

07-017

Pricing Liquidity: The Quantity Structure of Immediacy Prices

George C. Chacko

Jakub W. Jurek

Erik Stafford

Copyright © 2006 by George C. Chacko, Jakub W. Jurek, and Erik Stafford

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Pricing Liquidity: The Quantity Structure of Immediacy Prices*

George C. Chacko

Jakub W. Jurek

Erik Stafford

Abstract

This paper develops a model for understanding liquidity via the pricing of limit orders. Limit orders can be well defined and priced with the tools of option pricing, allowing the complex tradeoff between transaction size and speed to be reduced to a single price. The option-based framework allows the properties of liquidity to be characterized as functions of the fundamental value and the order flow processes. In the special case when immediate execution is desired, the option strike price at which immediate exercise is optimal determines the effective bid/ask price. A model with full-information, but imperfect market making, is able to describe many of the known properties of transaction costs.

First draft: May 2005

This draft: August 2006

JEL Classification: G11, G12

Keywords: Liquidity, limit order, American option, early exercise, transaction cost

*Chacko: Harvard Business School and Integrated Finance LTD - e-mail: gchacko@hbs.edu. Jurek: Harvard Department of Economics and Harvard Business School - e-mail: jjurek@hbs.edu. Stafford: Harvard Business School - e-mail: estafford@hbs.edu. We thank Joshua Coval, Robin Greenwood, Will Goetzmann, Bob Merton, André Perold, David Scharfstein, Anna Scherbina, Halla Yang, and seminar participants at Harvard Business School, INSEAD, The Bank of Italy, and State Street Bank.

1 Introduction

Researchers and participants in the market for securities have long been interested in the costs of transacting and the notion of liquidity as a performance measure of market structure. Corporations rely on these insights as they engage in large transactions that will affect prices except in the special situation where demand curves for securities are perfectly elastic. In real world capital markets, investors and corporations generally do not expect to transact at fundamental value. Rather, market participants face some degree of illiquidity, where they must sacrifice price, trade size, or speed of execution, forcing them to transact at prices away from fundamental value. The magnitude of this wedge as a function order quantity—the *price of liquidity*—is unknown. Despite widespread interest, a useful characterization of transaction prices for securities has been elusive.

This paper develops a parsimonious model of the price of liquidity via the pricing of limit orders. Limit orders are essentially American-type options, offering the right to trade up to a fixed quantity at a constant price for some period. The use of limit orders as a means of offering trade is common across many different markets, and their value is a natural measure of the cost of transacting.¹ Our model allows us to analytically characterize the schedule of price concessions at different transaction sizes for an investor desiring immediate execution—the *quantity structure of immediacy prices*—by exploiting the optimal early exercise conditions of the limit order. We find that our option-based model can describe many of the known properties of transaction costs and provides a convenient method for estimating the price of liquidity.

Liquidity has three important dimensions: price, quantity, and immediacy. A market is considered liquid if an investor can quickly execute a significant quantity at a price near fundamental value. In the limiting case of infinite liquidity, a perfect capital market is reached, where investors can instantaneously transact any quantity at precisely the security’s fundamental value. While this idealized description of capital markets may be appropriate in many settings, there are surely situations where it is inaccurate.

An inherent friction that limits liquidity in capital markets is the asynchronous arrival of

¹When a prospective buyer makes an offer to purchase up to a specific quantity at a specific price, the buyer is effectively submitting a limit order. Limit order is the name given to this type of transaction offer in financial markets, but the mechanism is common in many other markets. Therefore, even in markets without a formalized limit order structure, many transaction mechanisms can be viewed as limit orders.

buyers and sellers, each demanding relatively quick transactions. Grossman and Miller (1988) argue that the demand for immediacy in capital markets is both urgent and sustained.² This creates a role for an intermediary, or market maker, who supplies liquidity by standing ready to transact when order imbalances arise (Demsetz (1968)). The market maker's charge for providing immediacy services represents a potentially important component of transaction costs. There is a large literature exploring the nature of these costs, largely focusing on the market maker's marginal cost of holding inventory (see for example, Stoll (1978) and Ho and Stoll (1980)). Recent research shifts focus to the costs of adverse selection in market making, which arise when investors have access to information that is not yet reflected in the price.³ A common approach in the transaction cost literature is to make an auxiliary assumption of perfect competition in market making, so that the price of immediacy can be determined as the marginal cost of market making. In contrast, we assume imperfect competition in market making, allowing a single market maker to have exclusive rights to be continually present in the market for the security. The privileged position of the market maker, combined with the asynchronous arrival of immediacy demanding buyers and sellers, gives him some pricing power in setting transaction prices (or immediacy prices). The degree of pricing power is determined by the intensity of order arrivals, and collapses to zero, as in perfect competition, when arrival rates are infinite.

We develop a partial equilibrium model to study transaction costs from the perspective of an impatient individual seeking to transact Q units of a security. The individual can always trade via a standardized limit order, and when available, at the prices bid and asked by a market maker. In the spirit of individual portfolio choice problems, we assume that the fundamental value process and the arrival of other investors are unaffected by the individual's trading decisions.

A request to transact via limit order is essentially equivalent to writing an American option. For example, consider a seller placing a limit order. The seller can be viewed as offering the right to buy at a specific price at some point prior to an expiration date. This is effectively an American call option, requiring delivery of the underlying block of shares upon execution. Similarly, a request to buy is like an American put option. Unlike a typical option contract, the offer is

²Empirical evidence on order submission strategies generally supports this view (e.g. Bacidore, Battalio, and Jennings (2001); Werner (2003); He, Odders-White, and Ready (2006)).

³Bagehot (1971) was one of the first to consider the role of information in determining transaction costs in a capital market setting. Copeland and Galai (1983), Glosten and Milgrom (1985), and Kyle (1985) are important early models of the information component of transaction costs. See O'Hara (2004) for an overview of these models.

available to many counterparties, being extinguished as others transact against it or upon maturity. From the individual's perspective, the relevant potential counterparties are other individuals with liquidity demands as well as specialized liquidity providers, simply labeled here as market makers.⁴ Therefore, the market structure, including the existence and competitiveness of the market making function, determines the early exercise policy of an American-type limit order.

The notion that limit orders can be viewed as contingent claims is not new (see Copeland and Galai (1983) for a specific option-based model of prices bid and asked by a market maker; and Harris (2003) for general examples).⁵ The major innovation in this paper involves using the contingent claims approach to determine the prices of buy and sell transactions. To ensure immediate execution, the initiator of a transaction offer (the option writer) must offer a price at which it is currently optimal for the receiver of the transaction offer (the option owner) to exercise the option early. These strike prices, where immediate exercise is optimal, represent immediately transactable prices, and therefore are functionally equivalent to the prices bid and asked by the market maker.

We find that the option-based model of limit orders captures many important features of capital market transaction prices. The model points to the competitiveness of market making as a potentially important driver of transaction costs. In the case of a monopolist market maker, bid-ask spreads are increasing in the volatility of fundamental value, and in the level and persistence of order imbalances. Additionally, the model predicts that liquidity prices are nonlinear concave functions of transaction size. An attractive feature of the model is that the resulting formula for liquidity prices is a function of variables that can be estimated relatively easily. Calibration exercises demonstrate how the model can be used to estimate the quantity structure of transaction prices for individual securities, including very large transactions like corporate issues and takeovers. For example, using data through 2004, we estimate that a 1,000 share order in Philip Morris (Altria; MO) will cost 2bps; a 10,000 share order will cost 7bps, and immediately buying all shares outstanding will cost 22.6%.

The remainder of the paper is organized as follows. Section 2 describes the model. Section

⁴The function of providing immediacy is performed by whoever is currently available and paying attention. This may be an actual market maker, a dealer, a specialist, a "local" at the CBOT, day-trader, etc. We label all of these investor types, market makers.

⁵Longstaff (1995) uses an options-based framework to explore the cost of marketability at various horizons for an individual with perfect market timing.

3 discusses the properties of the quantity structure of immediacy prices. Section 4 proposes a method for implementing the model and suggests some empirical applications. Section 5 extends the basic model to allow two-states of market conditions via a regime switch, and introduces finite duration limit orders. Section 6 discusses liquidity in capital market transactions, and Section 7 concludes.

2 The Pricing of Limit Orders

A common feature of transaction offers across many markets is that they pre-specify price and quantity, and remain available for some potentially unknown amount of time. In financial markets, these offers are referred to as limit orders. So long as the value of the underlying asset can change over the life of the offer, viewing offers of this type as options is reasonable. The value of this option is naturally interpretable as a cost of transacting, since it represents the value foregone to obtain the desired execution terms. In particular, a limit order to sell (buy) Q shares at price K , gives arriving buyers (sellers) the right to purchase (sell) at a pre-specified limit price at some point prior to the expiration date of the limit order, and is therefore like an American call (put) option, with the Q -share block of the security acting as the underlying. By placing a limit order, the trade initiator can be viewed as surrendering an American call (put) option on the desired quantity of the underlying to the remaining market participants. Although the offer is potentially available to many counterparties, it is extinguished as soon as anyone exercises it or upon maturity. The option writer receives liquidity when the limit order is exercised. From the perspective of someone evaluating whether or not to exercise the option, the important considerations are their own liquidity demands and the potential for competition from other market participants.

