose of 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 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 it was concluded that market cap and the elasticity of liquidity under this proxy seem to be negatively related, indicating that small caps seem to be relatively more liquid. 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 liquidity. The liquidity parameters appeared to be fast mean-reverting and were well explained by some standard types of these processes.

As a part of future research there exist a number of further extensions to and applications of the model other than those already mentioned. Firstly, at least numerically, it should be straightforward to extend the model by a non-constant, possibly stochastic, volatility framework. This would imply the introduction of the traded options market into the model. Whereas [Jar2] and [Frey] have done it in their particular models, they did not introduce finite liquidity into the traded options market as well. As it was mentioned before, the latter would in practice be much less liquid than the cash market. Nonetheless, options are in practice often used as static hedges, so that continuous rebalancing would not be required and there may exist a better strategy than the dynamic Delta-hedge. For this purpose, in discrete time, additional constraints would have to be introduced and the objective function modified to find the cheapest possible hedging strategy of a particular position.

Another application, which creates a link to market microstructure theory or Econophysics as by e.g. [Kyle] and [F&J], would be to employ the model for strike, barrier, stop-loss or, in general, position detection after observing sequences of large trades. This would represent the inverse problem of hedging option positions. Essentially it would entail attempting to decompose observed volatility into a predictable part and noise, as is for instance done by [L&W] or [P&D], by using techniques

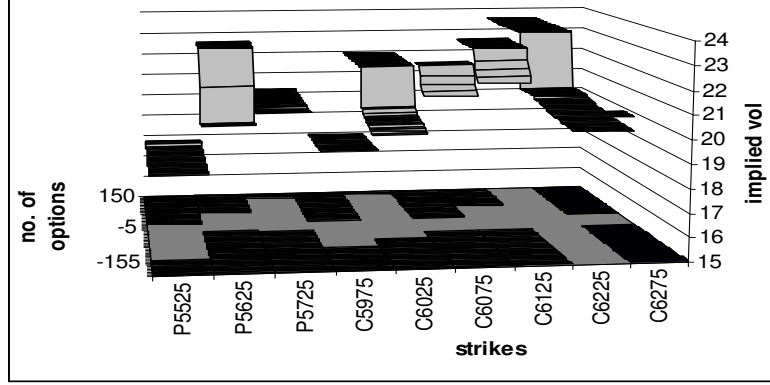


Figure 13: Implied volatility surface of FTSE100 index options, derived from a reconstructed order book snapshot of 1 March 2001 at around 8:15am. The underlying spot level moved around 5910. Contracts with strikes below spot were Calls, with strikes above spot Puts.

like maximum likelihood analysis or optimal filtering to estimate the Delta and Gamma of the large market participant and predict the resulting impact of their future actions. This may then introduce further feedback effects of traders trying to arbitrage the predictable agent.

Finally, in general, liquidity may be represented as an additional dimension in many, if not all, asset markets, by returning to first principles, namely recalling that price of and demand for an asset are positively related. As Figure 3 shows 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, an interest rate yield curve or an implied volatility smile/skew as Figure 13 indicates. Then, for instance, by observing the latter the question has to be asked how precisely, say, a volatility surface has to be estimated, when the non-arbitrage bounds around it are substantial. We hope that this paper is an initial step towards an approach of finding answers to these questions.

A Solving the nonlinear systems of equations

For a generic terminal condition $C(S, T)$ it is necessary to solve the system of simultaneous implicit nonlinear functions

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

$j = 1, \dots, n$, 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 Gamma 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:

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

$i = 1, \dots, n, j = 1, \dots, i - 1$. Solving for the holding process $(H_t(\omega), \hat{H}_t(\omega))_{\forall t, \omega}$ an algorithm is required that will converge to the roots of the respective simultaneous equations. One standard possibility is the Newton method¹⁴

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

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

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

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

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

¹⁴See e.g. [Str] for a detailed description.

