School of Social Sciences
CITY UNIVERSITY LONDON

Department of Economics

Optimal Trading Strategies in a Limit Order Market with Imperfect Liquidity

Polina Kovaleva*
City University London

Giulia Iori
City University London

Department of Economics
Discussion Paper Series
No. 12/05

* Corresponding Author: *Department of Economics, School of Social Science, City University London, Northampton Square; polina.kovaleva.1@city.ac.uk*

Optimal Trading Strategies in a Limit Order Market with Imperfect Liquidity

Polina Kovaleva* and Giulia Iori†

20 September, 2012

Abstract

We study the optimal execution strategy of selling a security. In a continuous time diffusion framework, a risk-averse trader faces the choice of selling the security promptly or placing a limit order and hence delaying the transaction in order to sell at a more favorable price. We introduce a random delay parameter, which defers limit order execution and characterizes market liquidity. The distribution of expected time-to-fill of limit orders conforms to the empirically observed exponential distribution of trading times, and its variance decreases with liquidity. We obtain a closed-form solution and demonstrate that the presence of the lag factor linearizes the impact of other market parameters on the optimal limit price. Finally, two more stylized facts are rationalized in our model: the equilibrium bid-ask spread decreases with liquidity, but increases with agents risk aversion.

JEL classification: D4, D81, D84, G1, G12
Keywords: order submission, execution delay, first passage time, risk aversion, liquidity traders

1 Introduction

The problem of optimal order placement is the kernel of the successful implementation of an investment strategy since the optimal trade execution reduces the associated transaction costs and augments expected returns. Traders construct their submission strategies to benefit from particular market properties and order types hence the architecture of the market defines their expectations about the future price dynamics and trade execution efficiency. In pure quote-driven dealer markets small orders typically execute at the best opposing dealer quote regardless of the order type. In public limit order books, market orders may encounter price improvement, whereas limit orders execution is conditional upon where traders place their limit prices relative to the prevailing bid and ask. Market orders trade at the best price currently available at the market and are filled instantaneously. The actual execution price of a market order is subject to the current market situation –

*Corresponding author: email polina.kovaleva.1@city.ac.uk, tel + 44 (0)20 7040 4645, fax + 44 (0)20 7040 8580, address: City University, Department of Economics, Northampton Square, London EC1V 0HB
†City University, Department of Economics, Northampton Square, London EC1V 0HB

this aspect of trading is referred to as execution price uncertainty. Limit orders are instructions to trade at the best price available but only if it is not worse than the limit price specified by a trader. The probability that a particular order will be filled depends on its limit price. There exist two main risks associated with limit order trading: execution uncertainty, when market moves away from the submitted limit price hence the agent never trades, and ex post regret, when for various reasons prices move towards and through the limit price. The goals pursued by traders operating in a competitive environment of non-intermediated double auction markets are distinct from those of market makers. The fundamental distinction between the models of dealer quoting and a generic trader problem is that the former is essentially indifferent to execution, therefore his objective function is a zero profit. In limit order markets agents have multiple reasons to trade and various financial instruments available to execute a particular trade. Primarily, though, each trader entering a non-intermediated market must decide upon the type of order to use – a market order or a limit order.

Substantial empirical evidence suggests limit orders play a dominant role in the markets. For instance, Harris and Hasbrouck (1996) found that limit orders generally perform best even in the presence of a non-execution penalty and market price improvement. According to Guéant et al. (2011), more than half of all trades, approximately 60%, are as passive as possible, that is they fill the queue to trade rather than consume the liquidity. Biais et al. (1995) present an empirical analysis of the order flow of the Paris Bourse which is a pure limit order market. They find that traders' strategies vary with market conditions, with more limit orders submitted at times when spreads are wide and market orders prevailing when spreads are narrow. In real markets the motivation to choose a limit order over a market order is obfuscated by various subtle effects, for example: the discrepancy in transaction fees,[1] the option to submit multiple orders simultaneously, the existence of several distinct markets for the same asset and the possibility to withdraw the order. Many previous studies were based on the division of the traders into two main groups: liquidity suppliers who trade with limit orders, and liquidity demanders who have higher immediacy priority. In contrast, we develop a framework where depending on the present market conditions the trader sets out either to provide or to consume liquidity.

In practice, even though investment and trading decisions are formulated jointly, they are usually analyzed and accomplished somewhat in isolation. Our model reflects precisely such allocation of tasks since we do not take into account a portfolio the trader holds and how the outcome of his trading operations affects the balance and the value of this portfolio. Instead, we assume that the agent enters the market with a given trading goal and his task is to devise an optimal execution strategy given market characteristics at the time of his arrival. The information set available to traders is an important issue in this context. There is abundant empirical evidence in the market microstructure literature suggesting that market movements are often triggered by information updates and the presence of information asymmetries. A number of early studies showed that the information component of the bid-ask spread is fractional; according to Huang and Stoll (1997), on average it compounds less than 12% of the spread. Yet, more contemporary findings indicate that the information component of the spread is significant and amounts for as much as 80% (Gould et al., 2010). Information asymmetry is usually defined in a fairly broad sense and it does not necessarily imply, for instance, any form of legal or illegal insider trading. Simply because the tools with which market participants assess the market vary, they draw different predictions from the same market conditions. However, superior or rather

[1] Maker/taker market microstructure, for instance, encourages liquidity provision by offering a fee discount when a limit order is submitted, while consuming liquidity entails higher charges.

heterogeneous information is not the only basis for trading as was demonstrated by Milgrom and Stockey (1982). Guided by this line of thought, we do not analyze information effects per se; we incorporate traders' individual expectations of the future asset price in our model, while leaving the grounds for this valuation beyond the scope of the present study.

In the present study we consider a problem of a trader who has to liquidate a position in an actively traded asset within a given period of time and forms his strategy based upon the market dynamics represented by the bid and ask prices. Building upon the model in Iori et al. (2003), we conceptually improve it by incorporating an exogenous limit order execution factor – an exponential random delay. Moreover, we use a quadratic utility function and examine how different risk perception affects the optimal strategy of a trader in a limit order market and his average waiting time. This formulation proves more advantageous both in terms of interpretation and the ease of potential calibration to data. Further we provide an explicit static solution to the limit order trading problem for the quadratic utility preferences and subsequently identify the key determinants of the limit order attractiveness to the trader.

The remainder of this paper proceeds as follows. Section 2 reviews relevant theoretical and empirical literature. In Section 3 we describe the market in which traders submit their orders, outline the clearing mechanism and formulate the problem of a risk-averse trader who operates in this market. In Section 4 we look at the distribution of limit orders trading times implied by the market design. In subsequent Sections 5 and 6 we examine the properties of the optimal strategy and two special cases respectively. Section 7 reports comparative statics analysis for the parameters of the model. The existence of equilibrium spread and the appropriate conditions are discussed in Section 8. The paper concludes in Section 9 with a brief summary and a discussion of issues for further research.

2 Literature Review

There is an extensive literature on the subject of the optimal order submission strategy in limit order markets. The main distinction between the theoretical approaches adopted in various studies lies in the definition of the limit order execution mechanism and, consequently, the resulting probability distribution.

Equilibrium analysis of order-driven markets has been realized by Kumar and Seppi (1994); Chakravarty and Holden (1995); Parlour (1998); Foucault (1999); Foucault et al. (2005); Goettler et al. (2005) and Rosu (2009). All of these models are variants of a dynamic multi-agent sequential bargaining game where heterogeneous traders derive their best response order submission strategies. Parlour (1998) assumes that the probability of execution of a sell limit order depends on the arrival of buy market orders and the relative attractiveness of buy market orders depends on relative attractiveness of buy limit orders, thus execution probabilities of buy and sell limit orders are determined jointly over time. In situations when prices are fixed, which holds in equilibrium, optimal order placement is contingent upon a single factor – the distribution of agents' impatience characteristic. Allowing for price movements, Foucault (1999) describes the asset price via binomial model and assumes that limit orders are valid only for one period, therefore, at each point in time the book is either full or empty. In this setting the probability of a limit order execution is endogenous and a part of execution risk arises from the next trader's order type. The focus on the optimal behavior in equilibrium yields numerous implications coherent with the documented market observations and these models proved especially useful for policy-makers. However, order-driven markets do not seem ever to attain these

conditions: while in equilibrium all market participants should get zero profits, the depth of real limit order book is usually insufficient to drive average expected profits to zero.

A separate branch of optimal order placement literature, where the individual traders' order submissions are aggregated by the asset price dynamics, was initiated by Cohen et al. (1981). In their seminal paper, Cohen et al. (1981) propel the theoretical analysis of the optimal choice between market and limit orders in a framework with the probability of order execution contingent upon future price movements and the associated probability densities. Trading takes place when the trajectory of the best quote first crosses the barrier determined by the limit price. Cohen et al. (1981) model the security price with a compound Poisson process, which, by the very definition, invokes a jump in the probability of execution: if a time-constrained trader is willing to buy a stock via a limit order and sets a price infinitely close to the current best ask, the probability of trading never attains unity. This property permitted to establish a so-called "gravitational pull effect": when the bid-ask spread is narrow, the benefit of a price improvement with a limit order becomes small compared to the risk of non-execution so traders are pulled to use market orders instead. Consequently, Cohen et al. (1981) argue that a limit order strategy is not always superior to trading with market orders and to refrain from submitting any orders might even be the best. This model adequately captures the trade-off between a favorable price and a higher order execution probability and, in this sense, draws a line between market and limit orders.[2] Cohen et al. (1981) further demonstrate that as the order arrival rate increases the Poisson process converges to the Wiener process eliminating the discontinuity in the execution probability function. The model setup, however, is too complex to obtain a closed-form solution and their analysis remains qualitative for the main part.

Langnau and Punchev (2011) concentrate deliberately on the issue of adequate price modeling in a public limit order book. Adopting the results of Kou and Wang (2003), they compare the implications of a pure diffusion and a double exponential jump diffusion (DEJD) mid-point price specifications. The appealing properties of a double exponential jump distribution include the memorylessness of the price and the ability to accommodate the leptokurtic nature of returns. They arrive at a closed-form solution for the first passage time with distinct expressions for buy and sell strategies in the latter case. The compelling result of this paper is that the DEJD case accommodates the asymmetric shape of a limit order book as well as a fat-tailed distribution of log returns, it is compatible with the equilibrium conditions whereas log-normal prices are not. Langnau and Punchev (2011) do not address directly the optimal trading problem. Although DEJD prices hinder the parameter calibration, it is potentially beneficial from the tractability viewpoint and we reserve it for further research.

Given the analytical complexity of the setting with the jumps in prices, it seems reasonable to examine the problem of optimal strategy within the context of a continuous time diffusion price. Several contributions can be found in the financial and econometric literature. For instance, Iori et al. (2003) show that despite its obvious shortcomings, the log-normal price still suggests that the optimal limit order strategy is coherent with the traders behavior observed from the market. A mean-reverting price specification, particularly relevant for commodities, depicts an interesting cross-over effect: the optimal value of the strategy increases with the speed of reversion for small expiry times, while decreases for longer expiry times. Nevertheless, pure first passage time models such as this do not justify the existence of the spread due to inability to differentiate between a marketable limit order and a market order, or more precisely, to distinguish among the time-to-

[2]The orders placed inside the current spread are also interpreted as limit orders.

first-fill, time-to-completion, and time-to-censoring when a limit order is withdrawn.

An attempt to preserve the appealing mathematical lightness of first passage time models and bridge it with the notion of imperfect liquidity was made by Harris (1998). Harris (1998) improves the framework with a pure diffusion prices by adding a supplemental criterion – an aggregated factor of degree of execution difficulty. The degree of difficulty in limit order execution is defined as an additional barrier which a limit price has to pass before a limit order is filled. Therefore it is not sufficient to become the best price on the same side of the market, a limit order has to supersede this best price: for instance, a sell limit order is executed when submitted limit order price is lower than the best ask less the difficulty parameter. Two execution mechanisms were juxtaposed: with certainty and with some probability. Through the comparative statics analysis based on a numerical solution Harris (1998) confirms that the probability of order execution depends positively on this difficulty factor, in the certainty case as well. An alternative stylized interpretation of the problem was given by Hasbrouck (2006). Assuming that traders on the opposite side of the book are ascribed with unobserved reservation prices, a random collateral barrier following exponential distribution is analyzed. Hasbrouck (2006) argues that the number of potential counterparties is decreasing in the intensity rate of reservation prices distribution, thus optimal strategy becomes more aggressive. Although very valid, theoretical approaches presented by Harris (1998) and Hasbrouck (2006) have certain limitations. It appears to be a challenging task to infer the reservation prices or the degree of difficulty in limit order distribution from data due to the apparent difficulty in disentangling the impact of these external factors from the order aggressiveness determinants and the actual trading times. Instead we introduce a tractable parameter to characterize limit orders which can actually be inferred from observable data.[3]

3 The Model

We consider the problem of an investor who has a position in a traded asset which has to be liquidated within a pre-specified time horizon. One option the agent has is to use a market order and trade at the best available price at once. Alternatively, he can submit a limit order and hope to trade at a more favorable price. In a double auction market a transaction occurs when a market order hits the quote on the opposite side of the book. We assume that there is no information asymmetry and future price dynamics depend only upon public information. Without loss of generality, we further focus on the problem of a seller.

We model a trader who is alloted the task of selling one unit of the asset and has to complete the trade by a certain deadline. The current price of the security is determined by the best bid b_0, the highest buying price, and the best ask a_0, the lowest selling price. Choosing to trade at the market, the agent receives an immediate profit. If he adopts a limit order strategy, he will optimize the limit ask price K_a while being aware of a penalty in the non-execution event. We assume that in order to optimize the price he is willing to adopt a limit order strategy but is aware that there is a penalty for the non-execution event. We introduce the possibility of converting to a market order at maturity if his limit order was not filled. Therefore, if his limit order does not reach the front of the queue in the limit order book within the horizon T, the agent has to sell the asset trading at the best available price that guarantees immediate trading.[4]

[3]For instance, in 2005 SEC adopted a new regulation Rule 605 that requires all market centers to disclose certain order execution information, facilitating market transparency. This regulation requires, among other things, to make publicly available the information on the order execution speed.