The value of the limit order and its optimal exercise policy depend crucially on three factors: (1) the arrival rate of shares eligible for execution against the order; (2) the evolution of the fundamental value of the underlying security or basket of securities; and (3) the structure of the market. Because these factors are likely to have complex dynamics in reality, our model is best interpreted as a reduced-form characterization of transaction costs.

The challenge is to specify a suitable market structure that allows the demand and supply of immediacy to be isolated. Generally, each party to a trade is both demanding and supplying

immediacy to some extent. To simplify, we assume that the limit order writer (the trade initiator) is impatient and demands immediate execution. In order to have the limit order filled instantaneously, he must write an option that is sufficiently deep in-the-money to make immediate exercise optimal. Although the option is available to both the market maker and opposing order flow, only the market maker can be relied upon to supply immediacy at any given time because order flow arrives stochastically. For the market maker, the threat of losing the order to opposing order flow acts like a stochastic dividend on the underlying block of shares, creating an incentive for the market maker to exercise the option early.

An attractive feature of this setup is that limit prices for which immediate exercise is optimal represent instantaneously transactable prices, and therefore are functionally equivalent to the prices bid and asked by a market maker. This allows us to characterize the generally unobserved bid and ask prices for large quantities (i.e. larger than the quantity posted at the best bid and ask). Moreover, the option-based model of transaction prices inherits the properties of ordinary options. The two drivers of transaction costs for any given quantity are the fundamental volatility and the effective option maturity, which is determined by the order flow arrival rate. A quantity structure of instantaneously transactable prices arises because larger trade sizes require the trade initiator to write options with longer effective maturities.

2.1 A Reduced Form Model of Transaction Costs

To develop a model of transaction costs, we adopt a partial equilibrium framework similar in spirit to the one used for studying individual portfolio choice (Merton (1969, 1971)). In particular, we impose the following market structure and initial assumptions:

- A1.** Fundamental value is observed by all participants, and is the value at which the security would trade in a *competitive capital market*. It represents the outcome of the rational expectations equilibrium price formation process in a market with asymmetrically- (Wang (1993)) or heterogeneously-informed investors (He and Wang (1995)). The dynamics for fundamental value, V_t , can be described by a diffusion-type stochastic process:

$$\frac{dV_t}{V_t} = \alpha dt + \sigma dZ_t$$

where α is the instantaneous expected rate of return on fundamental value; σ^2 is the instantaneous variance of the return on fundamental value; and dZ_t is a standard Gauss-Wiener process.

- A2.** The price formation process giving rise to fundamental value, V_t , pins down the price of risk, γ_V , for exposure to the shocks dZ_t , and implies a pricing kernel of the form:

$$\frac{d\Lambda_t}{\Lambda_t} = -r dt - \gamma_V dZ_t \quad (1)$$

where r is the instantaneous riskless rate and $\gamma_V = \frac{\alpha-r}{\sigma}$. If markets are incomplete, this pricing kernel will not be the unique kernel of the economy, but it will be the unique kernel in the span of dZ_t , allowing us to price any asset whose value is exposed only to innovations in dZ_t .

- A3.** The market for a security is organized around a single market maker, who acts both as an *intermediary*, facilitating trades between third parties; and as an *agent*, filling outstanding orders from his inventory. The trading mechanism (game) facing a party interested in buying (selling) \tilde{Q} shares is as follows. In the first stage, the trader observes whether an imbalance exists; if there is a sell (buy) imbalance of q shares, that quantity of the trade is crossed at fundamental value. In the second stage, the trader submits a buy (sell) limit order to the market maker for the residual quantity, $Q = \tilde{Q} - q$, who can either fill the order, or include it in the next round of trading. If there is a sell (buy) imbalance of Q shares in the next round, the order is filled, and so on.

- A4.** The instantaneous probability of observing a buy (sell) imbalance of Q shares during the next instant is given by $\lambda^B(Q)dt$ ($\lambda^S(Q)dt$). Given this assumption, the expected time to the completion of a Q -share limit order to sell (buy) is distributed exponentially with mean $\frac{1}{\lambda^B(Q)} \left(\frac{1}{\lambda^S(Q)} \right)$.⁶

To preserve tractability and abstract from modeling the evolution of the limit order book, we focus on the special case in which all limit order traders have zero patience and only place

⁶The λ parameters can alternatively be interpreted as *search intensities* for eligible counterparties, in the spirit of Duffie, Garleanu and Pedersen (2005) or Vayanos and Wang (2002).

orders that are immediately exercisable by a profit maximizing market maker.⁷ In order to obtain immediacy, an impatient limit order trader must set the limit price, K , such that the option embedded in the order is sufficiently in-the-money to make immediate exercise optimal. In general, the schedule of limit prices guaranteeing immediacy will depend on the factors determining the value of the option: the riskless rate, r ; the volatility of the underlying, σ ; and the arrival rate of opposing order flow, $\lambda^i(\cdot)$, which itself is a function of the order quantity, Q . We will denote the schedules of immediacy prices for Q -share sell and buy limit orders by, $K_B(Q, \tau = 0)$ and $K_A(Q, \tau = 0)$, respectively, with the spreads between fundamental value and these prices having the interpretation of the *price of liquidity*.⁸ These schedules represent prices at which transactions can take place instantaneously and are functionally equivalent to immediacy prices.

Proposition 1 *The strike price at which it is optimal to immediately exercise a sell (buy) limit order for Q shares determines the effective bid (ask) price for Q shares.*

In our baseline specification we assume that limit orders are not subject to cancellation by the limit order writer. This implies that the limit order option is *perpetual*, albeit subject to a stochastic liquidating dividend (A4) in the form of order execution.⁹ The main virtues of the perpetual limit order feature are its analytical tractability and the fact that it provides an upper bound to immediacy costs. Since the value of the American option implicit in the limit order is monotonically increasing in time, a limit order writer forced to trade in perpetual limit orders is effectively surrendering options with the highest possible time value. Consequently, immediacy costs are maximized. In Section 5 we relax this assumption and consider limit orders subject to random cancellation by the limit order writer, as well as, finite duration limit orders. We find that the qualitative predictions of the model are unaltered.

The presence of the liquidating dividend (A4) is crucial in that it makes an early exercise strategy for the monopolist market maker optimal and facilitates the interpretation of option exercise as liquidity provision. The particular structure of the dividend process, controlled by a Poisson

⁷Grossman and Miller (1988) argue that there is high demand for immediacy in capital markets. Empirically evidence supports this view. Bacidore, Battalio and Jennings (2001) and Werner (2003) report that between 37-47% of all orders submitted on the NYSE are liquidity demanding orders, comprised of market orders or marketable limit orders.

⁸The investor's patience level, $\tau = 0$, is included to emphasize that immediacy is being demanded.

⁹It is important to keep in mind that observationally all limit orders will have finite lifetimes due to the presence of the liquidating dividend.

random variable with a quantity-dependent arrival intensity, is chosen for analytical tractability. In particular, the memoryless feature of the inter-arrival process preserves the time-stationary feature of the perpetual option valuation problem. This allows us to intuit that the optimal exercise boundary will be a barrier rule, which optimally trades off the preservation of the time-value of the option with the adverse consequences of the dividend.

This structure for the liquidating dividend implicitly assumes that limit orders are only subject to *one-shot execution*—there is no possibility for a limit order to be filled by a sequence of partial fills. The execution of the limit order is simply controlled by innovations to a Poisson counter. This has the added attraction that the mean inter-arrival time of opposing orders of size Q , $\frac{1}{\lambda^i(Q)}$, can be readily calibrated from empirical signed order flow data. Although the simplicity of this limit order execution process can obviously be challenged, relaxing this assumption can only be accomplished at the expense of analytical tractability. Numerical simulations show that the pricing of limit orders under a more sophisticated order flow process allowing for partial fills, yields results which are qualitatively indistinguishable from those obtained under the analytical model.¹⁰ Under the augmented specification used for the numerical simulation the random maturity of the finite-lived limit order option is determined by the joint dynamics of order imbalance, N_t , and fundamental value, V_t . These dynamics imply a time-varying instantaneous survival probability for the limit order and lead to a distribution of the times to completion that is not analytically tractable. In turn, it is not possible to obtain a closed-form expression for the value of the limit order option or its optimal early exercise rule, a feature which is shared by most American-type

¹⁰The numerical simulation replaces assumption (A4) with the following specification for the market order flow process:

A4a. Asynchronous investor arrivals into the market place can be described by a residual order flow process representing excess demand at fundamental value and is assumed to follow a simple mean-reverting process:

$$dN_t = \kappa(\theta - N_t)dt + \gamma dX_t$$

where N_t denotes the instantaneous arrival rate of buys (sells); θ is the equilibrium arrival rate for orders; κ is a constant governing the rate of mean reversion; γ^2 is the instantaneous variance of the order arrival rate; and dX_t is a standard Gauss-Weiner process (assumed to be uncorrelated with dZ_t).