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 asset 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

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

and the self-financing constraints

$$\begin{aligned} (H_{t_i} - H_{t_{i-1}})e^{\lambda(H_{t_i} - H_{t_{i-1}})} (1 + \text{sign}(H_{t_i} - H_{t_{i-1}})\gamma) \tilde{u}^{i-j-1} \tilde{d}^j \tilde{S}_{t_{i-1}} \\ + (\hat{H}_{t_i} - \hat{H}_{t_{i-1}})B_{t_{i-1}} \leq 0, \quad j = 0, \dots, i-1; i = 1, \dots, n. \end{aligned} \quad (49)$$

C Proofs of Theorems 3.1 and 3.2

For either of the PDEs in these theorems to hold, it must be invariant under the choice of starting point within a time interval (the set-up costs are dealt with in the corollary to Theorem 3.1). We assume an already hedged portfolio that will need to be rehedgeed after observing the asset price diffusion; the proof of an initially unhedged position is analogous, except that the sequence of events is reversed. As before, for notational convenience, we drop the time subscript.

Assuming that H is already the correct hedge quantity, we start with a position

$$\Pi = V(S, t) - HS = \hat{H}B, \quad (50)$$

where by the self-financing requirement the right-hand cash position is exactly equal to the value of the left-hand portfolio. The change in V is first due to the exogenous diffusion of the asset, which is assumed to follow geometric Brownian motion, and subsequently due to one's own trading, i.e. a change in H . The first step is to observe the evolution of the left-hand side of (50), i.e. the mark-to-market value of the contract and the stock over a small but not infinitesimal time interval

δt , namely

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

due to the exogenous diffusion of the asset. Next, the portfolio needs to be rehedge, thus inducing the portfolio dynamics

$$\overline{\Pi + \delta\Pi} = V(\overline{S + \delta S}, t + \delta t) - (H + \delta H)(S + \delta S)(1 + \gamma \text{sign}(\delta H))^{(1-\alpha)} e^{\lambda(1-\alpha)\delta H}. \quad (51)$$

Likewise, the change in the cash portion is

$$\overline{\Pi + \delta\Pi} = (1 + r\delta t)(V - HS) - \delta H(S + \delta S)(1 + \gamma \text{sign}(\delta H))e^{\lambda\delta H}, \quad (52)$$

where the last term is the cash flow from rehedge, which does not involve α because the trading occurs before the market impact. We analyse these under the assumptions that δS and δH are small. We begin with the case $\gamma = 0$, in which there is a true continuous-time limit in which we have the Itô result $(\delta S)^2 \rightarrow \sigma^2 S^2 \delta t$ as $\delta t \rightarrow 0$.

C.1 No bid-ask spread (Theorem 3.1)

We expand (51) and (52) by Taylor series, with the usual series for the exponential, namely

$$e^{\lambda(1-\alpha)\delta H} = 1 + \lambda(1-\alpha)\delta H + \frac{1}{2}\lambda^2(1-\alpha)^2(\delta H)^2 + \dots$$

We anticipate that δH is of the same order of magnitude as δS , that is $O(\sqrt{\delta t})$, and so we retain terms up to and including quadratic in these quantities. The result of expanding the right-hand side of (51) is, after rearrangement,

$$\begin{aligned} V - HS + \frac{\partial V}{\partial t}\delta t + \left(\frac{\partial V}{\partial S} - H\right)\delta S + \left(\lambda(1-\alpha)\left(\frac{\partial V}{\partial S} - H\right) - 1\right)S\delta H \\ + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(\delta S)^2 + \left(\lambda(1-\alpha)\left(\frac{\partial V}{\partial S} - H + S\frac{\partial^2 V}{\partial S^2}\right) - 1\right)\delta S\delta H \\ + \left(\frac{1}{2}\lambda^2(1-\alpha)^2 S\left(\frac{\partial V}{\partial S} - H + S\frac{\partial^2 V}{\partial S^2}\right) - \lambda(1-\alpha)S\right)(\delta H)^2, \end{aligned}$$

in which V , H etc are evaluated at (S, t) , while (52) gives

$$V - HS + r(V - HS)\delta t - S\delta H - \delta S\delta H - \lambda S(\delta H)^2.$$

We first note that the choice

$$H = \frac{\partial V}{\partial S}$$

cancels the δS risk exactly. Moreover, even though δH is also random, this choice also removes the term linear in δH . Next, our choice of H means that

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

so we can simplify the quadratic terms using the rule $(\delta S)^2 = \sigma^2 S^2 \delta t$. After rearrangement, the result of Theorem 3.1 follows.