[4]The assumption that order size is one unit ensures that the market will absorb the trade. However, one should bear in mind that

In a perfectly liquid market even if the initial limit order at a certain price $K_a > a_0$ was not picked before T, a sell limit order submitted at the best ask is filled straightaway. Thus, a time-zero discounted payoff of the agent's strategy equals the liquidation value of the asset:

$$V\left(K_a;\vec{\nu}\right) = e^{-\delta\tau}K_a I_{\{\tau\leq T\}} + e^{-\delta T}a_T I_{\{\tau>T\}}, \tag{1}$$

where $\vec{\nu}$ is a vector of market parameters and $\tau = \inf\{t \geq 0 : \ a_t = K_a\}$. Essentially, in a perfectly liquid market, $a_t = b_t$ must hold at all times.

Once a random delay is introduced, the only sure immediate sell is at the best bid. In a situation with a random delay the agent cannot bear additional risk at maturity. Hence submitting a limit order at a_T is not a sure trade and he is forced to use a market order instead and sell the security at the best bid b_T. This strategy yields the following expression for the discounted terminal payoff:

$$V_\lambda\left(K_a;\vec{\nu}\right) = e^{-\delta(\tau+\varepsilon)}K_a I_{\{\tau+\varepsilon\leq T\}} + e^{-\delta T}b_T I_{\{\tau+\varepsilon>T\}}, \tag{2}$$

with $\tau = \inf\{t \geq 0 : \ a_t = K_a\}$ and $\varepsilon \sim Exp\left(\lambda\right)$.

Numerous studies solve for the optimal strategy from the standpoint of a risk-neutral agent. However, if risks cannot be hedged away, the trader is concerned not only about the expected payoff but also about the range where the future payoff might lay, therefore he should take into account the variance in the future wealth. We assume that the trader is ascribed with a φ degree of risk-aversion; he determines the optimal limit price K_a^* by maximizing a mean-variance utility function:

$$EU_\lambda(K_a;\vec{\nu},\varphi) \equiv \max_{K_a\geq a_0} \ \left\{E\left[V_\lambda(K_a;\vec{\nu})\right] - \varphi\cdot Var\left[V_\lambda(K_a;\vec{\nu})\right]\right\}. \tag{3}$$

The final decision of a trader is formed through comparison of b_0 and the maximum expected utility that can be attained via a limit order $EU_\lambda(K_a^*;\vec{\nu},\varphi)$.

The market we study is sufficiently liquid, implying that the price dynamics can be described by a continuous process. A pair of stochastic log-normal processes b_t and a_t describe the trajectories of the bid and ask prices in the book respectively:

$$da_t = \mu_a a_t dt + \sigma_a a_t dW_t^a \tag{4}$$

$$db_t = \mu_b b_t dt + \sigma_b b_t dW_t^b \tag{5}$$

where W_t^a and W_t^b are ρ-correlated standard Brownian motions, $E\left[dW_t^a\ dW_t^b\right] = \rho dt$. We analyze short-term decisions when market conditions do not change substantially, so the assumption deterministic price trends is viable.[5]

When a trader implements a limit order and submits to the book an order at his preferred ask price K_a, he is "competing" with other potential sellers, or the best ask process a_t. Placing a limit order far from the current quotes implies a more intense competition and increases the chances that the opportunity to trade will not arise before expiry time. Limit orders in the book are executed in first-in, first-out rule. However,

this does not necessarily hold for large orders and use market orders at maturity might inflict significant price discounts.

[5]The probability of a negative spread equals $Pr[b_t > a_t] = N\left(d_p\right)$, where $d_p = \frac{\ln\left(b_0/a_0\right)-\left|\bar{\sigma}_{ba}\right|^2 T/2}{\left|\bar{\sigma}_{ba}\right|\sqrt{T}}$ and $\bar{\sigma}_{ba} = (\rho\sigma_b - \sigma_a, \sigma_b\sqrt{1-\rho^2})$. In our numerical example we choose the parameter value in a way that this probability is low.

once the limit order of an agent becomes the best price in the market, a random delay before trading ensues. The primary source of the delay comes from the possibility of somebody else placing an order ahead of the trader in question. An impatient trader can arrive and put a market order at a "better" price. For instance, assume the trader intends to sell an asset and his order reaches the front of the queue, then another order arrives at a marginally lower price. If it happens, the patient trader loses the price priority while retaining the time priority at his price. Unless there is a fundamental shock, this price deviation should quickly recover and his order will trade soon. In other words, the delay indicates the time it will take the price to return to the trend value. Also, the delay in trading occurs due to the fact that somebody else might have put an order at exactly the same price but earlier than the agent in question. This is a salient feature of markets with hidden liquidity since traders normally have no information about invisible quote depth. As a result, placing an order at a given price they are unable to assess how long it will take to fill more aggressive orders in front of them. Another argument for the delay, albeit a minor one in the context of modern electronic markets, is an operational delay. The delay is usually small relative to the time horizon, but there is a small probability that the delay will be sufficiently long and hence distort the schedule of the agent's trading operations. Thus, in our market the transaction occurs with an unforeseen delay ε which we model as independent of the asset prices and sampled from an exponential distribution with constant intensity λ. Clearly, in most of the cases, once the market is trading close enough to the quote that an agent has previously submitted, it will not move away swiftly. This market design implies, as we will show later, that the existence of a spread is related to the costs of waiting.

In their empirical study based on survival analysis Lo et al. (2002) observe exponentially distributed trading times. The absence of the peak near zero can by partially attributed to the discrete nature of the data, whereas our result applies to continuous time. Cho and Nelling (2000) argue that given market orders arrive in a non-homogeneous Poisson process the waiting times of a limit order follow a Weibull distribution. The estimations from a duration model (based on TORQ data from Harris and Hasbrouck (1996)) suggest that a histogram of empirical observations resembling an exponential probability density as well as Weibull. Notably, our theoretical result is compatible with a Weibull distribution specification for a shape parameter smaller or equal to unity. Cincotti et al. (2005), using tick data from 7 different US financial markets, find that the distribution of trade waiting times is well approximated by a mixture of exponential processes and verify this implication on an agent-based artificial market model. In the next section we concentrate on the timescale of trading implied by our framework and confirm the consistency with the empirical findings mentioned above.

4 The Distribution of Trading Times

In this section we examine the properties of the waiting time implied by the market dynamics and the trader's submissions. We define the time-to-fill of a limit order as $\theta = \tau + \varepsilon$, where τ is the time it takes to become the best price and ε is a random delay in order execution. The probability density function of the time-to-fill, as depicted in Fig. 1, exhibits a negative exponential shape with a peak. We notice that as the intensity rate λ increases, the peak becomes more pronounced, implying that faster order execution is more likely other things equal.

The relationship between the distribution of a trading time in a market without delay and in a market with

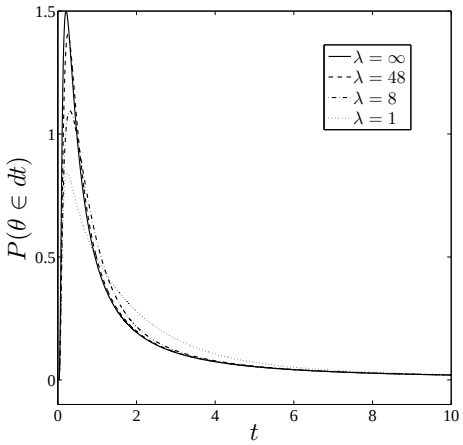

Figure 1: The probability density function of the time-to-fill of a limit order. Results are shown for parameters in Table II and $K_a/a_0 = 1.01$.

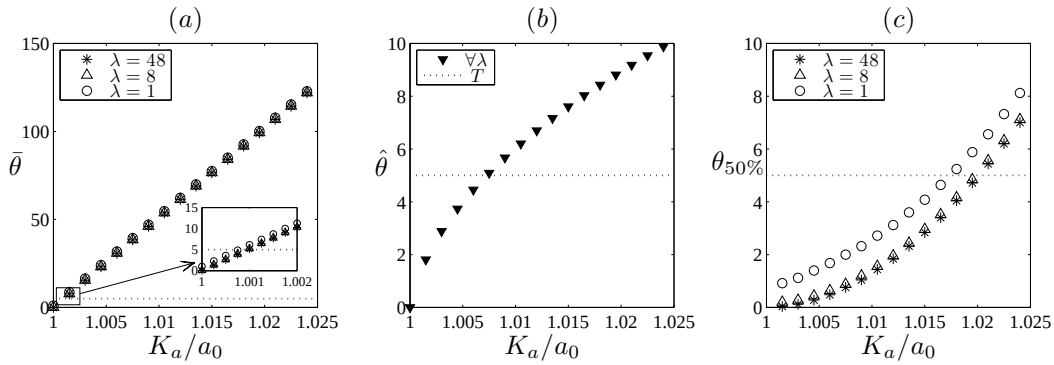

Figure 2: The characteristics of the most common time-to-fill of sell limit orders submitted at different prices $K_a \geq a_0$: (a) the average time-to-fill, (b) the mode time-to-fill, (c) the median time-to-fill. Results are shown for parameters in Table II.

a random delay is stated in the following proposition.

Proposition 1. *Let $P(\tau \leq t)$ be the cumulative distribution of limit order time-to-fill in a perfectly liquid market, then the cumulative distribution function of waiting times in a market with small average delay $P(\theta \leq t)$ is approximately*

$$P(\theta \leq t) = P(\tau \leq t) - \frac{x/\lambda}{\sigma_a t \sqrt{t}}\, n\left(\frac{x - A_1\sigma_a^2 t}{\sigma_a \sqrt{t}}\right). \tag{6}$$

The relationship between the probability density functions for two types of markets is:

$$P(\theta \in dt) = P(\tau \in dt)\left[1 - \frac{(x^2 - A_1^2\sigma_a^4 t^2 - 3t)/\lambda}{2\sigma_a^2 t^2}\right]. \tag{7}$$

Proof. See A.1. □

Since random delay ε follows an exponential distribution, this proposition reveals that the discrepancy

between waiting time in a market with infinite liquidity and a less liquid market is subject to an exponential component. The result complies with the intuition that the difference in limit order execution is most sensitive to market liquidity when medium-length maturities are concerned; both equations (6) and (7) show a negative dependence on a time factor.

In addition to that, we look at the impact of a random delay on the time-to-fill. The statistics are presented in Table I. Taking into account that prices are independent of delays, we find the mean time-to-fill as $E[\theta] = E[\tau] + E[\varepsilon]$. As Figure 2a reveals, the average waiting time $\bar{\theta}$ grows linearly with the distance-to-fill of a limit order. However, the contribution of the market liquidity parameter λ is very small and there is no marked difference in the submitted limit price, especially if the price grid is coarse.

Table I: The descriptive statistics of the time-to-first of a limit order and of a random delay.

Statistic	first passage time	delay	time-to-fill
mean	$\bar{\tau} = \frac{\ln\left(\frac{K_a}{a_0}\right)}{\mu_a - \sigma_a^2/2}$	$\bar{\varepsilon} = \frac{1}{\lambda}$	$\bar{\theta} = \frac{\ln\left(\frac{K_a}{a_0}\right)}{\mu_a - \sigma_a^2/2} + \frac{1}{\lambda}$
mode	$\hat{\tau} = \frac{\sqrt{4\ln\left(\frac{K_a}{a_0}\right)^2 (\mu_a - \sigma_a^2/2)^2 + 9\sigma_a^4} - 3\sigma_a^2}{2(\mu_a - \sigma_a^2/2)^2}$	$\hat{\varepsilon} = 0$	$\hat{\theta} = \frac{\sqrt{4\ln\left(\frac{K_a}{a_0}\right)^2 (\mu_a - \sigma_a^2/2)^2 + 9\sigma_a^4} - 3\sigma_a^2}{2(\mu_a - \sigma_a^2/2)^2}$
median	$P(\tau < \tau_{50\%}) = \frac{1}{2}$	$\varepsilon_{50\%} = \frac{\ln 2}{\lambda}$	$P(\theta < \theta_{50\%}) = \frac{1}{2}$

Given that the distribution of the time-to-fill is considerably skewed, the mean is not an informative statistic. Fig. 2b shows the mode time-to-fill $\hat{\theta}$, i.e. the most common value, for a range of limit prices. For any level of liquidity, the more aggressive a limit order strategy, the shorter is the mode of the execution time. The mode of the sum of independent variables is the sum of their modes and since the mode of exponentially distributed random variable is zero, the framework does not distinguish between the most common time-to-fill $\hat{\theta}$ and the most common first passage time $\hat{\tau}$. Further, we plot the median of the time-to-fill $\theta_{50\%}$ – a-50%-probability outcome (Fig. 2c). For instance, if a trader operating in a perfectly liquid market submits a sell order at 2% above the ask, then it is equally likely to execute earlier or later than his horizon $T = 5$ days; whereas, in a market with imperfect liquidity $\lambda = 1$ he must use a more aggressive order at 1.7% above a_0 to achieve this. As Fig. 2c demonstrates, the median trading time increases as passive orders fill the book. Also, the higher is the liquidity λ, the shorter is the median time-to-fill.

Empirical research also reveals that there is a correlation between execution time and limit order prices and the causality of this relationship is bilateral. Tkatch and Kandel (2006) find a significant causal impact of expected execution time on investors' decisions of which orders to submit. Lo et al. (2002) demonstrate that limit order execution times increase as limit prices become more passive and move further away from the quotes, which is partially an outcome of the price priority rule. According to Fig. 2a, the expected time-to-fill is longer for sell limit orders at higher prices K_a. Moreover, Lo et al. (2002) estimate the cumulative probability densities of the actual limit order execution times and compare to their hypothetical counterpart calculated as the first hitting times of geometric Brownian motion. The latter appears to understate largely expected trading times. Histograms of time-to-execution for limit orders exhibit exponential distribution and a comparison reveals that they differ not only in one or two moments but over their entire support. We observe changes in the median time-to-fill (Fig. 2c) in a market with a random delay as compared to

the perfect liquidity case. Without analysing directly how a particular trader's order placement affects the market, our model captures the statistical properties of trading times and relates the expected time-to-fill to order aggressiveness.