A4b. Limit orders to sell (buy) are transacted against by positive excess demand if fundamental value is greater (less) than or equal to the limit price until the quantity remaining is zero or the limit order is canceled. The quantity outstanding for a sell limit order evolves according to the rule:

$$\begin{aligned} dQ_t &= -N_t^* dt \\ N_t^* &= \begin{cases} \min(Q_t, \max(0, N_t)) & \text{if } V_t \geq K \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

options.

2.2 Model Solution

Under assumptions A1-A4, the value of the Q -share limit order with a strike price K , $L(V_t, Q, K, t)$, can be shown to satisfy the following ordinary differential equation (ODE):

$$L_F \cdot (rF_{Q,t}) + \frac{1}{2}L_{FF} \cdot (\sigma F_{Q,t})^2 - (r + \lambda^i(Q)) \cdot L = 0$$

where subscripts are used to denote partial derivatives and $F_{Q,t} = Q \cdot V_t$ represents the fundamental value of the underlying block of shares. This ODE is solved subject to three boundary conditions. The first boundary condition is determined by the asymptotic behavior of the value of limit order as a function of $F_{Q,t}$, and the second pair of conditions arises from the value matching and smooth pasting at the optimal early exercise threshold. The equidimensional structure of the ODE suggests that the solution will be a linear combination of power functions in $F_{Q,t}$ with exponents given by:

$$\phi_{\pm}(\lambda^i) = \left(\frac{1}{2} - \frac{r}{\sigma^2} \right) \pm \sqrt{\left(\frac{1}{2} - \frac{r}{\sigma^2} \right)^2 + \frac{2(r + \lambda^i(Q))}{\sigma^2}}$$

Economic intuition allows us to exclude one of the two roots in both the case of a sell limit order and a buy limit order. In particular, since the value of a sell (buy) limit order is increasing (decreasing) in $F_{Q,t}$ we can exclude the negative (positive) root. Finally, to pin down the value of the constant of integration we make use of the fact that the optimal exercise rule for the option is a barrier rule. Consequently, the value of the limit order at optimal exercise is given by $Q \cdot (V_t^* - K)$ for a sell limit order and $Q \cdot (K - V_t^{**})$, where V_t^* and V_t^{**} are the optimal exercise thresholds for sell and buy limit orders, respectively. The expressions for the values of the limit orders and the associated optimal exercise thresholds are collected in the following proposition.

Proposition 2 *The value of a Q -share sell limit order is given by:*

$$L^S(V_t, Q, K, t) = \frac{QK}{\phi_+(\lambda^B) - 1} \cdot \left(\frac{\phi_+(\lambda^B) - 1}{\phi_+(\lambda^B)} \cdot \frac{V_t}{K} \right)^{\phi_+(\lambda^B)} \quad V_t < V_t^*$$

and it is optimal for the market maker to exercise the implicit call option whenever fundamental

value reaches the threshold $V_t^* = K \cdot \left(\frac{\phi_+(\lambda^B)}{\phi_+(\lambda^B)-1} \right)$ from below. The value of Q -share buy limit order is given by:

$$L^B(V_t, Q, K, t) = \frac{QK}{1 - \phi_-(\lambda^S)} \cdot \left(\frac{\phi_-(\lambda^S) - 1}{\phi_-(\lambda^S)} \cdot \frac{V_t}{K} \right)^{\phi_-(\lambda^S)} \quad V_t > V_t^{**}$$

and it is optimal for the market maker to exercise the implicit put option whenever fundamental value reaches the threshold $V_t^{**} = K \cdot \left(\frac{\phi_-(\lambda^S)}{\phi_-(\lambda^S)-1} \right)$ from above.

In order to induce immediate exercise of a sell (buy) limit order, the limit price (i.e. the option strike price) has to be set such that the prevailing fundamental value, V_t , is exactly equal to V_t^* (V_t^{**}), making it optimal for the market maker to exercise the order instantaneously. The analytical expressions for the immediacy prices depend on the order quantity, Q , through $\phi_+(\lambda^B)$ and $\phi_-(\lambda^S)$ and are summarized below.

Proposition 3 *The bid, $K_B(Q, \tau = 0)$, and ask, $K_A(Q, \tau = 0)$, prices are given by:*

$$K_B(Q, \tau = 0) = V_t \cdot \left(\frac{\phi_+(\lambda^B) - 1}{\phi_+(\lambda^B)} \right) \quad (2)$$

$$K_A(Q, \tau = 0) = V_t \cdot \left(\frac{\phi_-(\lambda^S) - 1}{\phi_-(\lambda^S)} \right) \quad (3)$$

and imply that the percentage immediacy costs for sales and purchases are given by:

$$\frac{K_B(Q, \tau = 0) - V_t}{V_t} = -\frac{1}{\phi_+(\lambda^B)} \quad (4)$$

$$\frac{K_A(Q, \tau = 0) - V_t}{V_t} = -\frac{1}{\phi_-(\lambda^S)} \quad (5)$$

2.3 Discussion

The reduced-form model in Section 2.1 is simple, but delivers predictions that appear to be reasonably consistent with real world transaction prices. The necessary ingredients for the nonlinear quantity structure of transaction prices are (1) the limit order has a fixed strike price and can remain outstanding for a sufficiently long period that fundamental value can change, allowing the limit order to be interpreted as an option; and (2) the market is structured such that the

market maker is a monopolist in the instantaneous supply of immediacy, competing with opposing order flow only when present. This accomplishes two things. First, the market maker is effectively granted ownership of the option. Second, the threat of loss to opposing order flow acts like a stochastic dividend, creating an incentive for the market maker to exercise the option early. In the interest of producing a simple closed-form solution, we consider a perpetual American option with a Poisson liquidating dividend. The specific form of the dividend and the perpetual nature of the option are chosen for mathematical tractability, and can be relaxed considerably, as shown in Section 5.

3 The Quantity Structure of Immediacy Prices

Immediacy is the limiting case, when patience goes to zero. The model imposes this condition to produce a quantity structure of instantaneously transactable prices—immediacy prices. At each point in time, the market maker stands ready to provide liquidity—the joint supply of quantity and immediacy. In the model, the two primary drivers of immediacy prices are the volatility of fundamental value and the time rate of arrivals of opposing order flow. Matching intuition, the model predicts that bid-ask spreads are increasing in fundamental volatility and that there are economies of scale in transactions.

Before it is possible to illustrate the above results graphically, it is necessary to pre-specify the dependence between quantity and the order arrival intensity, or equivalently, the order arrival time. Driven by empirical observations, we assume a simple proportional structure between $\frac{1}{\lambda^i(Q)}$ and Q , with the coefficient of proportionality corresponding to the expected arrival time of a single share. Using this auxiliary assumption, Figure 1 graphs the schedule of percentage immediacy prices, (4) and (5), as a function of order quantity. In particular, we assume the annual volatility of fundamental value is 15% or 35%, the riskfree rate of interest is 5% per year, and that orders arrive at a rate of one share per second.

Figure 1 shows that immediacy prices are nonlinear functions of the transaction size. Using the above definition of liquidity cost, these costs are increasing and concave in transaction size. This is in contrast to most information-based models of liquidity, which typically produce constant marginal costs, or linear price functions of quantity (for example, Kyle (1985)).

3.1 Effect of Order Flow Arrival Rates

Demsetz (1968) argues that it is reasonable to expect scale economies in transactions. As order flow arrival rates for a security increase, the waiting times for transaction execution in that security decrease. In the limiting case of infinite arrival rates, waiting times go to zero. In the more typical case of finite arrivals, the waiting time of a transaction can make up a significant portion of the total transaction cost. When investors demand immediacy, the waiting time can be transferred to the market maker (or marginal supplier of liquidity) who specializes in providing this service, but the waiting time cannot be eliminated.

The key friction in the model is that order flow arrivals are finite, which gives rise to a positive waiting time for transaction execution. In the model, there is a direct mapping of waiting times to option maturity. The time rate of arrivals of opposing order flow determines the expected waiting time of any given order, which in turn, determines the distribution of maturities of the options written by those demanding immediacy. Option values are generally increasing in their time-to-maturity, and this comparative static of option values flows through to the optimal immediate exercise strike price. In the limiting case, when arrivals are infinite, the waiting time is zero. The options written by the transaction initiator have zero time-to-maturity, and therefore have zero value. In this case, the bid-ask spread collapses and traders can buy and sell as much as they desire at fundamental value. However, the waiting time is generally positive, which gives positive option value to the provider of immediate transactions.

This intuition is formally captured in expressions (4) and (5). First and foremost, as the arrival rate of order flow eligible for execution against the outstanding limit order, $\lambda^i(\cdot)$, increases, the market maker faces more competition from order flow and the percentage immediacy costs decline. In the perfectly liquid market, $\lambda^i(\cdot) \rightarrow \infty$, the market maker possesses no pricing power and the costs of immediacy collapse to zero. Conversely, as competition from exogenous order flow declines, $\lambda^i(\cdot) \rightarrow 0$, the market becomes progressively more illiquid, allowing the monopolist market maker to charge a wider bid-ask spread to counterparties seeking immediacy. When trading by other market participants ceases altogether, $\lambda^i(\cdot) = 0$, the market maker is the sole provider of liquidity, and the asset market breaks down completely. The value of the sell limit order converges to the value of the underlying, V_t , implying that, in order to obtain immediacy, the seller must part

with the asset at a zero price. Intuitively, in this scenario, the market is a pure monopoly in which the market maker captures the entire surplus. On the other hand, buy transactions still remain possible, but only at significant premia to fundamental value. In the limiting case when $\lambda^i(\cdot) = 0$, the smallest percentage premium to fundamental value guaranteeing immediate execution is given by $\frac{\sigma^2}{2r}$.