C.2 The case $\gamma > 0$: Theorem 3.2

Bearing in mind the remarks before Theorem 3.2, we first write $\gamma = \bar{\gamma} \sigma \sqrt{\delta t}$. As is standard in transaction cost analyses, we no longer have a perfect hedge, but instead we make the same choice

$$H = \frac{\partial V}{\partial S}$$

so that again

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

and then equate the *expected* values of the cash flows on both sides of (51) and (52). Firstly, though, we construct an expansion similar to that above, noting that

$$\left(1 + \text{sign} \left(\frac{\partial^2 V}{\partial S^2} \delta S \right) \bar{\gamma} \sigma \sqrt{\delta t} \right)^{(1-\alpha)} = 1 + (1-\alpha) \text{sign} \left(\frac{\partial^2 V}{\partial S^2} \delta S \right) \bar{\gamma} \sigma \sqrt{\delta t} + \dots,$$

and that

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

On taking expectations, in which we use the results that

$$E[|\delta S|] = \sigma S \sqrt{\frac{2\delta t}{\pi}}, \quad E[\delta S] = O(\delta t),$$

the result of Theorem 3.2 follows.

References

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

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

- [BLPS] Bensaid B., Lesne J.P., Pagès H. & Scheinkman J., 1992, *Derivative Asset Pricing with Transaction Costs*, *Mathematical Finance*, 2, 63-86
- [B&L] Bertsimas D. & Lo A., 1998, *Optimal Control of Execution Costs*, *Journal of Financial Markets*, 1, 1-50
- [BHK] Breen W., Hodrick L. & Korajczyk R., 2000, *Predicting Equity Liquidity*, Kellogg Graduate School of Management, Working Paper
- [B&S] Black F. & Scholes M., 1973, *The Pricing of Options and Corporate Liabilities*, *Journal of Political Economy*, 81, 637-654
- [B&Sa] Borodin A. & Salminen P., 1996, *Handbook of Brownian Motion*, Birkhäuser
- [B&V] Boyle P. & Vorst T., 1992, *Option Replication in Discrete Time with Transaction Costs*, *The Journal of Finance*, 47, 271-293
- [CRS] Chordia T., Roll R. & Subrahmanyam A., 2000, *Order Imbalance, Liquidity and Market Returns*, *Journal of Financial Economics*, forthcoming
- [Con] Constantinides G., 1986, *Capital Market Equilibrium with Transactions Costs*, *Journal of Political Economy*, 94, 842-862
- [Con2] Constantinides G., 1997, *Transaction Costs and the Pricing of Financial Assets*, *Multinational Finance Journal*, 2, 93-99
- [CIR] Cox J., Ingersoll J. & Ross S., 1985, *A Theory of the Term Structure of Interest Rates*, *Econometrica*, 53, 385-407
- [CRR] Cox J., Ross S. & Rubinstein M., 1979, *Option Pricing: A Simplified Approach*, *Journal of Financial Economics*, 7, 229-263
- [DFIS] Daniels M., Farmer D., Iori G. & Smith E., 2002, *Demand Storage, Market Liquidity and Price Volatility*, Santa Fe Institute, Working Paper
- [ENU] Edirisinghe C., Naik V. & Uppal R., 1993, *Optimal Replication of Options with Transactions Costs and Trading Restrictions*, *Journal of Financial and Quantitative Analysis*, 28, 117-138
- [Eva] Evans L., 1998, *Partial Differential Equations*, American Mathematical Society
- [F&J] Farmer D. & Joshi S., 2000, *The price dynamics of common trading strategies*, Santa Fe Institute, Working Paper