5 Optimal Strategy

The optimal strategy of a trader consists of two decisions. Essentially, he is choosing between placing a limit order at the optimal price K_a^* and a market order at b_0. In other words, the trader goes with the strategy which yields the highest utility. In the following proposition we derive the power function of a limit order payoff.

Proposition 2. *Denote the power function of the profit from selling a unit of security as* $G(K_a;\vec{v},\gamma) \equiv [V_\lambda(K_a;\vec{v})]^\gamma$. *Given that market prices are positively correlated log-normal processes (4) and (5), the expected value of G for selling at a limit price* $K_a \geq a_0$ *in a market with a random delay in limit order execution is given by expression:*[6]

$$\begin{aligned} EG(K_a;\vec{v},\gamma) &= K_a^\gamma \frac{\lambda}{\lambda+\gamma\delta}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2} N\left(\frac{-x-A_2\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_2} N\left(\frac{-x+A_2\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)\right. \\ &- e^{-(\lambda+\gamma\delta)T}\left(\left(\frac{K_a}{a_0}\right)^{A_1+A_3} N\left(\frac{-x-A_3\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_3} N\left(\frac{-x+A_3\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)\right)\Bigg] \\ &+ b_0^\gamma\, e^{\gamma\left(\mu_b+(\gamma-1)\sigma_b^2/2-\delta\right)T}\left[N\left(\frac{x-A_4\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)-\left(\frac{K_a}{a_0}\right)^{2A_4} N\left(\frac{-x-A_4\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)\right. \\ &+ e^{-\lambda T}\left(\left(\frac{K_a}{a_0}\right)^{A_4+A_5} N\left(\frac{-x-A_5\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_4-A_5} N\left(\frac{-x+A_5\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)\right)\Bigg] \end{aligned} \tag{8}$$

where $x = \ln\left(\frac{K_a}{a_0}\right)$ *and the constants are calculated as follows:*
$A_1 = \frac{\mu_a-\sigma_a^2/2}{\sigma_a^2}$, $A_2 = \frac{\sqrt{(\mu_a-\sigma_a^2/2)^2+2\gamma\delta\sigma_a^2}}{\sigma_a^2}$, $A_3 = \frac{\sqrt{(\mu_a-\sigma_a^2/2)^2-2\lambda\sigma_a^2}}{\sigma_a^2}$, $A_4 = \frac{\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b}{\sigma_a^2}$, $A_5 = \frac{\sqrt{(\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b)^2-2\lambda\sigma_a^2}}{\sigma_a^2}$;
with the parameter constraints $\lambda \in \left[0; \frac{(\mu_a-\sigma_a^2/2)^2}{2\sigma_a^2}\right]$ *and* $\rho \geq 0$.

Proof. See A.2. □

Given log-normal asset prices, the first passage time has an inverse Gaussian distribution which belongs to the exponential family. The upper bound on the parameter $\bar{\lambda} \equiv \left(\mu_a - \sigma_a^2/2\right)^2/2\sigma_a^2$ coincides with one of the two natural parameters of the inverse Gaussian distribution.[7]

Consider the case when a limit order to sell is submitted at the current ask. The expected value of this strategy is

$$EV_\lambda(K_a = a_0;\vec{v}) = a_0\frac{\lambda}{\lambda+\delta}\left(1-e^{-(\lambda+\delta)T}\right) + b_0\, e^{(\mu_b-\delta-\lambda)T}. \tag{9}$$

[6]The mean and the variance of the expected profit from selling the asset are: $EV_\lambda(K_a;\vec{v}) = EG(K_a;\vec{v},1)$ and $Var[V_\lambda(K_a;\vec{v})] = EG(K_a;\vec{v},2) - [EG(K_a;\vec{v},1)]^2$. Therefore, the expected utility equals $EU_\lambda(K_a;\vec{v},\varphi) = EG(K_a;\vec{v},1) - \varphi\left(EG(K_a;\vec{v},2) - [EG(K_a;\vec{v},1)]^2\right)$.

[7]The set of values for which the probability density function is finite on the entire support is called the natural parameter space. In particular, for the random variable with inverse Gaussian probability density $f_X(x;\omega_1,\omega_2)$, there are two natural parameters $\eta = \{-\frac{\omega_2}{2\omega_1^2}; -\frac{\omega_2}{2}\}$.

This expression suggests that the longer is the mean delay $\bar{\varepsilon} = 1/\lambda$, the smaller is the weight attached to the profit from using a limit order at the current ask $a_0$and the larger is the contribution of a market order submitted at maturity. In the extreme case of illiquidity when $\lambda = 0$, that is the delay is infinitely long, the expected profit from selling the asset equals the discounted expected best bid at maturity $\lim_{\lambda\to 0} EV_\lambda(K_a = a_0; \vec{v}) = e^{-\delta T} E[b_T]$. In fact, no limit order will trade before the maturity if $\lambda = 0$. More precisely, through formula (8) we reveal that the expected utility does not depend on K_a and equals exactly the expected utility of a market order at time T:

$$\lim_{\lambda\to 0} EU_\lambda(K_a; \vec{v}, \varphi) = b_0\, e^{(\mu_b-\delta)T} - \varphi b_0^2\, e^{2(\mu_b-\delta)T}\left[e^{\sigma_b^2 T} - 1\right]. \tag{10}$$

For an agents to prefer to sell with a market order upon maturity at b_T, instead of submitting a market order at b_0, the condition must hold:

$$\varphi \le \frac{e^{(\mu_b-\delta)T} - 1}{e^{2(\mu_b-\delta)T} b_0 \left(e^{\sigma_b^2 T} - 1\right)}. \tag{11}$$

It follows from here that the patience of the trader, represented by his risk-aversion coefficient φ, is a function of the trend in quotes trend on the opposite side of the limit order book: if the drift of the best bid μ_b is high enough many traders are motivated to provide liquidity by filling the limit order book with orders to sell.

Regarding the opposite situation when the market is perfectly liquid and λ is infinitely high, the utility of selling with a limit order converges to the perfect liquidity level given by zero bid-ask spread.

Proposition 3. *Denote the power function of the profit from selling a unit of security as $G(K_a; \vec{v}, \gamma) \equiv [V_\lambda(K_a; \vec{v})]^\gamma$. Given the market prices are ρ-correlated log-normal processes (4) and (5), the expected value of function G for selling at a limit price $K_a \ge a_0$ in a market with a small random delay in limit order execution converges to the value:*

$$\begin{aligned}\lim_{\lambda\to\infty} EG(K_a; \vec{v}, \gamma) &= K_a^\gamma \left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2} N\left(\frac{-x - A_2\sigma_a^2 T}{\sigma_a\sqrt{T}}\right) + \left(\frac{K_a}{a_0}\right)^{A_1-A_2} N\left(\frac{-x + A_2\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)\right] \\ &+ b_0^\gamma\, e^{\gamma(\mu_b + (\gamma-1)\sigma_b^2/2 - \delta)T} \left[N\left(\frac{x - A_4\sigma_a^2 T}{\sigma_a\sqrt{T}}\right) - \left(\frac{K_a}{a_0}\right)^{2A_4} N\left(\frac{-x - A_4\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)\right]\end{aligned} \tag{12}$$

where $x = \ln\left(\frac{K_a}{a_0}\right)$ and $A_1 = \frac{\mu_a - \sigma_a^2/2}{\sigma_a^2}$, $A_2 = \frac{\sqrt{(\mu_a - \sigma_a^2/2)^2 + 2\gamma\delta\sigma_a^2}}{\sigma_a^2}$, $A_4 = \frac{\mu_a - \sigma_a^2/2 + \gamma\rho\sigma_a\sigma_b}{\sigma_a^2}$.

Proof. See A.3. □

Notably, this result replicates the valuation formula derived in Iori et al. (2003) for the market where limit orders trade upon achieving the beginning of the queue on the relevant side. However, in accordance with the payoff in equation (1), in an infinitely liquid market a trader will resolve to a limit order at a_T in the non-execution event rather than pick the best buying order at b_T. Nonetheless, this setting actually requires a zero bid-ask spread, or $a_t = b_t,\ \forall t$.

A closed-form representation (8) given in Proposition 2 is valid for $0 \le \lambda \le \bar{\lambda}$. However, the maximum value $\bar{\lambda}$ poses a strong liquidity restriction, as we later discuss in the numerical example. In order to assess the range of values that the expected utility takes, we compute its upper and lower bounds.

Proposition 4. *Denote the power function of the profit from selling a unit of security as* $G(K_a;\vec{\nu},\gamma) \equiv [V_\lambda(K_a;\vec{\nu})]^\gamma$. *Given the market prices are* ρ*-correlated log-normal processes (4) and (5), the expected value of function* G *for selling at a limit price* $K_a \geq a_0$ *in a market with a random delay lies in the interval between its upper bound*

$$\begin{aligned} EG^U(K_a;\vec{\nu},\gamma) &= K_a^\gamma \frac{\lambda}{\lambda+\gamma\delta}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2} N\left(\frac{-x-A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_2} N\left(\frac{-x+A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right. \\ &- \left. e^{-(\lambda+\gamma\delta)T}\left(\left(\frac{K_a}{a_0}\right)^{A_1+A_6} N\left(\frac{-x+A_6\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_6} N\left(\frac{-x-A_6\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right)\right] \\ &+ b_0^\gamma e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T} \end{aligned} \tag{13}$$

and its lower bound

$$\begin{aligned} EG^L(K_a;\vec{\nu},\gamma) &= K_a^\gamma \frac{\lambda}{\lambda+\gamma\delta}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2} N\left(\frac{-x-A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_2} N\left(\frac{-x+A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right. \\ &- \left. e^{-\gamma\delta T}\left(N\left(\frac{-x-A_1\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{2A_1} N\left(\frac{-x+A_1\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right)\right] \\ &+ b_0^\gamma e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}\left[N\left(\frac{x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)-\left(\frac{K_a}{a_0}\right)^{2A_4} N\left(\frac{-x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right. \\ &+ \left. e^{-\lambda T}\left(\left(\frac{K_a}{a_0}\right)^{A_4+A_7} N\left(\frac{-x+A_7\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_4-A_7} N\left(\frac{-x-A_7\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right)\right] \end{aligned} \tag{14}$$

with $x = \ln\left(\frac{K_a}{a_0}\right)$ *and* $A_1 = \frac{\mu_a-\sigma_a^2/2}{\sigma_a^2}$, $A_2 = \frac{\sqrt{(\mu_a-\sigma_a^2/2)^2+2\gamma\delta\sigma_a^2}}{\sigma_a^2}$, $A_4 = \frac{\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b}{\sigma_a^2}$, $A_6 = \frac{\sqrt{(\mu_a-\sigma_a^2/2)^2+2\lambda\sigma_a^2}}{\sigma_a^2}$, $A_7 = \frac{\sqrt{(\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b)^2+2\lambda\sigma_a^2}}{\sigma_a^2}$.

Proof. See A.4. □

The values of the power function $EG(K_a;\vec{\nu},\gamma)$ fall, by construction, between the boundary functions (13) and (14). However, the interval given by $EG^U(K_a;\vec{\nu},\gamma)$ and $EG^L(K_a;\vec{\nu},\gamma)$ is rather wide, and these boundary functions do not produce an accurate solution to the problem (3). For intermediate values of $\bar{\lambda} < \lambda \ll \infty$ we use numerical integration methods to calculate the expected utility and find the optimal limit price.

6 Special Cases

We focus on two important special cases that allow us to separate two types of risks inherent to limit orders. First, we examine the situation when the time constraint is removed. Second, we look into the limiting behavior of the quotes chosen by the trader when the volatility of a traded asset is extremely low.

6.1 Infinite Time Horizon

Consider the case of infinitely long time horizon. We expect to detect a monotonic increasing relationship between the limit price K_a and the expected profit. The intuition behind this reasoning is that once the time

pressure is removed, the penalty which incurs paying the spread at maturity, simply vanishes. Both for illiquid and liquid markets,[8] we derive that when the bid price growth is slower than the discount rate ($\mu_b - \sigma_b^2/2 < \delta$), the expected profit from selling an asset is

$$\lim_{T\to\infty} EV_\lambda(K_a;\vec{v}) = \frac{\lambda}{\lambda+\delta} K_a \left(\frac{K_a}{a_0}\right)^{A_1 - A_2}. \tag{15}$$

It is easy to see that expression in (15) is smaller than K_a, since $K_a \geq a_0$ and $A_1 > A_2$ for any values of the parameters, the discrepancy being especially stark when the liquidity λ is low.

6.2 No Price Uncertainty

When the volatility of a security which the trader has to liquidate approaches zero, the price of this security will be changing at a constant rate per unit of time. Therefore, the maximum that the best ask can attain until maturity is known to be $a_T = a_0 e^{\mu_a T}$, and the best bid is $b_T = b_0 e^{\mu_b T}$. The moment when the limit sell order K_a will hit the quote is calculated as $\tau = \frac{\ln(K_a/a_0)}{\mu_a}$. In this situation the trader bears no price risk and the only risk he faces is linked to the non-execution of his orders due to delays.[9]

$$\lim_{\sigma\to 0} EV_\lambda(K_a;\vec{v}) = \frac{\lambda}{\lambda+\delta} K_a \left[\left(\frac{K_a}{a_0}\right)^{-\delta/\mu_a} - e^{-(\lambda+\delta)T}\left(\frac{K_a}{a_0}\right)^{\lambda/\mu_a}\right] + b_0 e^{(\mu_b-\delta-\lambda)T}\left(\frac{K_a}{a_0}\right)^{\lambda/\mu_a}. \tag{16}$$

Considering the deterministic nature of prices, the trader will optimize his strategy only over the subset of limit prices from the interval $a_0 \leq K_a \leq a_0 \cdot e^{\mu_a T}$ to ensure that $\tau \leq T$.