Figure 2 displays the immediacy prices for fixed transaction sizes as a function of the expected arrival time of one share, or equivalently, the order flow arrival rate. In general, immediacy prices do not equal fundamental value. As order flow arrival rates increase, expected waiting times shrink, and the bid and ask prices converge towards fundamental value. The increase in efficiency is largest when arrival rates begin low and increase. The figure shows a changing rate of convergence in immediacy prices towards fundamental value—initially very fast at low arrival rates, then becoming more gradual as arrival rates increase.

3.2 Effect of Fundamental Volatility

In our model, immediacy prices offered by the market maker deviate farther from fundamental value as the volatility of fundamental value increases, for any given quantity (an illustration is presented in Figure 1). This is a direct consequence of the option-based approach. Option values are increasing in volatility, and this property flows through to the strike price at which immediate exercise is optimal. The more valuable the option's value, the larger the distance must be between the strike price and fundamental value for the market maker to exercise immediately. In particular, in the limit as $\sigma \rightarrow \infty$, the value of a Q -share sell limit order with a limit price of K approaches Q times the fundamental value. A similar buy limit order approaches Q times the limit price. Because immediate exercise requires that the limit order writer give the market maker an option that is in-the-money, the percentage immediacy cost for sell orders goes to 100%. Buy limit orders, on the other hand, are never executed. Conversely, in the absence of any price risk, i.e. when the volatility of fundamental value is zero ($\sigma = 0$), the options implicit in the order flow have no value, so no premium is required to induce the market maker to exercise immediately.

3.3 Liquidity Events

The analytical model presented in Section 2 allows us to examine how shocks to the arrival rate of buy/sell orders and the fundamental value of the underlying may compound during a liquidity crisis to affect immediacy prices. The arrival rate of buy (sell) orders will determine the expected maturity of the options written by a seller (buyer) demanding immediacy. Therefore, from the seller's (buyer's) perspective, a liquidity crisis is likely to involve a significant decrease in the current time rate of buy (sell) order arrivals, relative to the equilibrium rate. This asymmetry in arrival rates may become more severe if the current time rate of sell order arrivals also increases. This captures the notion that a liquidity crisis involves some sort of order imbalance. As a consequence of a temporary order imbalance a significant asymmetry in buy and sell immediacy prices may emerge at all quantities, causing the midpoint of the bid-ask spread to become a biased estimator of the fundamental value.

Figure 3 displays the effects of an order imbalance on the quantity structure of immediacy prices. In particular, the figure assumes that the current rate of sell order arrivals increases fivefold, while the current rate of buy order arrivals falls by this factor. This represents a major "running for the exit" in the security. Immediacy prices for buyers become much more elastic, such that an investor wishing to buy can now immediately transact very large quantities at a price much closer to fundamental value. However, investors wishing to sell immediately must pay a large premium, even for relatively small quantities. In other words, the immediacy prices facing sellers are now less elastic at all quantities.

Figure 3 also displays immediacy prices in the case when an order imbalance coincides with an increase in fundamental volatility. The increased volatility offsets the reduced waiting time for buy orders, attenuating the increased elasticity of immediacy ask prices slightly. On the other hand, the higher volatility further increases the premium for immediacy for sellers, making prices even less elastic at all quantities.

4 Empirical Calibration

In this section we propose some empirical calibration schemes for implementing the model introduced in Section 2, and examine the resulting model-based predictions for immediacy prices. Calibrating our model requires the estimation of three parameters: the riskless rate, r ; the volatility of fundamental value, σ ; and the arrival rate of price insensitive order flow, $\lambda^i(Q)$. Obtaining values for the first two parameters is simple and only requires the use of daily data. An estimate for the riskless rate can be readily obtained from data on the 30-day Treasury bill, available from CRSP and the Federal Reserve. Similarly, the volatility of fundamental value, σ , can be obtained by using the volatility of daily returns, available from CRSP. Estimation of the quantity-dependent order arrival rates, $\lambda^B(Q)$ and $\lambda^S(Q)$, is only slightly more demanding in that it requires transaction level (Trade and Quote, TAQ) data.

4.1 Estimating $\lambda^i(Q)$

The estimation of the order arrival rates and their quantity dependence proceeds in two steps. In the first step, the transaction level data is signed into buyer- and seller-initiated trades. We accomplish this by utilizing the Lee and Ready (1991) tick-signing algorithm. Under this tick-signing scheme, trades occurring at prices above (below) the prevailing midquote are considered to be buyer-initiated (seller-initiated). In the second step, we use the time-series of the the signed order flow to obtain an estimate of the order arrival rates. Here, we exploit the fact that the order arrival intensity, $\lambda^i(Q)$, is simply the inverse of the mean interarrival time between Q -share increments in the signed order flow. In turn, examining the mean interarrival times for Q -share blocks of buys and sells, as a function of Q , provides a non-parametric estimate of $\frac{1}{\lambda^i(Q)}$, which can be inverted to obtain $\lambda^i(Q)$.

Casual examination of the functional relationship between the mean interarrival times and demanded quantity indicates that a proportional relationship is reasonable. This form of relation is expected whenever the arrival rate for buy (sell) *volume* is constant.¹¹ Mathematically, the

¹¹This result suggests that the measured arrival rates of buy and sell orders are constant after re-scaling business time by signed volume. It does *not* indicate that the arrival rates of individual buy and sell transactions are constant.

observed relation indicates that:

$$\frac{1}{\lambda^i(Q)} \approx \frac{1}{\lambda^i(1)} \cdot Q$$

The coefficient of proportionality, $\frac{1}{\lambda^i(1)}$, is simply the mean waiting time for the arrival of one share of buy (sell) volume.¹² It can be estimated through two simple methods. Under the naïve method, one simply takes the total realized buy (sell) volume and divides it by the total amount of time that has elapsed (e.g. in seconds), to obtain a (e.g. *per second*) arrival rate of buy (sell) volume. Alternatively, one estimates an OLS regression of the sample mean interarrival times of Q -share blocks of signed volume as a function of Q . Figure 4 plots the in-sample relation between the mean inter-arrival times as a function of quantity, Q , for Altria (ticker: MO) using signed order flow data from 2004. The left two panels present the buy arrival rates, and the right two panels display the sell arrival rates; the top panels are plotted in the natural scale, while the bottom panels use log scales. The plot confirms that expected waiting times are proportional to size and that this relation holds over a wide range of order magnitudes. Moreover, the fitted relation obtained using the OLS estimated slope parameters and the naïve parameter estimate, obtained by simply dividing total order flow by total elapsed time, are quite similar. In the case of Altria the OLS implied *per second* order arrival rates are $\hat{\lambda}^B(1) = 105.45$ and $\hat{\lambda}^S(1) = 101.70$, whereas their naïve counterparts are $\hat{\lambda}^B(1) = 105.65$ and $\hat{\lambda}^S(1) = 96.35$. Because these parameter estimates are relatively similar, we use the naïve estimator in what follows.

4.2 Estimated Quantity Structures of Immediacy Prices

Figure 5 plots the calibrated quantity structures of immediacy prices for four firms: Altria Group (ticker: MO), Corning (ticker: GLN), Harrahs Entertainment (ticker: HET) and Hecla Mining (ticker: HL). These firms represent a varied sample in terms of the estimated fundamental volatility and order arrival rates. The wide range of underlying parameters is reflected in the resulting *price of liquidity* for the various firms. For example, demanding immediacy on a \$10,000 transaction in Altria requires a 1 basis point concession relative to fundamental value, while the same transaction in Hecla Mining requires a 16 basis point concession. For a \$1,000,000 transaction, the immediacy costs grow to 9 basis points and 159 basis points, respectively.

¹²The theoretically predicted intercept of the relationship is zero since the waiting time for zero shares should trivially be equal to zero.

The calibrated model also generates reasonable predictions for more extreme capital market transactions. For example, a cash tender offer can be thought of as a demand for the instantaneous acquisition of all outstanding shares of the target company. The takeover premia produced by our model (24% for Altria, 56% for Corning, 18% for Harrahs and 45% for Hecla Mining) are in the ballpark with those reported in the merger literature.¹³ In a similar spirit, our model also allows one to evaluate the liquidity costs of seasoned equity offers. Here, the issuing firm is the liquidity demanding party and must offer a price concession relative to fundamental value to induce market participants to absorb the new shares.

In order to gain a sense for the cross-sectional dispersion in the underlying model parameters and the model implied transaction costs, we compute these values for all NYSE stocks using data for 2004. Their mean values, by size deciles, are collected in Table 1. There is strong relation between our liquidity measures and firm size. The smallest firms have relatively low order flow arrival rates and high volatility, producing large estimated transaction costs. Specifically, the model estimates the average transaction cost associated with a \$100,000 trade in a stock from the smallest size decile is 1.34%. The estimate is 0.04% for the same trade in a stock from the largest size decile.