- [Fig] Figlewski S., 1989, *Options Arbitrage in Imperfect Markets*, Journal of Finance, 44, 1289-1311
- [Frey] Frey R., 1996, *The Pricing and Hedging of Options in Finitely Elastic Markets*, University of Bonn, Discussion Paper
- [F&S] Frey R. & Stremme A., 1995, *Market Volatility and Feedback Effects from Dynamic Hedging*, University of Bonn, Discussion Paper
- [FPS] Fouque J.P., Papanicolaou G. & Sircar R., 1999, *Financial Modeling in a Fast Mean-Reverting Stochastic Volatility Environment*, Asia-Pacific Financial Markets, 6, 37-48
- [G&L] Gennotte G. & Leland H., 1990, *Market Liquidity, Hedging and Crashes*, The American Economic Review, 80, 999-1021
- [HWW] Hoggard T., Whalley E. & Wilmott P., 1994, *Hedging Option Portfolios in the Presence of Transaction Costs*, Adv. Fut. Opt. Res., 7, 21-35
- [H&S1] Huberman G. & Stanzl W., 2000, *Arbitrage-free Price-update and Price-Impact Functions*, Yale, Working Paper
- [H&S2] Huberman G. & Stanzl W., 2000, *Optimal Liquidity Trading*, Yale, Working Paper
- [Jae] Jaeckel P., 2002, *Monte-Carlo Methods in Finance*, Wiley
- [Jar1] Jarrow R., 1992, *Market Manipulation, Bubbles, Corners, and Short Squeezes*, Journal of Financial and Quantitative Analysis, 27, 311-336
- [Jar2] Jarrow R., 1994, *Derivative Security Markets, Market Manipulation and Option Pricing Theory*, Journal of Financial and Quantitative Analysis, 29, 241-261
- [Kra] Krakovsky A., 1999, *Pricing Liquidity into Derivatives*, Risk, Dec.
- [Kyle] Kyle A., 1985, *Continuous Auctions and Insider Trading*, Econometrica, 53, 1315-1336
- [Lel] Leland H., 1985, *Option Pricing and Replication with Transactions Costs*, The Journal of Finance, 40, 1283-1301
- [L&W] Lo A. & Wang J., 1995, *Implementing Option Pricing Models When Asset Returns Are Predictable*, The Journal of Finance, 50, 87-129
- [Lon] Longstaff F., 1995, *Option Pricing and the Martingale Restriction*, Review of Financial Studies, 8, 1091-1124

- [Mit] Mitton M., 2003, *DPhil Transfer Thesis*, OCIAM, Oxford University
- [M&R] Musiela M. & Rutkowski M., 1997, *Martingale Methods in Financial Modelling*, Springer
- [Per] Persaud A., 2000, *The Liquidity Puzzle*, Risk, June
- [P&D] Putyatin V. & Dewynne J., 1999, *Market Liquidity and its Effects on Option Valuation and Hedging*, Philosophical Transactions of the Royal Society, 357, 2093-2108
- [Rut] Rutkowski M., 1998, *Optimality of Replication in the CRR Model with Transaction Costs*, Applicationes Mathematicae, 25, 29-53
- [S&W] Schönbucher P. & Wilmott P., 2000, *The Feedback Effect of Hedging in Illiquid Markets*, SIAM Journal of Applied Mathematics, 61, 232-272
- [S&P] Sircar R. & Papanicolaou G., 1998, *General Black-Scholes Models Accounting for Increased Market Volatility from Hedging Strategies*, Applied Mathematical Finance, 5, 45-82
- [SSC] Soner M., Shreve S. & Cvitanic J., 1995, *There is No Nontrivial Hedging Portfolio for Option Pricing with Transaction Costs*, Annals of Applied Probability, 5, 327-355
- [Str] Strang G., 1986, *Introduction to Applied Mathematics*, Wellesley Cambridge Press
- [Tal] Taleb N., 1997, *Dynamic Hedging*, Wiley
- [W&W] Whalley E. & Wilmott P., 1997, *An Asymptotic Analysis of an Optimal Hedging Model for Option Pricing with Transaction Costs*, Mathematical Finance, 7, 307-324
- [WHD] Wilmott P., Howison S. & Dewynne J., 1993, *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press