7 Numerical Example

7.1 Baseline Parameters

The values of parameters that we use in our numerical example are given in Table II. We set the time-zero best ask in the book equal to $a_0 = 1000$ with the initial spread of 5. The discount factor in this market is $\delta = 5\%$ per annum while the expected drift parameters are $\mu = 10\%$ and the annual volatility of $\sigma = 20\%$ with a small positive correlation in prices $\rho = 0.1$. It follows immediately from the assumption of no information asymmetry that all traders are equally informed about the value of the security. The bid and ask prices represent valuation on demand and supply sides of the same asset and should not diverge significantly at any point in time. Given these considerations, we work with the cases when the drifts of bid and ask are the same in order to preserve stationarity of the spread.[10] We choose a time horizon $T = 5$ days in a benchmark case. The delays are characterized by the intensity rate $\lambda = 8$ per day, and risk-aversion parameter is $\varphi = 0.01$ if not specified otherwise.

[8] In case $\lambda \leq \bar{\lambda}$ it follows from the result in Proposition 2 and for the case $\lambda > \bar{\lambda}$ we apply the Laplace integral approximation in 1 for fixed λ and $T \to \infty$.

[9] $EV_\lambda(K_a;\vec{v}) = E\left[K_a e^{-\delta(\tau+\varepsilon)} I_{\{\varepsilon\leq T-\tau\}} + b_T^\gamma e^{-\delta T} I_{\{\varepsilon > T-\tau\}}\right] = K_a e^{-\delta\tau}\int_0^{T-\tau} e^{-\delta\varepsilon}\lambda e^{-\lambda\varepsilon} d\varepsilon + b_0 e^{(\mu_b-\delta)T}\int_{T-\tau}^{\infty}\lambda e^{-\lambda\varepsilon} d\varepsilon = K_a e^{-\delta\tau}\frac{\lambda}{\lambda+\delta}\left(1 - e^{-(\lambda+\delta)(T-\tau)}\right) + b_0 e^{(\mu_b-\delta)T} e^{-\lambda(T-\tau)}$ and substituting τ yields the result in (16).

[10] Condition $\mu_a = \mu_b$ implies that a process b_t/a_t is a martingale and $E\left[\frac{b_t}{a_t}\right] = \frac{b_0}{a_0}$.

Table II: Baseline parameters.

Parameter	Value	Unit
initial prices	ask: $a_0 = 1000$	£
	bid: $b_0 = 995$	£
price correlation	$\rho = 0.1$	
drift	$\mu_a = 10\%$	per annum
	$\mu_b = 10\%$	per annum
volatility	$\sigma_a = 20\%$	per annum
	$\sigma_b = 20\%$	per annum
discount factor	$\delta = 5\%$	per annum
time horizon	$T = 5$	days
delay intensity	$\lambda = 8$	day^{-1}
risk aversion	$\varphi = 0.01$	

Using analytical expressions (12) and (8) the expected payoff in a perfectly and imperfectly liquid markets respectively, we find the optimal limit price which a trader should submit to achieve the maximum level of expected mean-variance utility. Further, we implement numerical integration with recursive adaptive Simpson quadrature rule to obtain the solution to the optimization problem for medium levels of liquidity.[11] We then examine the impact of various model parameters on the optimal decision of a trader. In order to increase the efficiency we set a precision grid equal to 0.25 and discretize the solution obtained from the continuous time model.

7.2 Comparative Static Effects

Prior to the discussion of the effects of various market parameters on the optimal order placement we look at the shape of the expected quadratic utility function. There are three possible market patterns: sideway market, up-trend market and down-trend market. A straightforward interpretation of the drift is to link it to the extent of traders' optimism or pessimism about the future price movements based on the past market performance. The greater the expected surge in the asset price, the higher the chances of a sell limit order getting filled before expiry.[12] Our analysis concerns trading in a risky asset, so a drift higher than the discount factor $\delta = 5\%$ implies a rise in stock price. This, in turn, makes limit orders more attractive for the investor who is willing to sell.

We compare the expected utilities of a risk-neutral (Fig. 3) and risk-averse (Fig. 4) traders. The expected profit that a risk-neutral agent attains trading in a perfectly liquid limit order book is the same for any limit price he might choose to submit (Fig. 3a) if the expected return of the underlying security equals risk-free rate. If the price trend is downward sloping – the best choice would be to sell at the current prevailing price, while upward trend implies that a trader should use an infinitely high price. This relationship matches the

[11] Our choice of baseline market parameters implies the median delay $\varepsilon_{50\%} = 0.0866$ days, which is sufficiently lower than the trader's time horizon.

[12] If $\mu \leq \sigma^2/2$ then the probability density function of the first passage time is defective and its integral over $[0, \infty)$ does not attain unity, consequently, the unconditional probability that a limit order will never get filled is strictly positive.

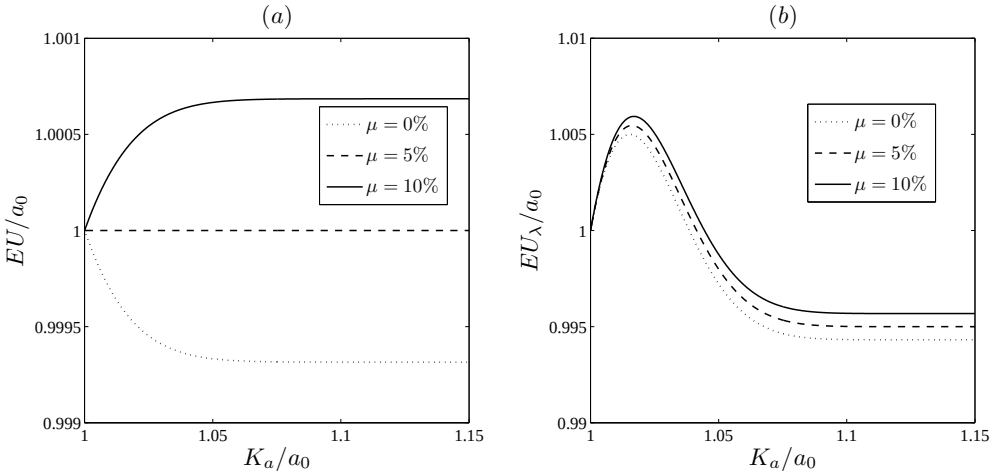

Figure 3: The expected utility of a sell limit order strategy for a risk-neutral trader ($\varphi = 0$) as a function of the limit price: (a) no delay in execution ($a_0 = b_0 = 1000$), (b) in presence of a random delay in limit order execution ($a_0 = 1000$, $b_0 = 995$). Results are shown for parameters in Table II.

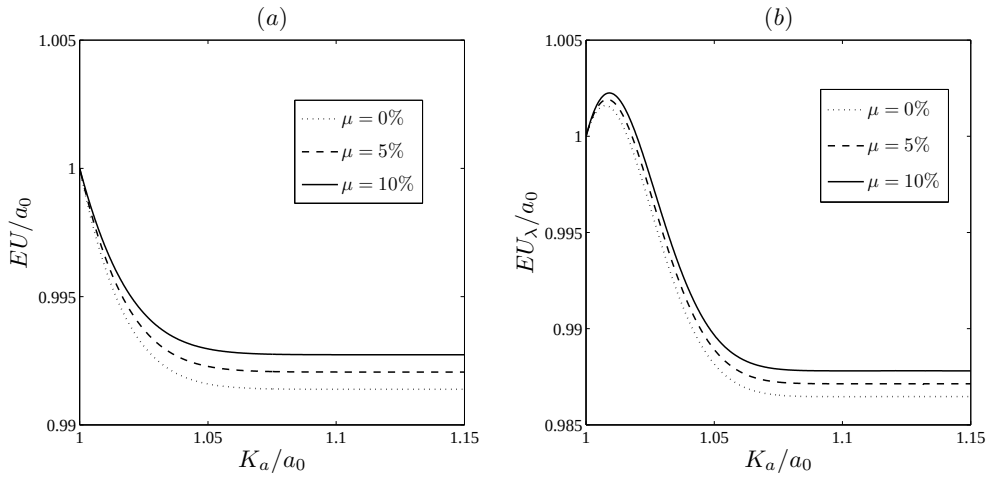

Figure 4: The expected utility of a sell limit order strategy for a risk-averse trader ($\varphi = 0.01$) as a function of the limit price: (a) no delay in execution ($a_0 = b_0 = 1000$), (b) in presence of a random delay in limit order execution ($a_0 = 1000$, $b_0 = 995$). Results are shown for parameters in Table II.

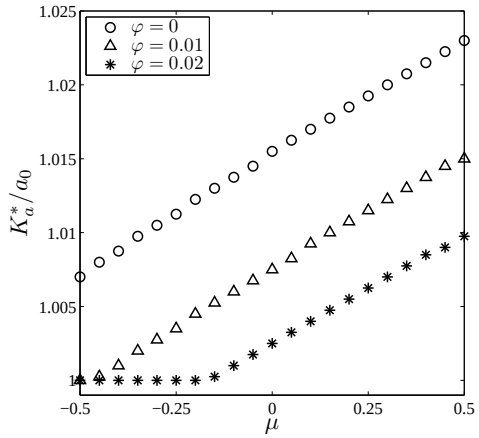

Figure 5: The optimal limit sell price for a risk-averse trader as a function of the price drift in presence of a random delay in limit order execution. Results are shown for parameters in Table II.

result obtained by Iori et al. (2003), who prove for a market without spread that: if $\mu > \delta$ the trader always waits till maturity ($K_a^* = \infty$), if $\mu < \delta$ the trader sells today with a market order, and if prices are martingales ($\mu = \delta$) the trader can submit any price $K_a \geq a_0$. In contrast, the expected utility of a trader selling via limit orders in an illiquid market retains a concavo-convex shape for a range of risk aversion coefficients, including $\varphi = 0$ (Fig. 3b and Fig. 4b).

Focusing on the effect of the drift parameter we notice that the optimal K_a^* in a situation without delay and no risk aversion is much more sensitive to the change in the price drift μ. In effect, the conclusion to be drawn from Fig. 3a and Fig. 4a is that perfect liquidity implies binary choice: either to trade at the current quote or post an infinitely high limit sell price.

Once a random delay is introduced to the market, the optimal limit price K_a^* increases linearly with μ, as Fig. 5a clearly reveals, while perfect liquidity assumes limit prices $K_a^* = \{a_0 \vee \infty\}$. Denote $\hat{\mu}$ the highest price drift for which the optimal selling strategy is a limit order at $K_a = a_0$. Other things being equal, the threshold $\hat{\mu}$ after which an agent switches to a more passive one is predictably moving rightward as risk aversion raises: it is below zero for $\varphi \leq 0.02$ (Fig. 5b). Thereupon, in a market with imperfect liquidity a trader tends to choose a more passive strategy for certain values of the price drift, while the optimal decision of the same trader in an absolutely liquid market under the same circumstances is not uniquely defined.

The limit price is diminishing as the bid-ask spread at the time of order placement increases Fig. 6. Using Paris Bourse order flow data, Biais et al. (1995) find that limit orders prevail at times of wide spreads and market orders – at times of narrow spreads. This tendency is compatible with the trading pattern of arbitrageurs and high frequency traders. Our result is in line with Ranaldo (2004) who observes that order aggressiveness of patient traders increases as spread widens.

Fig. 7 demonstrates the effect of the delay intensity on K_a^*.[13] The more risk-averse the trader, the smaller is an absolute impact of delay characteristic on the optimum. For relatively high levels of risk tolerance ($\varphi = \{0; 0.01; 0.02\}$) there is a positive relationship between liquidity λ and the optimal limit price K_a^*. This is an intuitively appealing result: an increase in the intensity rate λ implies a decrease in the mean delay (as well

[13] Here we use the exact formula for the power utility in (8) for small $\lambda \leq \bar{\lambda}$ to obtain the optimal submission price.

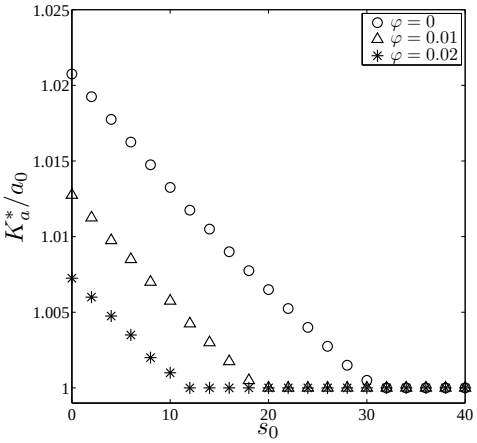

Figure 6: The optimal limit sell price for a risk-averse trader as a function of initial spread s_0. Results are shown for parameters in Table II.

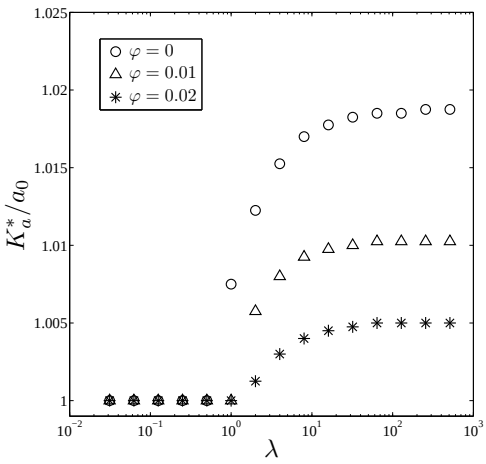

Figure 7: The optimal limit sell price for a risk-averse trader as a function of market liquidity. Results are shown for parameters in Table II.