An alternative approach to using our model is to use it as a device for inferring one of the underlying parameters, in the same spirit that the Black-Scholes formula is used to “back out” implied volatility. Typically the parameter of interest will be an order arrival rate for which there may not be enough empirical data for a direct calibration, as may be the case for thinly traded stocks or bonds. After backing out the implied order arrival rate, the analytical structure of the model can be used to infer the rest of the quantity structure of immediacy prices. To do this one simply matches the model predicted immediacy cost to a particular data point for which one has an ample number of observations, e.g. empirical data on the price impact of the most frequently observed order quantity. Fixing the fundamental volatility from the daily return series, one can solve for the implied order arrival rate and use that information to generate the entire, unobserved quantity structure of immediacy prices.

¹³ Andrade, Mitchell, and Stafford (2001) report that the median merger premium from 1973 to 1998 is 38%.

5 Extensions

We consider two extensions to the model presented in Section 2. The first, augments the option valuation model with regime switching in the volatility of the underlying and the arrival rates of opposing order flow. The regime-switching feature allows us to formally examine the effect of changing market conditions on immediacy prices, and to incorporate the non-normality of high-frequency return data. The second extension relaxes the perpetual option feature of the limit order, thereby allowing agents to submit limit orders that are subject to random cancellation. We show that a model with Poisson order cancellation is isomorphic to the baseline model from Section 2, with the order arrival intensity, $\lambda^i(Q)$, augmented by the limit order cancellation intensity, η . Lastly, we examine limit orders with an Erlang distributed terminal date. Using an asymptotic argument in the spirit of Carr (1998), we derive recursive expressions for the value of a limit order whose maturity date has a degenerate probability distribution with all mass concentrated at a single future date, T .

5.1 Regime Switching in Liquidity and Volatility

Short-lived, but extreme, changes in market conditions have the potential to significantly affect the cost of liquidity, as is illustrated in Section 3.3. In anticipation of these dislocations, market makers are likely to rationally adjust the premia (discounts) relative to fundamental value at which they are willing to provide immediacy. To examine the magnitude of these effects we embed the model from Section 2 in a continuous-time Markov chain framework with two states: a high liquidity state (state h) and a low-liquidity state (state l). State h can be interpreted as corresponding to the “normal” market conditions, and state l to a market experiencing a liquidity crisis (e.g. a flight to quality). The transitions between the two states are governed by a Markov chain with infinitesimal generator, Ξ :

$$\Xi = \begin{bmatrix} -q_l & q_l \\ q_h & -q_h \end{bmatrix}$$

Under this regime-switching model, the market is expected to stay in state i for an exponentially distributed interval of time, with mean duration $\frac{1}{q_i}$. Given our assumption that state h corresponds

to the “normal” market conditions, and state l to the comparatively rare, and short-lived, liquidity crises, we expect q_h (q_l) to be low (high). To retain the greatest generality, we allow the volatility of fundamental value to be state dependent, and equal to σ_h and σ_l , in the high and low liquidity states, respectively. We set $\sigma_h < \sigma_l$, to reflect the potentially greater uncertainty about fundamentals during market crises. In what follows, we focus on the case of the sell limit order, relegating the symmetric case of a buy limit order to Appendix B.

In the presence of regime-switching, the value of a sell-limit order will be state dependent, and given by $L^{h,S}$ ($L^{l,S}$) in the high-liquidity (low-liquidity) state. While the sell limit order retains the features of a call option, as in Section 2, its underlying is now subject to regime-switching in volatility and the arrival rate of the liquidating dividend. Because the joint evolution of fundamental value, V_t , and the liquidity state, $i \in \{h, l\}$, is Markov, it is possible to show that the optimal exercise rule for this option will be a state-dependent barrier rule.¹⁴ In state l , the intensity of the liquidating dividend is lower than in state h (longer effective life of the limit order option) and the volatility of fundamental volatility is higher. Both of these features increase the value of the limit order option, and therefore the required inducement to illicit early exercise. Consequently, the optimal exercise threshold in state l , V_l^* , will exceed the optimal exercise threshold in state h , V_h^* .

The two early exercise barriers effectively partition the state space into three regions: (1) a region in which it is not optimal to fill the limit order independent of the prevailing liquidity state ($0 < V_t < V_h^*$); (2) a region in which early exercise is only optimal in the high-liquidity state ($V_h^* \leq V_t < V_l^*$); and (3) a region in which it is optimal to exercise the option independent of the prevailing liquidity state ($V_l^* \leq V_t$). The value of the sell limit order in the third region is simply $Q \cdot (V_t - K)$. The value functions in the first two states are given by the solution to the following system of ODEs. When $V_t < V_h^*$ the state-dependent limit order value satisfies the following system of equations:

$$\begin{aligned} L_F^{(h,S)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(h,S)} \cdot (\sigma_h F_{Q,t})^2 - rL^{(h,S)} &= \lambda_h^B(Q) \cdot (L^{(h,S)} - 0) + q_h \cdot (L^{(h,S)} - L^{(l,S)}) \\ L_F^{(l,S)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(l,S)} \cdot (\sigma_l F_{Q,t})^2 - rL^{(l,S)} &= \lambda_l^B(Q) \cdot (L^{(l,S)} - 0) + q_l \cdot (L^{(l,S)} - L^{(h,S)}) \end{aligned}$$

¹⁴A formal proof of the optimality of this exercise rule is provided by Guo and Zhang (2004).

and when $V_t \in [V_h^*, V_l^*]$ the value functions satisfy:

$$\begin{aligned} L_F^{(l,S)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(l,S)} \cdot (\sigma_l F_{Q,t})^2 - rL^{(l,S)} &= \lambda_l^B(Q) \cdot (L^{(l,S)} - 0) + q_l \cdot (L^{(l,S)} - Q \cdot (V_t - K)) \\ L^{(h,S)} &= Q \cdot (V_t - K) \end{aligned}$$

Appendix B shows that this system of ODEs can be solved in closed-form, yielding our next result.

Proposition 4 *The value of a sell limit order subject to regime-switching in the volatility of innovations to fundamental value and the arrival rate of the stochastic liquidating dividend is given by the following, state-dependent value function:*

$$L^{(l,S)} = \begin{cases} A_3 F_{Q,t}^{\beta_3} + A_4 F_{Q,t}^{\beta_4} & V_t < V_h^* \\ C_1 F_{Q,t}^{\phi_+ (\lambda_l^B)} + C_2 F_{Q,t}^{\phi_- (\lambda_l^B)} + \nu_l(F_{Q,t}) & V_h^* \leq V_t < V_l^* \\ Q \cdot (V_t - K) & V_t \geq V_l^* \end{cases} \quad (6)$$

$$L^{(h,S)} = \begin{cases} B_3 F_{Q,t}^{\beta_3} + B_4 F_{Q,t}^{\beta_4} & V_t < V_h^* \\ Q \cdot (V_t - K) & V_t \geq V_h^* \end{cases} \quad (7)$$

with the values of $(A_3, A_4, C_1, C_2, V_h^*, V_l^*)$ given by the solution to a system of algebraic equations provided in Appendix B. The value of a buy limit order has a symmetric structure and its derivation is relegated to Appendix B.

An added benefit of the regime-switching framework is that it allows us to examine the impact of fat tails in the distribution of the underlying on the pricing and optimal exercise rules of the limit orders. Although the evolution of fundamental value follows a geometric Brownian motion conditional on a given state of the Markov chain, the *unconditional* distribution of returns is generated by a mixture of normals. This distribution exhibits excess kurtosis, capturing a salient feature of returns measured at high-frequencies. Consequently, the limit order values and optimal exercise rules obtained from the regime-switching model can also be interpreted as measuring the effects of non-normalities in the evolution of fundamental value on the quantity structure of immediacy prices.

5.2 Finite Maturity Limit Orders

The baseline model of Section 2 is solved under the assumption that limit order writers never cancel a submitted order. This feature allows us to treat the limit order option as a perpetual option, subject to a stochastic liquidating dividend in the form of execution by the opposing order flow. However, since this assumption is clearly open to challenge, we devote this section to showing that it can be relaxed considerably without any effect on the qualitative features of the quantity structure of immediacy prices.

We begin our analysis by considering limit orders that are subject to random cancellation by the limit order writer, and then turn our attention to limit orders with a finite maturity date. Although limit orders are *de facto* unlikely to be canceled at randomly selected times, random cancellation will be observationally indistinguishable from a deterministic cancellation rule, so long as the market maker cannot infer this rule. We therefore assume that a limit order is canceled by the order writer at the N -th arrival time of a Poisson process with intensity η . Under this auxiliary assumption, the order cancellation time, τ , will have an Erlang distribution with:

$$Pr\{\tau \in dt\} = \frac{\eta^N}{(N-1)!} t^{N-1} e^{-\eta t} dt$$

and the expectation and variance of the cancellation time, τ , will be given by:

$$E[\tau] = \frac{N}{\eta} \quad Var[\tau] = \frac{N}{\eta^2}$$

In the base case, when $N = 1$, the Erlang distribution collapses to an exponential distribution. In this case it is easy to show that the value of the buy and sell limit orders continues to be given by the expressions provided in Section 2, but with slightly modified cancellation intensities.

Proposition 5 *The value of a sell (buy) limit order that is subject to cancellation by the limit order writer at the first jump time of a Poisson process with intensity η , is equivalent to the value of a limit order that is not subject to cancellation, but is subject to a stochastic liquidating dividend*

arriving at rate $\tilde{\lambda}^B(Q)$ ($\tilde{\lambda}^S(Q)$), is given by:

$$\begin{aligned}\tilde{\lambda}^S(Q) &= \lambda^S(Q) + \eta \\ \tilde{\lambda}^B(Q) &= \lambda^B(Q) + \eta\end{aligned}$$

A formal proof of this result can be found in Appendix C.

The simple isomorphism between limit orders that are not subject to cancellation by the limit order writer (i.e. perpetual limit orders) and those that are, shows that the qualitative features of the quantity structure of transaction prices will be unaffected by the introduction of the cancellation feature.

To establish that our results continue to hold in the case of finite duration limit orders (i.e. orders that will be cancelled at a future date T), we exploit the randomization device of Carr (1998). This mathematical device takes advantage of the scaling of the moments of an Erlang distributed random variable in the Poisson arrival intensity, η , to synthesize a random variable with a pre-specified mean and zero variance. To see this, suppose we let $\eta = \frac{N}{T}$, and allow $N \rightarrow \infty$. Asymptotically, the moments of the limit order cancellation time, τ , collapse to $E[\tau] \rightarrow T$ and $Var[\tau] \rightarrow 0$. In other words, the limit order is canceled at time T with unit probability.

To determine the value of a limit order subject to cancellation at time T , it is therefore sufficient to determine the value of the limit order subject to Erlang cancellation when $\eta = \frac{N}{T}$, and $N \rightarrow \infty$. Under the Erlang cancellation scheme, the value of a limit order will depend on the fundamental value of the underlying, $F_{Q,t}$; the arrival intensity of the opposing order flow, $\lambda^i(Q)$; and the number of periods left to the termination of the option, n . Given these assumptions, the value of the limit order will be given by the solution to the following system of N ($n = N \dots 1$) ordinary differential equations:

$$L_F^{(n)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(n)} \cdot (\sigma F_{Q,t})^2 - r \cdot L^{(n)} = \lambda^i(Q) \cdot (L^{(n)} - 0) + \eta \cdot (L^{(n)} - L^{(n-1)})$$

with $L^{(0)} = 0$ (i.e. the limit order becomes worthless upon cancellation). The terms on the left hand side of the equality represent the evolution of the limit order value in the absence of jumps, while the terms on the right hand side represent the probability weighted losses from order exercise by

oncoming order flow and the passage of time, as measured by the jumps in the $Poisson(\eta)$ variable. To solve this system of ODEs we proceed by backwards recursion, starting with state $n = 1$. The solution is comprised of a sequence of state-dependent value functions and the associated optimal early exercise thresholds. A characterization of the complete, recursive solution is given in the following proposition.

Proposition 6 *The value of a sell limit order in state n is given by:*

$$L^{(n,S)} = \begin{cases} \alpha_{0,n} \cdot F_{Q,t}^{\phi_+(\lambda^B)} + \alpha_{1,n} \cdot F_{Q,t}^{\phi_-(\lambda^B)} + \left(\frac{\eta}{\eta + \lambda^B(Q)}\right) \cdot F_{Q,t} - \left(\frac{\eta}{r + \eta + \lambda^B(Q)}\right) \cdot QK & V_t \geq V_{n-1}^* \\ \beta_{0,n} \cdot F_{Q,t}^{\phi_+(\lambda^B)} + L_p^{(n,S)}(V_t < V_{n-1}^*) & V_t < V_{n-1}^* \end{cases} \quad (8)$$

where V_n^* denotes the optimal early exercise threshold for state n and $L_p^{(n,S)}(V_t < V_{n-1}^*)$ is an analytical expression related to the value function, $L^{(n-1,S)}$, in the continuation region for state $n - 1$. The values for $(\alpha_{0,n}, \alpha_{1,n}, \beta_{0,n}, V_n^*)$ can be determined by solving a system of equations provided in the Appendix. The corresponding solution for a buy limit order can be found in Appendix C.

It is possible to show that the sequence of optimal exercise thresholds for a sell limit order, V_n^* , is increasing in n , reflecting the increasing time-value of the limit order option. Despite the complexity of the full solution, the form of the value function characterizing the limit order option in an arbitrary state, n , is closely related to the solution from Section 2. The recursive analytical solution, combined with numerical solution of the system of equations parameterizing the coefficients of the value function and optimal exercise threshold, confirms that the quantity structure of transaction prices (now indexed by state n) retains all the qualitative features examined in Section 3.

6 Liquidity in Capital Market Transactions

Real world markets are less than perfectly liquid, meaning that there is typically a gap between fundamental value and transaction prices. The price of liquidity, $p(Q)$, measures this gap. There are many potential reasons for a positive price of liquidity, and a large literature investigates how this price is determined in equilibrium. Recognizing the unique role of the market

maker in capital markets, much research focuses on the costs of market making. These costs arise from market makers' sub-optimal portfolio holdings and from the adverse selection they face when trading against informed investors. With a characterization of the market maker's cost function and the assumption of perfect competition, transaction costs can be defined: $p(Q) = mc(Q)$.

We approach the pricing of liquidity from a different perspective. Starting with the observation that limit orders are options, we focus on the valuation of a limit order, $L(Q, K)$. Since limit orders are common means of trade, their values represent the cost of transacting. Moreover, the limit price that induces a profit maximizing agent to immediately exercise early, $K^*(Q)$, is functionally equivalent to the prices bid and asked by a market maker—immediacy price for Q -units. Therefore, the price of liquidity can be defined from these values, $p(Q) = L(Q, K^*) = V_t - K^*(Q)$.

The assumption of a monopolistic market maker plays an important role in our model. This effectively grants ownership of the limit order to the market maker. Competition for the limit order arrives stochastically from opposing order flow, which from the market maker's perspective, acts like a dividend on the underlying block of shares. Our market maker has a monopoly in the right to "hang around", while other market participants must take an action and move on. Consequently, there is competition in the supply of liquidity through time, but instantaneous competition cannot be guaranteed at all times. In the case of order flow imbalances, the market maker enjoys pricing power. However, this pricing power diminishes when frequent order flow arrivals discipline the market maker. As arrival rates increase, transaction prices converge to fundamental value, and the market is essentially competitive.

The introduction of a competitive market making function alters the pricing of a limit order through the early exercise rule. An individual placing a limit order in this market structure expects their limit order to be exercised by opposing order flow, as before, but now also by the market maker any time the intrinsic value of the option exceeds the marginal cost of the market maker's adjustment to inventory, modifying the early exercise boundary, $V_t - K^*(Q) = mc(Q)$. To be useful, this requires a characterization of the market maker's cost function. On the other hand, if the market maker is a monopolist, we can determine the price of liquidity through the optimal exercise policy of the limit order, with no explicit knowledge of the market maker's cost function.

Two simple implications of our basic setup are that market makers are long volatility and

that they earn profits in the presence of order imbalances. This seems to fit with common intuition and empirical evidence on supplying liquidity. For example, in the price pressure hypothesis proposed by Scholes (1972), prices can temporarily diverge from their information-efficient values because of uninformed shifts in excess demand to compensate those that provide liquidity. Mechanically, this occurs when prices return to their information-efficient values, presumably over a short horizon. This should not occur with perfectly competitive market making in the absence of imperfect information. Our model captures this notion of price pressure through imperfect competition. A market maker is able to extract more rents from a large order imbalance when the primary source of competition comes from investors with liquidity demands (i.e. the residual order flow process itself) rather than other market makers. The order imbalance represents a weakening of competition for the monopoly market maker, allowing him to extract larger rents. In other words, the investors with common liquidity demands are forced to write options with longer effective maturities (i.e. more valuable options) when order imbalances grow and/or become somewhat persistent. A consequence of this type of price pressure is that supplying immediacy in these situations is profitable. Coval and Stafford (2006) present evidence consistent with impatient order flow imbalances of either direction significantly affecting transaction prices for relatively long periods, but eventually reversing once investors' demand for immediacy subsides. In particular, the immediacy arises from extreme capital flows to mutual funds, which can lead to order imbalances when many specialized funds, facing concurrent capital flows, have highly overlapping portfolios. Those who provide liquidity in these periods appear to earn significantly positive abnormal returns, consistent with the notion that competition in the provision of immediacy in these situations is less than perfectly competitive.

Another implication of the option-based framework is that idiosyncratic risk may play an important role in equilibrium asset pricing. In the model, transaction prices are generally not equal to fundamental value, but instead are the optimal exercise prices of call and put options. The optimal exercise prices of these options share the comparative static properties of option values, with respect to volatility. Option values are increasing in total volatility, with no distinction made between systematic or non-systematic risk. As such, the model predicts that the transaction prices we observe in real world markets will be influenced by non-systematic risk.