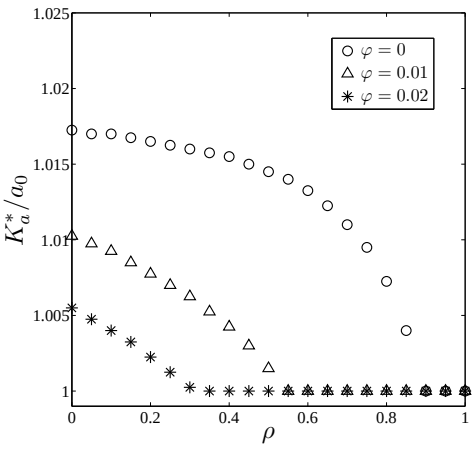

Figure 8: The optimal limit sell price for a risk-averse trader as a function of the bid and ask correlation in presence of a random delay in limit order execution. Results are shown for parameters in Table II.

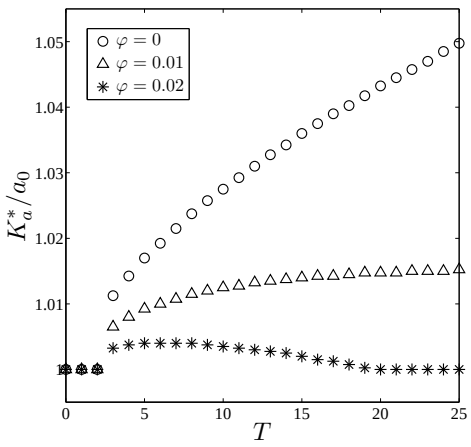

Figure 9: The optimal limit sell price for a risk-averse trader as a function of the expiry time in presence of a random delay in limit order execution. Results are shown for parameters in Table II.

as its variance), thus, other things being equal, leads to a shorter time-to-completion of a trade. The results for the different levels of risk tolerance suggest that an increase in φ decreases the optimal K_a^*, as expected.

Further, we examine the effect of the correlation between the paths of the best bid and ask prices. In a market with a random delay higher correlation coefficient reduces gradually the optimal limit sell price the agent should submit (Fig. 8). If the bid and ask are moving in parallel, if the order is not filled by T then the agent's penalty is only the spread. Essentially, the correlation controls the widths of the spread. When the drifts are equal and there is a perfect correlation in bid and ask trajectories, we essentially deal with the situation of a constant spread. High correlation implies less uncertainty about the spread size in the future, so the agent is more patient and trades with a limit order. When the correlation is low, the spread can narrow or widen at maturity, therefore the trader prefers more aggressive strategies to avoid this risk.

Lastly, the order aggressiveness decreases as the maturity extends (Fig. 9). This result confirms previous statements that longer time horizons imply higher chances that a_t reaches the barrier represented by the trader's limit price K_a^*. Again, this proves the consistency of our theoretical framework. Moreover, the optimal strategy of a risk averse trader in an illiquid market is apparently less respondent to a change in expiry time; so the impact of φ is that waiting for too long becomes equivalent to selling straightaway in terms of the expected utility. This is also in line with the result (15): when the time horizon T is very long, the order placement choice is not contingent upon it.

8 The Equilibrium Spread

It follows from the preceding discussion that in a continuous double auction the optimal decision of every market participant is made on the basis of existing best bid and ask prices which define the market spread. Meanwhile, the choices and actions that traders take now affect the current market spread and, as a direct consequence, influence future trading decisions. In fact, the market spread is the result of the interactions between heterogeneous agents who populate the market.

In the quote-driven markets the bid-ask spread is charged by the market maker in order to cover the expenses incurred by trading against better informed agents. The reasons for the bid-ask spread in a double-auction markets are more subtle; its presence is justified in several ways in the literature: information asymmetry which we have discussed before, gravitational pull effect depicted by Cohen et al. (1981), variations in the state of the book and traders valuations for the asset, trading costs, which in turn comprise direct costs such as commissions, transfer taxes order submission fees and account service fees, and indirect cost – the difference between the price at which the transaction was actually carried and a certain fair price. The delay parameter absorbs all these nuances which, in essence, make the difference between trading with a limit order rather than a market order.

Cohen et al. (1981) examine this issue in detail and show that a positive market spread is incumbent to this market microstructure. According to Cohen et al. (1981), the equilibrium spread in a dynamic trading system is "the bid-ask spread at which, for the next instant of time, the probability of the spread increasing is equal to the probability of the spread decreasing." Adopting his definition to our framework we arrive at the condition (see B.1):

$$\frac{\mu_a}{\mu_b} = \frac{b_t}{a_t}. \tag{17}$$

This condition, as Cohen et al. (1981) emphasize, does not guarantee that the market will eventually settle at this equality, rather that it is more likely that the price will move towards this condition than in the opposite direction. Since the drift parameters are constant, the condition in equation (17) simply states the equilibrium level of spread is constant and is determined by the gap between the growth rates of bid and ask prices. However, in order to preserve non-negative spread we must have $a_t \geq b_t, \ \forall t$, or $\mu_a \leq \mu_b$ in equilibrium. This slightly counter-intuitive contingency arises from the specific behavior of log-normal process – an increase and a decrease by the same amounts are not equally likely. If we impose for the purpose of stationarity $\mu_a = \mu_b$ then we must observe a zero absolute spread in equilibrium – equation (17) requires $a_t = b_t, \ \forall t$.

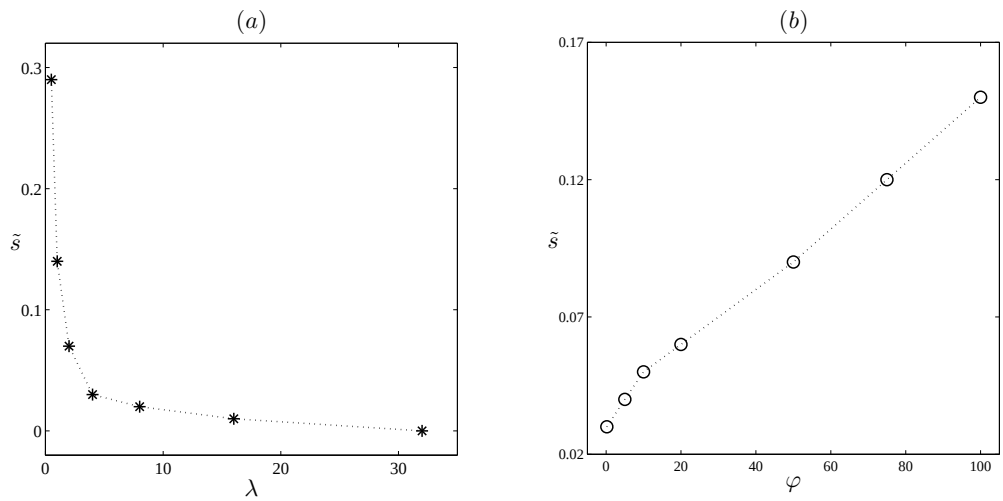

Figure 10: Equilibrium spread a s a function of liquidity and risk tolerance of agents: (a) for different value of delay intensity λ, (b) for different levels of risk aversion φ. Results are shown for parameters in Table II.

Nevertheless, this definition does not account for the notion of delay which is the crucial characteristic of the market we examine. It is precisely the deviation from the perfect liquidity case inflicted by large delays in limit order execution that determines the real spread size.

Definition (Equilibrium Spread)**.** *In a dynamic trading process the equilibrium market spread is the bid-ask spread such that the expected utility from trading via a limit order at the optimal limit price is equal to the utility from an immediate market order.*[14]

Applying numerical optimization we find the implied equilibrium spread given the expected delay.[15] It has been discussed in the empirical literature that the main determinants of the spread size are competition for liquidity and risk aversion of market participants (Ranaldo, 2004). The higher is the competition among traders to provide liquidity, the tighter is the observed bid-ask spread, whereas the degree of risk aversion of the traders has a positive impact on the spread size. As depicted in Fig. 10a, the size of spread $\tilde{s}$ diminishes and eventually attains zero as the value of λ increases, thereby decreasing the expected delay. In Fig. 10b we can see that the equilibrium spread increases, as expected, with the degree of risk aversion φ.

[14]This is an interpretation of a definition suggested by Harris (2003) (p.304): "The spread which ensures that traders are indifferent between using a limit order and a market order is the *equilibrium spread*."

[15]The necessary derivations are presented in B.2.

Although this result is obtained from the standpoint of a seller, the pattern for the buyer would be symmetrical in our setting. Moreover, since the limit orders are convertible in our model, seller has to monitor both sides of the book. In other words, a market design where a limit order is executed immediately the moment it becomes the best price in the market requires a zero bid-ask spread.

9 Conclusions

We have developed a logically consistent and empirically plausible model, which is easy to estimate. The central feature of the present study is the analysis of the impact of a random delay in limit order execution on the optimal strategy of a risk-averse trader. Based on an analytical solution, we examine the effects of various market parameters on his optimal selling strategy.

Our framework both benefits from transparency and explains the trade-off between immediacy and a favorable transaction price. In contrast with standard first passage time models of trading, it captures the fundamental difference between time required to reach the beginning of the queue on the relevant side of the market and the time-to-completion of a trade. The probability density of the expected time-to-fill of limit orders sharpens as liquidity increases and reveals an empirically observed exponential distribution of trading times. The discrepancy, as the model confirms, is due to imperfect liquidity, which in turn defers the trade execution. The main result suggests that the introduction of a random delay factor alleviates the impact of various market conditions on the optimal limit price the trader submits. Using comparative statics analysis, we demonstrate that the presence of a lag factor linearizes these effects. Notably, it is not the magnitude but the mere presence of delay that alters the nature of the relationship. Furthermore, we determine the equilibrium market spread as the bid-ask spread such that the expected utility from trading via a limit order at the optimal limit price is equal to the profit from an immediate market order. We subsequently prove that, consistent with real market phenomena, the equilibrium bid-ask spread increases both as liquidity decreases and agents' risk aversion increases.

In addition to that we have demonstrated that in distinction from the profit-maximizing case, the introduction of risk-aversion factor provides the results which are more coherent with empirical observations and seem to be more useful for the practical implementation. The mean-variance utility function permits adequate risk assessment for a strategy involving limit order trading, therefore our approach allows to model the trading trajectories of heterogeneous investors.

We provide a solution for a static problem which can be extended to multi-period submission steps and solved in a manner of Harris (1998). However, we expect that results will not alter qualitatively once a trader is allowed to revise his strategy a finite number of times. Whereas we analyzed only small trades, large trades should be examined differently since they have a price impact when market orders are used to execute the trade. There is a separate branch of literature on the order splitting issues which are closely related to our framework (Almgren and Chriss, 2000; Obizhaeva and Wang, 2005; Alfonsi et al., 2010; Løkka, 2011). In a recent paper by Guéant et al. (2011), the authors propose a novel approach of splitting a large trade using limit orders rather than market orders. This is the direction for the future development of our framework.

A Proofs

A.1 Proof of Proposition 1

We write a log-normal process as $a_t = a_0 e^{X_t}$ with $X_t = \left(\mu_a - \sigma_a^2/2\right)t + \sigma_a W_t$. Then the first passage time is $\tau = \inf\{t \geq 0 | X_t = x\}$ with $x = \ln(K_a/a_0)$. Denote the maximum of the X_t over a time period T as $M_T^X = \max\{X_t | 0 \leq t \leq T\}$. By the reflection principle $\forall\ x \geq 0$ we have

$$P(\tau \leq t) = P\left(M_t^X \geq x\right).$$

The cumulative distribution of the maximum of a Brownian motion with drift ξ and volatility ν and $x \geq 0$ is equal to

$$P\left(M_t^X \leq x\right) = N\left(\frac{x - \xi t}{\nu\sqrt{t}}\right) - e^{2\xi x/\nu^2} N\left(\frac{-x - \xi t}{\nu\sqrt{t}}\right), \tag{A.1}$$

and

$$P(\tau \in dt) = \frac{|x|}{\sqrt{2\pi\nu^2 t^3}} e^{\frac{-(x-\xi t)^2}{2\nu^2 t}}. \tag{A.2}$$

Since in our framework the prices follow geometric Brownian motion, the first passage time has an inverse Gaussian distribution with the probability density function $f(\tau)$ given by (A.2) with parameters $\nu = \sigma_a$ and $\xi = \mu_a - \sigma_a^2/2$. The delay variable follows an exponential distribution with a positive parameter λ, then $f(\varepsilon) = \lambda e^{-\lambda\varepsilon}$. In the presence of a random delay in limit order execution we denote time-to-fill as $\theta = \tau + \varepsilon$ and obtain its distribution:

$$\begin{aligned} P(\theta \leq t) &= \int_0^t f(\tau) \int_0^{t-\tau} f(\varepsilon) d\varepsilon d\tau \\ &= \int_0^t f(\tau) \int_0^{t-\tau} \lambda e^{-\lambda\varepsilon} d\varepsilon d\tau = \int_0^t f(\tau)\left(1 - e^{-\lambda(t-\tau)}\right) d\tau \\ &= \int_0^t f(\tau) d\tau - e^{-\lambda t} \int_0^t e^{\lambda\tau} f(\tau) d\tau. \end{aligned} \tag{A.3}$$

The first term of this expression is independent of λ and equals exactly the probability density function of the first passage time, while the second term accounts for the random delay. However, if the parameter λ takes large values, then the integrand of the second term is not finite. We find the limit of the second term for high λ using Laplace's method. Define an integral

$$\mathcal{I}(\lambda) = \int_a^b h(\tau) e^{\lambda g(\tau)} d\tau, \tag{A.4}$$

where $[a, b]$ is a finite interval and functions $h(\tau)$ and $g(\tau)$ are continuous.

Lemma 1. *Suppose the function $g(\tau)$ attains a maximum on $[a, b]$ at either endpoint, $\tau_0 = a$ or $\tau_0 = b$, and is differentiable in a neighborhood of τ_0, with $g'(\tau_0) \neq 0$ and $h(\tau_0) \neq 0$. Then the leading term of the asymptotic expansion of the integral (A.4) , as $\lambda \to +\infty$, is given by*

$$\mathcal{I}(\lambda) = \frac{h(\tau_0) \cdot e^{\lambda g(\tau_0)}}{\lambda |g'(\tau_0)|}. \tag{A.5}$$

We apply this lemma to functions $g(\tau) = \tau$ and $h(\tau) = \frac{|\alpha|}{\tau\sqrt{\tau}} n\left(\frac{\alpha+\beta\tau}{\sqrt{\tau}}\right)$ with integration limits $a = 0$, $b = t$. The function $g(\tau)$ attains the highest value at $\tau_0 = b = t$ and $g'(\tau_0) = 1$, therefore we obtain from formula (A.5):

$$\begin{aligned}\int_0^t e^{\lambda\tau} \frac{|\alpha|}{\tau\sqrt{\tau}} n\left(\frac{\alpha+\beta\tau}{\sqrt{\tau}}\right) d\tau &= \frac{\frac{|\alpha|}{t\sqrt{t}}\, n\left(\frac{\alpha+\beta t}{\sqrt{t}}\right) \cdot e^{\lambda t}}{\lambda \cdot |1|} \\ &= e^{\lambda t} \frac{|\alpha|}{\lambda t\sqrt{t}} \cdot n\left(\frac{\alpha+\beta t}{\sqrt{t}}\right). \end{aligned} \tag{A.6}$$

With $\alpha = x/\sigma_a$ and $\beta = -A_1\sigma_a$, we re-write (A.3) as follows:

$$P(\theta \le t) = P(\tau \le t) - \frac{x/\lambda}{\sigma_a t\sqrt{t}}\, n\left(\frac{x - A_1\sigma_a^2 t}{\sigma_a\sqrt{t}}\right), \tag{A.7}$$

under on the probability space $(\mathcal{P}, \Omega, \mathcal{F})$ with $x = \ln\left(\frac{K_a}{a_0}\right)$ and $A_1 = \frac{\mu_a - \sigma_a^2/2}{\sigma_a^2}$.

Note that the probability density of first passage time $P(\tau \le t)$ is defined in equation (A.2). Taking the derivative of the second term in expression (A.7) with respect to t, we arrive at the probability density function of limit order time-to-fill θ:

$$P(\theta \in dt) = \frac{x}{\sigma_a t\sqrt{t}}\, n\left(\frac{x - A_1\sigma_a^2 t}{\sigma_a\sqrt{t}}\right)\left[1 - \frac{(x^2 - A_1^2\sigma_a^4 t^2 - 3t)/\lambda}{2\sigma_a^2 t^2}\right] = P(\tau \in dt)\left[1 - \frac{(x^2 - A_1^2\sigma_a^4 t^2 - 3t)/\lambda}{2\sigma_a^2 t^2}\right]. \tag{A.8}$$

It follows immediately from (A.8) that when the expected delay approaches zero time-to-fill is equivalent to the time it takes to reach the front of the queue: $\lim_{\lambda\to\infty} P(\theta \in dt) = P(\tau \in dt)$.

A.2 Proof of Proposition 2

In order to find the analytical expression for the power of the limit order strategy payoff we need to calculate $EG(K_a; \vec{v}, \gamma) \equiv E\left[K_a^\gamma e^{-\gamma\delta(\tau+\varepsilon)} I_{\{\tau+\varepsilon\le T\}} + b_T^\gamma e^{-\gamma\delta T} I_{\{\tau+\varepsilon>T\}}\right]$.

We want to calculate $J_1 \equiv E\left[K_a^\gamma e^{-\gamma\delta(\tau+\varepsilon)} I_{\{\tau+\varepsilon\le T\}}\right]$ and $J_2 \equiv E\left[b_T^\gamma e^{-\gamma\delta T} I_{\{\tau+\varepsilon>T\}}\right]$. Assuming that the sopping time and delays are independent and implementing the integrated expectations formula[16] we arrive at the following expression:

$$\begin{aligned} J_1 &= K_a^\gamma E\left[e^{-\gamma\delta(\tau+\varepsilon)} I_{\{\tau\le T\}} I_{\{\varepsilon\le T-\tau\}}\right] = K_a^\gamma E\left[e^{-\gamma\delta\tau} I_{\{\tau\le T\}} E\left[e^{-\gamma\delta\varepsilon} I_{\{\varepsilon\le T-\tau\}} \middle| \tau\right]\right] \\ &= K_a^\gamma \int_0^T e^{-\gamma\delta\tau} f(\tau) \int_0^{T-\tau} e^{-\gamma\delta\varepsilon} f(\varepsilon)\, d\varepsilon d\tau. \end{aligned} \tag{A.9}$$

The probability density function of the first passage time $f(\tau)$ is given in (A.2); the delay variable follows

[16] $E(X) = E(E(X|Y))$

an exponential distribution with $f(\varepsilon) = \lambda e^{-\lambda\varepsilon}$. Therefore, we simplify the integral in (A.9):

$$\begin{aligned}\int_0^T e^{-\gamma\delta\tau} f(\tau) \int_0^{T-\tau} e^{-\gamma\delta\varepsilon} f(\varepsilon)\, d\varepsilon d\tau &= \int_0^T e^{-\gamma\delta\tau} f(\tau) \left[\int_0^{T-\tau} e^{-\gamma\delta\varepsilon} \lambda e^{-\lambda\varepsilon} d\varepsilon\right] d\tau \\ &= \int_0^T e^{-\gamma\delta\tau} f(\tau) \frac{\lambda}{\lambda+\gamma\delta}\left(1 - e^{-(\lambda+\gamma\delta)(T-\tau)}\right) d\tau \\ &= \frac{\lambda}{\lambda+\gamma\delta}\left[\int_0^T e^{-\gamma\delta\tau} f(\tau)\, d\tau - e^{-(\lambda+\gamma\delta)T} \int_0^T e^{\lambda\tau} f(\tau)\, d\tau\right].\end{aligned}$$

Using equation (A.2) with $\nu = \sigma_a$ and $\xi = \mu_a - \sigma_a^2/2$ we rewrite the first term as

$$J_1 = K_a^{\gamma} \frac{\lambda}{\lambda+\gamma\delta}\left[\int_0^T e^{-\gamma\delta\tau} \frac{|x|}{\sigma_a \tau \sqrt{\tau}} n\left(\frac{x - \left(\mu_a - \sigma_a^2/2\right)\tau}{\sigma_a \sqrt{\tau}}\right) d\tau - e^{-(\lambda+\gamma\delta)T} \int_0^T e^{\lambda\tau} \frac{|x|}{\sigma_a \tau \sqrt{\tau}} n\left(\frac{x - \left(\mu_a - \sigma_a^2/2\right)\tau}{\sigma_a \sqrt{\tau}}\right) d\tau\right]. \tag{A.10}$$

Since the following equality holds for the normal density

$$e^{-\phi t} n\left(\frac{\alpha + \beta t}{\sqrt{t}}\right) = e^{-\alpha\beta + \alpha\sqrt{\beta^2 + 2\phi}} n\left(\frac{\alpha + t\sqrt{\beta^2 + 2\phi}}{\sqrt{t}}\right), \tag{A.11}$$

with $\alpha = x/\sigma_a, \beta = -(\mu_a - \sigma_a^2/2)/\sigma_a$ and $\phi = \gamma\delta$ we rearrange the first component of J_1 as

$$\int_0^T e^{-\gamma\delta\tau} \frac{|x|}{\sigma_a \tau \sqrt{\tau}} n\left(\frac{x - \left(\mu_a - \sigma_a^2/2\right)\tau}{\sigma_a \sqrt{\tau}}\right) d\tau = \frac{|x|}{\sigma_a} \int_0^T \frac{1}{\tau\sqrt{\tau}}\left[e^{x(A_1+A_2)} n\left(\frac{x + A_2\sigma_a^2\tau}{\sigma_a\sqrt{\tau}}\right)\right] d\tau,$$

where $A_1 = \frac{\mu_a - \sigma_a^2/2}{\sigma_a^2}$ and $A_2 = \frac{\sqrt{(\mu_a - \sigma_a^2/2)^2 + 2\gamma\delta\sigma_a^2}}{\sigma_a^2}$.

Finally, using the identity

$$\int_0^T \frac{1}{t\sqrt{t}} n\left(\frac{\alpha + \beta t}{\sqrt{t}}\right) dt = \frac{1}{|\alpha|}\left[N\left(\frac{-|\alpha|}{\sqrt{T}} - sgn(\alpha)\beta\sqrt{T}\right) + e^{-2\alpha\beta} N\left(\frac{-|\alpha|}{\sqrt{T}} + sgn(\alpha)\beta\sqrt{T}\right)\right], \tag{A.12}$$

substituting $\alpha = x/\sigma_a, \beta = A_2\sigma_a$ and $K_a \geq a_0$, we get

$$\frac{|x|}{\sigma_a} \int_0^T \frac{1}{\tau\sqrt{\tau}}\left[e^{x(A_1+A_2)} n\left(\frac{x + A_2\sigma_a^2\tau}{\sigma_a\sqrt{\tau}}\right)\right] d\tau = \left(\frac{K_a}{a_0}\right)^{A_1+A_2} N\left(\frac{-x - A_2\sigma_a^2 T}{\sigma_a\sqrt{T}}\right) + \left(\frac{K_a}{a_0}\right)^{A_1-A_2} N\left(\frac{-x + A_2\sigma_a^2 T}{\sigma_a\sqrt{T}}\right). \tag{A.13}$$

The second integral in J_1 is modified via the following identity which holds for the values $\lambda \leq \beta^2/2$:[17]

$$e^{\lambda t} n\left(\frac{\alpha + \beta t}{\sqrt{t}}\right) = e^{-\alpha\beta + \alpha\sqrt{\beta^2 - 2\lambda}} n\left(\frac{\alpha + t\sqrt{\beta^2 - 2\lambda}}{\sqrt{t}}\right). \tag{A.14}$$

[17] $e^{\lambda t} n\left(\frac{\alpha+\beta t}{\sqrt{t}}\right) = e^{\lambda t}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha+\beta t}{\sqrt{t}}\right)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2+\beta^2 t^2+2\alpha\beta t-2\lambda t^2}{2t}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha+\sqrt{\beta^2-2\lambda}t}{\sqrt{t}}\right)^2 - \alpha\beta + \alpha\sqrt{\beta^2-2\lambda}t} = e^{-\alpha\beta+\alpha\sqrt{\beta^2-2\lambda}} n\left(\frac{\alpha+t\sqrt{\beta^2-2\lambda}}{\sqrt{t}}\right)$

Precisely, we obtain

$$\int_0^T e^{\lambda\tau}\frac{|x|}{\sigma_a\tau\sqrt{\tau}}n\left(\frac{x-\left(\mu_a-\sigma_a^2/2\right)\tau}{\sigma_a\sqrt{\tau}}\right)d\tau=\frac{|x|}{\sigma_a}\int_0^T\frac{1}{\tau\sqrt{\tau}}\left[e^{x(A_1+A_3)}n\left(\frac{x+A_3\sigma_a^2\tau}{\sigma_a\sqrt{\tau}}\right)\right]d\tau,$$

where $A_3=\frac{\sqrt{(\mu_a-\sigma_a^2/2)^2-2\lambda\sigma_a^2}}{\sigma_a^2}$. Applying formula (A.12) with $\alpha=x/\sigma_a,\beta=A_3\sigma_a$ and $K_a\geq a_0$

$$\frac{|x|}{\sigma_a}\int_0^T\frac{1}{\tau\sqrt{\tau}}\left[e^{x(A_1+A_3)}n\left(\frac{x+A_3\sigma_a^2\tau}{\sigma_a\sqrt{\tau}}\right)\right]d\tau=\left(\frac{K_a}{a_0}\right)^{A_1+A_3}N\left(\frac{-x-A_3\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_3}N\left(\frac{-x+A_3\sigma_a^2T}{\sigma_a\sqrt{T}}\right). \tag{A.15}$$

Combining the results (A.13) and (A.15) we arrive that at the final expression for the first term:

$$\begin{aligned}J_1 &= K_a^\gamma\frac{\lambda}{\lambda+\gamma\delta}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2}N\left(\frac{-x-A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_2}N\left(\frac{-x+A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right.\\ &- \left.e^{-(\lambda+\gamma\delta)T}\left(\left(\frac{K_a}{a_0}\right)^{A_1+A_3}N\left(\frac{-x-A_3\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_3}N\left(\frac{-x+A_3\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right)\right].\end{aligned}$$

Before calculating the second term we rewrite the bid and ask equations to express these processes in terms of a two-dimensional Brownian motion (W_1,W_2):

$$\begin{aligned}da_t &= a_t\left(\mu_a dt+\bar{\sigma_a}d\bar{W}_t\right)\\ db_t &= b_t\left(\mu_b dt+\bar{\sigma_b}d\bar{W}_t\right),\end{aligned}$$

where $\bar{\sigma_a}=(\sigma_a,\ 0)$ and $\bar{\sigma_b}=\left(\sigma_b\rho,\ \sigma_b\sqrt{1-\rho^2}\right)$. Therefore,

$$\begin{aligned}a_t &= a_0e^{(\mu_a-|\bar{\sigma_a}|^2/2)t+\bar{\sigma_a}\bar{W}_t}\\ b_t &= b_0e^{(\mu_b-|\bar{\sigma_b}|^2/2)t+\bar{\sigma_b}\bar{W}_t}.\end{aligned}$$

Notice that if b_t is a log-normal process then b_t^γ is also log-normally distributed. Applying Ito's formula to this process we can show that

$$db_t^\gamma=b_t^\gamma\left(\gamma\mu_b+\frac{\gamma(\gamma-1)}{2}\sigma_b^2\right)dt+\gamma b_t^\gamma\bar{\sigma_b}d\bar{W}_t. \tag{A.16}$$

Hence, applying Ito's lemma once again we get $b_T^\gamma=b_0^\gamma e^{(\gamma\mu_b+\gamma(\gamma-1)\sigma_b^2/2)T+\gamma\bar{\sigma_b}\bar{W}_T}$. Now we rewrite the second term

$$\begin{aligned}J_2 &= e^{-\gamma\delta T}E\left[b_0^\gamma e^{\left(\gamma\mu_b+\frac{\gamma(\gamma-1)}{2}\sigma_b^2\right)T+\gamma\bar{\sigma_b}\bar{W}_T}\cdot I_{\{\varepsilon>T-\tau\}}\cdot\left(I_{\{\tau>T\}}+I_{\{\tau\leq T\}}\right)\right]\\ &= b_0^\gamma\ e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}\ E^*\left[I_{\{\tau>T\}}+I_{\{\varepsilon>T-\tau\}}I_{\{\tau\leq T\}}\right]\\ &= b_0^\gamma\ e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}\ \left(E^*\left[I_{\{\tau>T\}}\right]+E^*\left[I_{\{\tau\leq T\}}E^*\left[I_{\{\varepsilon>T-\tau\}}\middle|\tau\right]\right]\right)\\ &= b_0^\gamma\ e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}\ \left(\int_T^\infty f(\tau)\,d\tau+\int_0^T f(\tau)\int_{T-\tau}^\infty f(\varepsilon)\,d\varepsilon d\tau\right),\end{aligned} \tag{A.17}$$

where expectation is calculated under probability measure P^* defined by Radon-Nikodym derivative $\eta_T = \frac{dP^*}{dP} = e^{\bar{\sigma}_b W_T - |\bar{\sigma}_b|^2 T}$. As Girsanov theorem states, $\bar{W}_t^* = \bar{W}_t - \gamma\bar{\sigma}_a t$ is a two-dimensional Brownian motion under P^*, so we rewrite

$$X_T = (\mu_a - \sigma_a^2/2)T + \bar{\sigma}_a\left(\bar{W}_T^* - \gamma\bar{\sigma}_a T\right) = (\mu_a + (\gamma - 1/2)\sigma_a^2)T + \bar{\sigma}_a\bar{W}_T^*,$$

implying a drift $\xi^* = \mu_a - \sigma_a^2/2 + \gamma\rho\sigma_a\sigma_b$ under P^*.