In a recent paper, Spiegel and Wang (2005) regress various liquidity measures on idiosyncratic risk (essentially residual variance from FF 3-factor model) and find a strong negative relation. Moreover, they find that in a cross-sectional analysis of returns, controlling for both idiosyncratic risk and various liquidity measures, that only the idiosyncratic risk has significant explanatory power. Our model produces a new functional relation between volatility and liquidity that may help explain this result.

7 Conclusion

This paper develops a model of instantaneously transactable prices as a function of transaction quantity through the pricing of limit orders. The quantity structure of immediacy prices represents the effective bid/ask price for various transaction sizes, and is derived as the limit prices that induce immediate early exercise of American-type limit orders. The market maker's early exercise policy is dependent on the competitiveness of the market making function. When competition in market making is imperfect, the option-based model of immediacy prices identifies the level and persistence of order imbalances and the volatility of the security as the key drivers of transaction costs.

Interestingly, our model with full-information, but imperfect market making, is able to qualitatively reproduce many of the properties of transaction costs that emerge from traditional models of costly market making. An attractive feature of the option-based framework is that the model can be estimated as a function of relatively observable variables. We propose a method for implementing the model using forecasts of volatility and order flow. Additionally, the predictions of the model can be tested with these data.

Finally, the model may also be a useful step towards a new measure of *liquidity risk*. The uncertainty over transactable prices, relative to fundamental value, produces a liquidity risk. As such, the time series variation of the price of liquidity is a natural measure of liquidity risk. This suggests extending the baseline model to incorporate time-varying arrivals and exploring the time dynamics of the resulting quantity structure of immediacy prices.

References

- [1] Admati, A. and P. Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies* 1, 3-40.
- [2] Admati, A. and P. Pfleiderer, 1989, Divide and conquer: A theory of intraday and day-of-the-week mean effects, *Review of Financial Studies* 2, 189-224.
- [3] Amihud, Y. and H. Mendelson, 1980, Dealership market: Market making with inventory, *Journal of Financial Economics* 8, 31-53.
- [4] Bacidore, J., R. Battalio, R. Jennings, 2003, Order submission strategies, liquidity supply, and trading in pennies on the New York Stock Exchange, *Journal of Financial Markets* 6, 337-362.
- [5] Bagehot, W., 1971, The only game in town, *Financial Analysts Journal* 22, 12-14.
- [6] Black, Fischer and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 617-654.
- [7] Brunnermeier, Markus K. and Lasse H. Pedersen, 2006, Market liquidity and funding liquidity, Princeton working paper.
- [8] Carr, Peter, 1998, Randomization and the American Put, *Review of Financial Studies* 11(3), 597-626.
- [9] Copeland, Thomas E., and Dan Galai, 1983, Information effects and the bid-ask spread, *Journal of Finance* 38, 1457-1469.
- [10] Coval, Joshua and Erik Stafford, 2006, Asset fire sales (and purchases) in equity markets, Harvard Business School working paper.
- [11] Demsetz, Harold, 1968, The cost of transacting, *Quarterly Journal of Economics* (82), 33-53.
- [12] Duffie, Darrell, Nicolae Gârleanu, and Lasse Pedersen, 2005, Over-the-counter markets, *Econometrica* 73, 1815-1847.
- [13] Easley, D. and M. O'Hara, 1987, Price, Trade Size, and Information in Securities Markets, *Journal of Financial Economics* 19, 69-90.

- [14] Garman, M., 1976, Market microstructure, *Journal of Financial Economics* 3, 257-275.
- [15] Glosten, L. and P. Milgrom, 1985, Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 13, 71-100.
- [16] Glosten, Lawrence R., 1989, Insider trading, liquidity, and the role of the monopolist specialist, *Journal of Business* 62, 211-235.
- [17] Grossman, Sanford J., and Merton H. Miller, 1988, Liquidity and market structure, *Journal of Finance* 43, 617-633.
- [18] Guo, X., and Zhang, Q., 2004, Closed-form solutions for perpetual American put options with regime switching, *SIAM Journal of Applied Mathematics* 64(6), 2034-2049.
- [19] Harris, Larry, 2003, *Trading and Exchanges: Market Microstructure for Practitioners*, Oxford University Press, New York.
- [20] Hasbrouck, Joel, 2005, Trading costs and returns for US equities: The evidence from daily data, NYU, Stern School of Business working paper.
- [21] He, Chen, Elizabeth R. Odders-White, and Mark J. Ready, 2006, The impact of preferencing on execution quality, *Journal of Financial Markets* 9, 246-273.
- [22] He, Hua, Jiang Wang, 1995, Differential information and dynamic behavior of stock trading volume, *Review of Financial Studies* 8, 919-972.
- [23] Ho, Thomas and Hans R. Stoll, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics* 9, 47-73.
- [24] Lee, Charles M. C. and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733-746.
- [25] Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1335.
- [26] Longstaff, Francis A., 1995, How much can marketability affect security values, *Journal of Finance* 50, 1767-1774.

- [27] Longstaff, Francis A., Eduardo S. Schwartz, 2001, Valuing American options by simulation: a simple least squares approach, *Review of Financial Studies* 14, 113-147.
- [28] Merton, Robert C., 1969, Lifetime portfolio selection under uncertainty: The continuous time case, *Review of Economics and Statistics* 51, 247-257.
- [29] Merton, Robert C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373-413.
- [30] Merton, Robert C., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science*, 141-183.
- [31] Merton, Robert C., 1977, On the pricing of contingent claims and the Modigliani-Miller theorem, *Journal of Financial Economics* 5, 241-249.
- [32] Merton, Robert C., 1989, On the application of the continuous-time theory of finance to financial intermediation and insurance, *The Geneva Papers on Risk and Insurance* 14, 225-262. (Reproduced as Chapter 14 in *Continuous-Time Finance*.)
- [33] O'Hara, Maureen, 2004, *Market Microstructure Theory*, Blackwell Publishing, Oxford, UK.
- [34] O'Hara, M. and G. Oldfield, 1986, The microeconomics of market making, *Journal of Financial and Quantitative Analysis* 21, 361-376.
- [35] Scholes, Myron, 1972, The market for corporate securities: Substitution versus price pressure and the effects of information on stock prices, *Journal of Business* 45, 179-211.
- [36] Seppi, D., 1990, Equilibrium block trading and asymmetric information, *Journal of Finance* 45, 73-94.
- [37] Spiegel, Matthew and Xiaotong Wang, 2005, Cross-sectional variation in stock returns: Liquidity and idiosyncratic risk, Yale School of Management working paper.
- [38] Stoll, Hans R., 1978, The supply of dealer services in securities markets, *Journal of Finance* 33, 1133-1151.

- [39] Vayanos, D, and T. Wang, 2002, Search and endogenous concentration of liquidity in asset markets, UBC working paper.
- [40] Wang, Jiang, 1993, A model of intertemporal asset prices under asymmetric information, *Review of Economic Studies* 60, 249-282.
- [41] Werner, Ingrid, 2003, NYSE order flow, spreads, and information, *Journal of Financial Markets* 6, 309-335.

A Pricing Limit Orders

We consider the pricing of limit orders under the following assumptions:

1. **A1: Limit Order.** The sell (buy) limit order is issued with a limit price of K , where K should be interpreted as a minimum price an agent is willing to accept *per share*. By placing a sell (buy) limit order the limit order writer effectively gives the other market participants a call (put) option. The value of this option provides a natural measure of liquidity costs.
2. **A2: Pricing Kernel and Fundamental Value.** The economy is equipped with a pricing kernel of the form:

$$\frac{d\Lambda_t}{\Lambda_t} = -r dt - \theta dZ \quad (1)$$

where $\theta = \frac{\mu - r}{\sigma}$. The fundamental value of a share to be acquired (sold) follows:

$$dV_t = \mu V_t dt + \sigma V_t dZ \quad (2)$$

and is assumed to be common knowledge in the economy.

3. **A3: Fill Dynamics.** The probability that a Q -share sell (buy) limit order is executed by the oncoming flow of buy (sell) market orders over an interval dt is given by $\lambda^B(Q)$ ($\lambda^S(Q)$). The arrival intensities of price-insensitive order flow are assumed to be constant over time, but quantity dependent reflecting the fact that the expected completion times for limit orders is a function of the order quantity, Q . Under the above assumptions, the cumulative survival probability for the time interval (t, T) , i.e. the probability that the limit order is not filled during this interval, is given by:

$$\pi(t, T) = \exp \left\{ - \int_t^T \lambda^i(Q) ds \right\} \quad (3)$$

The value of the limit order at time t , $L(\cdot)$, is uniquely determined by the prevailing fundamental value, P_t , the instantaneous arrival rate of opposing, price-insensitive order flow, $\lambda^i(Q)$, and the demanded quantity, Q . At any point in time, we allow for one of three possible events to occur:

1. The market maker fills the Q -share limit order in its *entirety* (i.e. we do not allow the market maker to deliver partial fills to the sell limit order writer). Should the market maker decide to execute the sell (buy) limit order at time t , he will realize a payoff of $Q \cdot (V_t - K)$ from a sell limit order and $Q \cdot (K - V_t)$ from a buy limit order. Clearly, since the sell (buy) limit order confers a *right* and not an obligation to acquire (sell) Q shares of a limited liability asset at price K , the value of the problem is bounded below by zero and it will never be optimal to exercise a sell (buy) limit order when it is out-of-the-money, $V_t < K$ ($V_t > K$).
2. The sell (buy) limit order is left unfilled by the market maker but is filled in its *entirety* by the oncoming flow of price-insensitive buy (sell) market orders.
3. The limit order survives in an unaltered form.