The first integral in (A.17) is the probability that the first passage time exceeds horizon T, which is given in formula (A.1) for $\nu = \sigma_a$ and ξ^*:

$$\int_T^\infty \frac{|x|}{\sigma_a\tau\sqrt{\tau}} n\left(\frac{x - \left(\mu_a - \sigma_a^2/2 + \gamma\rho\sigma_a\sigma_b\right)\tau}{\sigma_a\sqrt{\tau}}\right)d\tau = N\left(\frac{x - A_4\sigma_a^2 T}{\sigma_a\sqrt{T}}\right) - \left(\frac{K_a}{a_0}\right)^{2A_4} N\left(\frac{-x - A_4\sigma_a^2 T}{\sigma_a\sqrt{T}}\right), \quad \text{(A.18)}$$

where $A_4 = \frac{\mu_a - \sigma_a^2/2 + \gamma\rho\sigma_a\sigma_b}{\sigma_a^2}$. Furthermore, simplifying the last term in (A.17) yields

$$\int_0^T f(\tau)\int_{T-\tau}^\infty f(\varepsilon)\,d\varepsilon d\tau = \int_0^T f(\tau)\left[\int_{T-\tau}^\infty \lambda e^{-\lambda\varepsilon}d\varepsilon\right]d\tau = \int_0^T f(\tau)\,e^{-\lambda(T-\tau)}d\tau = e^{-\lambda T}\int_0^T e^{\lambda\tau}f(\tau)\,d\tau.$$

We substitute the probability density function $f(\tau)$ with volatility σ_a and drift $\xi^* = \mu_a - \sigma_a^2/2 + \gamma\rho\sigma_a\sigma_b$ and using property (A.14) obtain

$$\int_0^T e^{\lambda\tau}\frac{|x|}{\sigma_a\tau\sqrt{\tau}} n\left(\frac{x - \left(\mu_a - \sigma_a^2/2 + \gamma\rho\sigma_a\sigma_b\right)\tau}{\sigma_a\sqrt{\tau}}\right)d\tau = \frac{|x|}{\sigma_a}\int_0^T \frac{1}{\tau\sqrt{\tau}}\left[e^{x(A_4+A_5)} n\left(\frac{x + A_5\sigma_a^2\tau}{\sigma_a\sqrt{\tau}}\right)\right]d\tau, \quad \text{(A.19)}$$

where $A_5 = \frac{\sqrt{(\mu_a - \sigma_a^2/2 + \gamma\rho\sigma_a\sigma_b)^2 - 2\lambda\sigma_a^2}}{\sigma_a^2}$ and $\lambda \leq (\mu_a - \sigma_a^2/2 + \gamma\rho\sigma_a\sigma_b)^2/2\sigma_a^2$. Given the parameters values $\alpha = x/\sigma_a, \beta = A_5\sigma_a$ and $K_a \geq a_0$ we write an explicit expression for this integral as shown in formula (A.12)

$$\frac{|x|}{\sigma_a}\int_0^T \frac{1}{\tau\sqrt{\tau}}\left[e^{x(A_4+A_5)} n\left(\frac{x + A_5\sigma_a^2\tau}{\sigma_a\sqrt{\tau}}\right)\right]d\tau, = \left(\frac{K_a}{a_0}\right)^{A_4+A_5} N\left(\frac{-x - A_5\sigma_a^2 T}{\sigma_a\sqrt{T}}\right) + \left(\frac{K_a}{a_0}\right)^{A_4-A_5} N\left(\frac{-x + A_5\sigma_a^2 T}{\sigma_a\sqrt{T}}\right). \quad \text{(A.20)}$$

Using (A.18) and (A.20),

$$\begin{aligned} J_2 &= b_0^\gamma\, e^{\gamma(\mu_b + (\gamma-1)\sigma_b^2/2 - \delta)T}\left[N\left(\frac{x - A_4\sigma_a^2 T}{\sigma_a\sqrt{T}}\right) - \left(\frac{K_a}{a_0}\right)^{2A_4} N\left(\frac{-x - A_4\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)\right. \\ &+ \left. e^{-\lambda T}\left(\left(\frac{K_a}{a_0}\right)^{A_4+A_5} N\left(\frac{-x - A_5\sigma_a^2 T}{\sigma_a\sqrt{T}}\right) + \left(\frac{K_a}{a_0}\right)^{A_4-A_5} N\left(\frac{-x + A_5\sigma_a^2 T}{\sigma_a\sqrt{T}}\right)\right)\right]. \end{aligned}$$

Thus,

$$
\begin{aligned}
EG(K_a;\vec{v},\gamma) &= K_a^{\gamma}\frac{\lambda}{\lambda+\gamma\delta}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2} N\left(\frac{-x-A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_2} N\left(\frac{-x+A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right.\\
&- e^{-(\lambda+\gamma\delta)T}\left(\left(\frac{K_a}{a_0}\right)^{A_1+A_3} N\left(\frac{-x-A_3\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_3} N\left(\frac{-x+A_3\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right)\Bigg]\\
&+ b_0^{\gamma}\, e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}\left[N\left(\frac{x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)-\left(\frac{K_a}{a_0}\right)^{2A_4} N\left(\frac{-x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right.\\
&+ e^{-\lambda T}\left(\left(\frac{K_a}{a_0}\right)^{A_4+A_5} N\left(\frac{-x-A_5\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_4-A_5} N\left(\frac{-x+A_5\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right)\Bigg],
\end{aligned}
$$

with $A_1=\frac{\mu_a-\sigma_a^2/2}{\sigma_a^2}$, $A_2=\frac{\sqrt{(\mu_a-\sigma_a^2/2)^2+2\gamma\delta\sigma_a^2}}{\sigma_a^2}$, $A_3=\frac{\sqrt{(\mu_a-\sigma_a^2/2)^2-2\lambda\sigma_a^2}}{\sigma_a^2}$, $A_4=\frac{\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b}{\sigma_a^2}$, $A_5=\frac{\sqrt{(\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b)^2-2\lambda\sigma_a^2}}{\sigma_a^2}$ and $\forall\ \lambda\in\left[0,\bar{\lambda}\right]$, where $\bar{\lambda}=\min\left\{\frac{(\mu_a-\sigma_a^2/2)^2}{2\sigma_a^2},\frac{(\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b)^2}{2\sigma_a^2}\right\}$.

A.3 Proof of Proposition 3.

We find the limit of the expected utility using Laplace's integral approximation method. Using the result in (A.6), we approximate the second integral in (A.9) and the integral in (A.19) when $\lambda\geq\beta^2/2$ and identity (A.14) does not hold. It is easy to show that for high values of λ the power utility is approximately

$$
\begin{aligned}
EG(K_a;\vec{v},\gamma) &\cong K_a^{\gamma}\frac{\lambda}{\lambda+\gamma\delta}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2} N\left(\frac{-x-A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_2} N\left(\frac{-x+A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right.\\
&- e^{-\gamma\delta T}\frac{x/\lambda}{\sigma_aT\sqrt{T}}\, n\left(\frac{x-A_1\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\Bigg]\\
&+ b_0^{\gamma}\, e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}\left[N\left(\frac{x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)-\left(\frac{K_a}{a_0}\right)^{2A_4} N\left(\frac{-x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right.\\
&+ \frac{x/\lambda}{\sigma_aT\sqrt{T}}\, n\left(\frac{x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\Bigg],
\end{aligned}
\tag{A.21}
$$

where $x=\ln\left(\frac{K_a}{a_0}\right)$, $A_1=\frac{\mu_a-\sigma_a^2/2}{\sigma_a^2}$, $A_2=\frac{\sqrt{(\mu_a-\sigma_a^2/2)^2+2\gamma\delta\sigma_a^2}}{\sigma_a^2}$, $A_4=\frac{\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b}{\sigma_a^2}$. Further, noting that constants A_1, A_2 and A_4 are independent of λ, in a perfectly liquid market we obtain

$$
\begin{aligned}
\lim_{\lambda\to\infty} EG(K_a;\vec{v},\gamma) &= K_a^{\gamma}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2} N\left(\frac{-x-A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_2} N\left(\frac{-x+A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right]\\
&+ b_0^{\gamma}\, e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}\left[N\left(\frac{x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)-\left(\frac{K_a}{a_0}\right)^{2A_4} N\left(\frac{-x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right].
\end{aligned}
$$

A.4 Proof of Proposition 4

The general form of the integral we need to bound is

$$\int_0^T e^{-\lambda(T-\tau)} f(\tau) d\tau, \tag{A.22}$$

where the density is $f(\tau) = \frac{|\alpha|}{\tau\sqrt{\tau}} n\left(\frac{\alpha+\beta\tau}{\sqrt{\tau}}\right)$ and λ takes large values. It is easy to see that

$$\forall \tau \in [0, T]: \qquad e^{-\lambda(T+\tau)} \leq e^{-\lambda(T-\tau)} \leq 1,$$

and since $f(\tau)$ is a bounded function we have

$$\int_0^T e^{-\lambda(T+\tau)} f(\tau) d\tau \leq \int_0^T e^{-\lambda(T-\tau)} f(\tau) d\tau \leq \int_0^T f(\tau) d\tau.$$

For the upper bound we apply identity (A.12)

$$\begin{aligned} \int_0^T f(\tau) d\tau &= \int_0^T \frac{|\alpha|}{\tau\sqrt{\tau}} n\left(\frac{\alpha+\beta\tau}{\sqrt{\tau}}\right) d\tau \\ &= N\left(\frac{-\alpha+\beta T}{\sqrt{T}}\right) + e^{-2\alpha\beta} N\left(\frac{-\alpha-\beta T}{\sqrt{T}}\right). \end{aligned} \tag{A.23}$$

In order to calculate the lower bound we first use property (A.11)

$$e^{-\lambda(T+\tau)} f(\tau) = e^{-\lambda T} \frac{|\alpha|}{\tau\sqrt{\tau}} \left[e^{-\lambda\tau}\, n\left(\frac{\alpha+\beta\tau}{\sqrt{\tau}}\right)\right] = e^{-\lambda T} \frac{|\alpha|}{\tau\sqrt{\tau}} \left[e^{-\alpha\beta+\alpha\sqrt{\beta^2+2\lambda}} n\left(\frac{\alpha+\tau\sqrt{\beta^2+2\lambda}}{\sqrt{\tau}}\right)\right],$$

then property (A.12) to get

$$\begin{aligned} \int_0^T e^{-\lambda(T+\tau)} f(\tau) d\tau &= |\alpha|\, e^{\alpha(\psi-\beta)-\lambda T} \int_0^T \frac{1}{\tau\sqrt{\tau}}\, n\left(\frac{\alpha+\psi\tau}{\sqrt{\tau}}\right) d\tau \\ &= e^{-\lambda T}\left[e^{\alpha(\psi-\beta)} N\left(\frac{-\alpha+\psi T}{\sqrt{T}}\right) + e^{-\alpha(\psi+\beta)} N\left(\frac{-\alpha-\psi T}{\sqrt{T}}\right)\right], \end{aligned} \tag{A.24}$$

where $\psi = \sqrt{\beta^2+2\lambda}$.

The full expression for expected power utility is

$$
\begin{aligned}
EG(K_a;\vec{v},\gamma) &= K_a^{\gamma}\frac{\lambda}{\lambda+\gamma\delta}\left[\int_0^T e^{-\gamma\delta\tau}\frac{|x|}{\sigma_a\tau\sqrt{\tau}}n\left(\frac{x-\left(\mu_a-\sigma_a^2/2\right)\tau}{\sigma_a\sqrt{\tau}}\right)d\tau\right.\\
&- \left. e^{-(\lambda+\gamma\delta)T}\int_0^T e^{\lambda\tau}\frac{|x|}{\sigma_a\tau\sqrt{\tau}}n\left(\frac{x-\left(\mu_a-\sigma_a^2/2\right)\tau}{\sigma_a\sqrt{\tau}}\right)d\tau\right]\\
&+ b_0^{\gamma}e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}\left[\int_T^{\infty}\frac{|x|}{\sigma_a\tau\sqrt{\tau}}n\left(\frac{x-\left(\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b\right)\tau}{\sigma_a\sqrt{\tau}}\right)d\tau\right.\\
&+ \left. e^{-\lambda T}\int_0^T e^{\lambda\tau}\frac{|x|}{\sigma_a\tau\sqrt{\tau}}n\left(\frac{x-\left(\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b\right)\tau}{\sigma_a\sqrt{\tau}}\right)d\tau\right].
\end{aligned}
$$

As shown in the the proof of Proposition 2, the first and the third integrals have analytical expression for all values of parameters. Since the second term in this expression is negative with $\alpha = x/\sigma_a$ and $\beta = -A_1\sigma_a$ we apply formula (A.24) to determine its upper bound

$$
I_2 \le -e^{-(\lambda+\gamma\delta)T}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_6}N\left(\frac{-x+A_6\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_6}N\left(\frac{-x-A_6\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right], \tag{A.25}
$$

where $A_6 = \frac{\sqrt{(\mu_a-\sigma_a^2/2)^2+2\lambda\sigma_a^2}}{\sigma_a^2}$, and formula (A.23) to determine its lower bound

$$
I_2 \ge -e^{-\gamma\delta T}\left[N\left(\frac{-x-A_1\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{2A_1}N\left(\frac{-x+A_1\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right]. \tag{A.26}
$$

Similarly, for the fourth term in the expected power utility function $\alpha = x/\sigma_a$ and $\beta = -A_4\sigma_a$ we apply formula (A.23) to determine its upper bound

$$
I_4 \le N\left(\frac{-x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{2A_4}N\left(\frac{-x+A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right) \tag{A.27}
$$

and (A.24) to find the lower bound

$$
I_4 \ge e^{-\lambda T}\left[\left(\frac{K_a}{a_0}\right)^{A_4+A_7}N\left(\frac{-x+A_7\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_4-A_7}N\left(\frac{-x-A_7\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right], \tag{A.28}
$$

where $A_7 = \frac{\sqrt{(\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b)^2+2\lambda\sigma_a^2}}{\sigma_a^2}$.

Finally, substituting (A.25) and (A.27) we obtain an *upper bound* of expected power utility:

$$
\begin{aligned}
EG^U(K_a;\vec{v},\gamma) &= K_a^{\gamma}\frac{\lambda}{\lambda+\gamma\delta}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2}N\left(\frac{-x-A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_2}N\left(\frac{-x+A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right.\\
&- \left. e^{-(\lambda+\gamma\delta)T}\left(\left(\frac{K_a}{a_0}\right)^{A_1+A_6}N\left(\frac{-x+A_6\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_6}N\left(\frac{-x-A_6\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right)\right]\\
&+ b_0^{\gamma}e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}.
\end{aligned}
$$

Using results (A.26) and (A.28) we obtain a *lower bound* of expected power utility:

$$\begin{aligned}
EG^L(K_a;\vec{\nu},\gamma) &= K_a^\gamma \frac{\lambda}{\lambda+\gamma\delta}\left[\left(\frac{K_a}{a_0}\right)^{A_1+A_2} N\left(\frac{-x-A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_1-A_2} N\left(\frac{-x+A_2\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right.\\
&- \left. e^{-\gamma\delta T}\left(N\left(\frac{-x-A_1\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{2A_1} N\left(\frac{-x+A_1\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right)\right]\\
&+ b_0^\gamma e^{\gamma(\mu_b+(\gamma-1)\sigma_b^2/2-\delta)T}\left[N\left(\frac{x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)-\left(\frac{K_a}{a_0}\right)^{2A_4} N\left(\frac{-x-A_4\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right.\\
&+ \left. e^{-\lambda T}\left(\left(\frac{K_a}{a_0}\right)^{A_4+A_7} N\left(\frac{-x+A_7\sigma_a^2T}{\sigma_a\sqrt{T}}\right)+\left(\frac{K_a}{a_0}\right)^{A_4-A_7} N\left(\frac{-x-A_7\sigma_a^2T}{\sigma_a\sqrt{T}}\right)\right)\right]
\end{aligned}$$

with $A_1 = \frac{\mu_a-\sigma_a^2/2}{\sigma_a^2}$, $A_2 = \frac{\sqrt{(\mu_a-\sigma_a^2/2)^2+2\gamma\delta\sigma_a^2}}{\sigma_a^2}$, $A_4 = \frac{\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b}{\sigma_a^2}$, $A_6 = \frac{\sqrt{(\mu_a-\sigma_a^2/2)^2+2\lambda\sigma_a^2}}{\sigma_a^2}$, $A_7 = \frac{\sqrt{(\mu_a-\sigma_a^2/2+\gamma\rho\sigma_a\sigma_b)^2+2\lambda\sigma_a^2}}{\sigma_a^2}$.

B Equilibrium Spread Derivations

B.1 Derivation of Condition (17).

Denote $S_t \equiv a_t - b_t$, then the probability that the spread will increase in an infinitely small time increment dt is equivalent to saying that the change in ask will be greater than the change in bid, independently of the direction.

$$\begin{aligned}
Pr\left[da_t \geq db_t\right] &= Pr\left[a_t\left(\mu_a dt+\bar{\sigma}_a dW_t^a\right) \geq b_t\left(\mu_b dt+\bar{\sigma}_b dW_t^b\right)\right]\\
&= Pr\left[\frac{b_t}{a_t}\left(\mu_b dt+\left(\rho\sigma_b\psi_1+\sigma_b\sqrt{1-\rho^2}\psi_2\right)\sqrt{dt}\right) \leq \mu_a dt+\sigma_a\psi_1\sqrt{dt}\right]\\
&= Pr\left[\left(\frac{b_t}{a_t}\rho\sigma_b-\sigma_a\right)\psi_1+\frac{b_t}{a_t}\sigma_b\sqrt{1-\rho^2}\psi_2 \leq \left(\mu_a-\frac{b_t}{a_t}\mu_b\right)\sqrt{dt}\right]\\
&= Pr\left[\psi_3 \leq z_t\right] = N\left(z_t\right),
\end{aligned}$$

where $\psi_j \sim N(0,1)$ are independent random variables, $z_t \equiv \frac{\mu_a-\mu_b b_t/a_t}{\nu_s}\sqrt{dt}$ with $\nu_s^2 = \frac{b_t^2}{a_t^2}\sigma_b^2+\sigma_a^2-2\rho\sigma_a\sigma_b\frac{b_t}{a_t}$. It is easy to see that variance ν_s^2 is strictly positive for non-zero volatilities of the asset prices σ_a and σ_b and $|\rho| \leq 1$.

Similarly, the probability of a narrower spread is expressed as

$$Pr\left[da_t \leq db_t\right] = Pr\left[\psi_3 \geq z_t\right] = Pr\left[\psi_3 \leq -z_t\right] = N\left(-z_t\right) = 1-N\left(z_t\right).$$

Equating two expressions yields:

$$\begin{aligned}
Pr\left[da_t \geq db_t\right] &= Pr\left[da_t \geq db_t\right]\\
N\left(z_t\right) &= 1-N\left(z_t\right)\\
\frac{\mu_a}{\mu_b} &= \frac{b_t}{a_t}.
\end{aligned}$$

B.2 Numerical Calculation of the Spread in a Market with Delay.

The preferences of a trader are described by a mean-variance utility function (3). The equilibrium spread $\tilde{s}$ is defined as a situation when the expected utility from using a limit order at the optimal price K_a^* equals the utility of the profit from trading via immediate market order. Let $\tilde{s} = a_t - b_t, \forall$, then

$$\tilde{s} = a_0 - EU_\lambda(K_a^*; \vec{v} \,\big|\, \tilde{s}). \tag{B.1}$$

We solve this problem in three steps. First, we substitute $b_0 = a_0 - s$ and calculate the expected utility of limit orders submitted at various prices $K_a \geq a_0$ for a large range of spreads $s \geq 0$:

$$\forall s \geq 0 : K_a^*(s) = \arg\max_{K_a \geq a_0} EU_\lambda(K_a; \vec{v} \,\big|\, s). \tag{B.2}$$

Second, for each spread size we determine the optimal limit price, therefore, we obtain an optimal limit price as a function of spread $K_a^* = K_a^*(s)$ for a range $s \geq 0$. Third, we infer pair of limit price and spread that satisfy condition (B.1).

References

Alfonsi, A., Fruth, A., Schied, A., 2010. Optimal execution strategies in limit order books with general shape functions. Quantitative Finance 10, 143–157.

Almgren, R., Chriss, N., 2000. Optimal execution of portfolio transactions. Journal of Risk 3, 5–39.

Angel, J., 1994. Limit versus market orders. Working paper, School of Business Administration, Georgetown University.

Biais, B., Hillion, P., Spatt, C., 1995. An empirical analysis of the limit order book and the order flow in the paris bourse. The Journal of Finance 50, 1655–1689.

Bouchaud, J., Farmer, D., Lillo, F., 2009. Handbook of Financial Markets: Dynamics and Evolution. Handbooks in Finance. North-Holland, Ch. How Markets Slowly Digest Changes in Supply and Demand, pp. 57–130.

Chakravarty, S., Holden, C., 1995. An integrated model of market and limit orders. Journal of Financial Intermediation 4, 213–241.

Cho, J., Nelling, E., 2000. The probability of limit order execution. Financial Analysts Journal 56, 28–33.

Cincotti, S., Ponta, L., andE. Scalas, M. R., 2005. Poisson-process generalization for the trading waiting time distribution in a double auction mechanism. In: Proceedings of the International Conference on Noise and Fluctuations in Econophysics and Finance. The International Society for Optical Engineering, Austin, USA.

Cohen, K., Maier, S., Schwartz, R., Whicomb, D., 1981. Transaction costs, order placement strategy, and existence of the bid-ask spread. Journal of Political Economy 89, 287–305.

Dayanik, S., Karatzas, I., 2003. On the optimal stopping problem for one-dimensional diffusions. Stochastic Processes and their Applications 107, 173–212.

Foucault, T., 1999. Order flow composition and trading costs in a dynamic limit order market. Journal of Financial Markets 2, 99–134.

Foucault, T., Kadan, O., Kandel, E., 2005. Limit order book as a market for liquidity. The Review of Financial Studies 18, 1171–1217.

Glosten, L., 1994. Is the electronic open limit order book inevitable? The Journal of Finance 49, 1127–1161.

Goettler, R., Parlour, C., Rajan, U., 2005. Equilibrium in a dynamic limit order market. The Journal of Finance 60, 2149–2192.

Gould, M., Porter, M., Williams, S., McDonald, M., Fenn, D., Howison, S., 2010. Limit order books, available at ArXiv: 1012.0349v2.

Guéant, O., Lehalle, C., Tapia, J., 2011. Optimal portfolio liquidation with limit orders, available at ArXiv: 1106.3279v4.

Handa, P., Schwartz, R., Tiwari, A., 2003. Quote setting and price formation in an order-driven market. Journal of Financial Markets 6, 461–489.

Harris, L., 1998. Optimal dynamic order submission strategies in some stylized trading problems. Financial Markets, Institutions, and Instruments 7, 1–76.

Harris, L., 2003. Trading and Exchange. Market Microstructure for Practitioners. Oxford University Press.

Harris, L., Hasbrouck, J., 1996. Market versus limit orders: the superdot evidence on order submission strategy. Journal of Financial and Quantitative Analysis 31, 213–231.

Hasbrouck, J., 2006. Empirical Market Microstructure: the Institutions, Economics, and Econometrics of Securities Trading. Oxford University Press.

Hedvall, K., N. J. R. G., 1997. Do buyers and sellers behave similarly in a limit order book? a high-frequency data examination of the finnish stock exchange. Journal of Empirical Finance 4, 279–293.

Hollifield, B., R.A. Miller, R., Sandås, P., 1999. An empirical analysis of limit order markets. Working Paper 2843, Center for Economic Policy Research (CEPR), University of Pennsylvania.

Huang, R., Stoll, H., 1997. The components of the bid-ask spread: A general approach. The Review of Financial Studies 10, 995–1034.

Iori, G., Padilla, F., Zervos, M., 2003. Optimal limit order strategies. Working paper, King's College London.

Kou, S., Wang, H., 2003. First passage times of a jump diffusion process. Advanced Applied Probabilities 35, 504–531.

Kumar, P., Seppi, D., 1994. Limit and market orders with optimizing traders. Working paper, Carnegie Mellon University.

Langnau, A., Punchev, Y., 2011. Stochastic price dynamics implied by the limit order book, available at ArXiv: 1105.4789v1.

Lo, A., MacKinlay, A., Zhang, J., 2002. Econometric models of limit-order executives. Journal of Financial Economics 65, 31–71.

Løkka, A., 2011. Optimal liquidation in a limit order book for a risk averse investor. Working paper, London School of Economics.

Mike, S., Farmer, J., 2008. An empirical behavioral model of liquidity and volatility. Journal of Economic Dynamics and Control 32, 200–234.

Milgrom, P., Stockey, N., 1982. information, trade and common knowledge. Journal of Economic Theory 26, 17–27.

Obizhaeva, A., Wang, J., 2005. Optimal trading strategy and supply/demand dynamics, available at SSRN: 686168.

Omura, K., Tanigawa, Y., Uno, J., 2000. Execution probability of limit orders on the tokyo stock exchange. Working paper, Graduate School of Economics, Osaka University, available at SSRN: 252588.

Parlour, C., 1998. Price dynamics in limit order markets. The Review of Financial Studies 11, 789–816.

Parlour, C., Seppi, D., 2008. Handbook of Financial Intermediation and Banking. Vol. 5. Elsevier, Ch. Limit Order Markets : A Survey, pp. 1–14.

Ranaldo, A., 2004. Order aggressiveness in limit order book markets. Journal of Financial Markets 7, 53–74.

Rock, K., 1996. The specialist's order book and price anomalies. Working paper, Harvard University.

Rosu, I., 2009. A dynamic model of the limit order book. Review of Financial Studies 22, 4601–4641.

Seppi, D., 1997. Liquidity provision with limit orders and a strategic specialist. The Review of Financial Studies 10, 103–150.

Tkatch, I., Kandel, E., 2006. Demand for immediacy of execution: Time *is* money. Working paper, Georgia State University.