In this Appendix, we consider the case of a **perpetual** limit order. In other words, a limit order that is not subject to cancelation by the limit order writer. Despite their perpetual nature, the presence of oncoming, price-insensitive order flow will guarantee that the expected lifetime of any limit order will be finite. In particular, given our assumption about the arrival rates of price-insensitive order flow, the expected lifetime of a limit order will be distributed exponentially, with a mean of $\frac{1}{\lambda_i(Q)}$.

Because the Q -share sell (buy) limit order confers a right to buy (sell) a block of shares onto the market participants at a pre-specified price, it can be thought of as a perpetual call (put) option with a strike price, K , equal to the limit order price. Although the option implicit in the limit order is perpetual, the presence of the oncoming order flow acts as a dividend on the option, making early exercise optimal. Moreover, the memoryless property of the order fill process (i.e. order arrivals are Poisson) preserves the time-stationarity of the perpetual option valuation problem, and makes the optimal early exercise rule a barrier rule.

Let $F_{Q,t} = Q \cdot V_t$ denote the fundamental value of the block of shares underlying a Q -share limit order, and let the value of the limit order itself be denoted by $L(V_t, Q, K, t)$. Due to the possibility of order execution, the evolution of the option price will be described by the following law of motion:

$$\begin{aligned} dL(V_t, Q, K, t) &= \lim_{\Delta t \rightarrow 0} \left[\pi(t, t + \Delta t) \cdot (L(V_{t+\Delta t}, Q, K, t) - L(V_t, Q, K, t)) + \right. \\ &\quad \left. + (1 - \pi(t, t + \Delta t)) \cdot (0 - L(V_t, Q, K, t)) \right] \end{aligned}$$

where $\pi(t, t + \Delta t)$ denotes the probability that the limit order remains unexercised over an interval of length Δt . The first term in the above equation reflects the evolution of the limit order value in the *continuation* region. The second term represents

the loss in value due to the disappearance of the option as a results of its exercise by the oncoming, price-insensitive order flow. Rearranging slightly, we obtain:

$$\begin{aligned} dL(V_t, Q, K, t) &= \lim_{\Delta t \rightarrow 0} \left[L(V_{t+\Delta t}, Q, K, t) - L(P_t, Q, K, t) - (1 - \pi(t, t + \Delta t)) \cdot L(V_{t+\Delta t}, Q, K, t) \right] \\ &= L_F \cdot (dF_{q,t}) + \frac{1}{2} L_{FF} \cdot (dF_{q,t})^2 - (\lambda^i(Q) dt) \cdot L \end{aligned} \quad (4)$$

where we have used subscripts to denote partial derivatives and dropped the notation that explicitly depicts the dependence of $L(\cdot)$ on the underlying variables.

To derive the differential equation characterizing the evolution of the value of the Q -share option we impose the no arbitrage pricing condition:

$$E_t[d(\Lambda_t L)] = \Lambda_t E_t[dL] + L E_t[d\Lambda_t] + E_t[d\Lambda_t dL] = 0$$

or equivalently:

$$E_t\left[\frac{dL}{L}\right] + E_t\left[\frac{d\Lambda_t}{\Lambda_t}\right] = -E_t\left[\frac{d\Lambda_t}{\Lambda_t} \cdot \frac{dL}{L}\right]$$

Substituting in the pricing kernel and the relevant SDE for $dF_{Q,t}$ (from Ito's Lemma) we obtain the following ODE:

$$L_F \cdot (r F_{Q,t}) + \frac{1}{2} L_{FF} \cdot (\sigma F_{Q,t})^2 - (r + \lambda^i(Q)) L = 0 \quad (5)$$

The ODE is immediately recognizable as an Euler equidimensional equation whose solution is given by a linear combination of power functions:

$$L^j(V_t, Q, K, t) = \alpha_0 F_{Q,t}^{\phi_+} + \alpha_1 F_{Q,t}^{\phi_-}$$

where (α_0, α_1) are two constants of integration that will be determined from the boundary conditions. We use the superscript, j , on the value of the limit order to indicate that the same functional form applies to sell and buy limit orders ($j = S$ and $j = B$, respectively). The value of a sell (buy) limit order will depend on the quantity-dependent arrival intensity, $\lambda^i(Q)$, of buy (sell) market orders ($i = B$ and $i = S$, respectively) through the power coefficient, $\phi_{\pm}(\lambda^i)$. To solve for the power coefficients, $\phi_{\pm}(\lambda^i)$, substitute the guess $F_{Q,t}^{\phi}$ into (5) to obtain a quadratic equation in ϕ :

$$r \cdot \phi + \frac{\sigma^2}{2} \cdot \phi(\phi - 1) - (r + \lambda^i(Q)) = 0 \quad (6)$$

The power coefficients are given by the roots of this equation:

$$\phi_{\pm}(\lambda^i) = \left(\frac{1}{2} - \frac{r}{\sigma^2} \right) \pm \sqrt{\left(\frac{1}{2} - \frac{r}{\sigma^2} \right)^2 + \frac{2(r + \lambda^i(Q))}{\sigma^2}} \quad (7)$$

A.1 Sell limit orders

To solve for the value of a perpetual sell limit order subject to cancelation at a rate $\lambda^B(Q)$, we first conjecture that it will be optimal for the market maker to fill the limit order whenever fundamental value V_t reaches V^* (from below). We then impose the following boundary conditions:

$$\begin{aligned} \lim_{V_t \downarrow 0} L^{(S)} &= 0 \\ \lim_{V_t \uparrow V^*} L^{(S)} &= Q \cdot (V^* - K) \\ \lim_{V_t \uparrow V^*} L_F^{(S)} &= 1 \end{aligned}$$

The first conditions indicates that the call option becomes worthless as the value of the underlying tends to zero. The second and third conditions, respectively, correspond to the value matching and smooth pasting conditions at the optimal exercise threshold, V^* . This yields a solution to the ODE (5) of the form:

$$L^S(V_t, Q, K, t) = \begin{cases} \frac{QK}{\phi_+(\lambda^B)-1} \cdot \left(\frac{\phi_+(\lambda^B)-1}{\phi_+(\lambda^B)} \cdot \frac{V_t}{K} \right)^{\phi_+(\lambda^B)} & V_t < V^* \\ Q \cdot (V^* - K) & V_t \geq V^* \end{cases} \quad (8)$$

with the associated optimal barrier level, $V^* = \left(\frac{\phi_+(\lambda^B)}{\phi_+(\lambda^B)-1} \right) \cdot K$.

A.2 Buy limit orders

A symmetric analysis is used to solve for the value of the a perpetual buy limit order subject to cancelation at a rate $\lambda^S(Q)$. Once again, we first conjecture that it will be optimal for the market maker to fill the limit order whenever fundamental value V_t reaches V^{**} (from above). We then impose the standard boundary conditions:

$$\begin{aligned} \lim_{V_t \uparrow \infty} L^{(B)} &= 0 \\ \lim_{V_t \downarrow V^{**}} L^{(B)} &= Q \cdot (K - V^{**}) \\ \lim_{V_t \downarrow V^{**}} L_F^{(B)} &= 1 \end{aligned}$$

to obtain a solution of the form:

$$L^B(V_t, Q, K, t) = \begin{cases} \frac{QK}{1 - \phi_-(\lambda^S)} \cdot \left(\frac{\phi_-(\lambda^S) - 1}{\phi_-(\lambda^S)} \cdot \frac{V_t}{K} \right)^{\phi_-(\lambda^S)} & V_t > V^{**} \\ Q \cdot (K - V^{**}) & V_t \leq V^{**} \end{cases} \quad (9)$$

The associated optimal barrier level is given by, $V^{**} = \left(\frac{\phi_-(\lambda^S)}{\phi_-(\lambda^S) - 1} \right) \cdot K$.

B Incorporating Regime Switching in Liquidity and Volatility

In this section, we consider the pricing of limit orders in a setting where the volatility of fundamental value and the arrival intensities of price-insensitive order flow are regime-specific. We allow for two regimes - a low-liquidity regime, l , and a high-liquidity regime, h - and assume that the regime-shifts occur according to a continuous-time Markov chain with the following infinitesimal generator, Ξ :

$$\Xi = \begin{bmatrix} -q_l & q_l \\ q_h & -q_h \end{bmatrix} \quad (10)$$

Under these assumptions, the process sojourns in state i for a duration that is exponentially distributed with parameter, q_i .¹ Naturally, the joint-evolution of (V_t, i) , where i is the current liquidity state ($i \in (l, h)$), continues to be Markov, suggesting that the optimal exercise rule for the perpetual option will be a barrier rule, but with a state-dependent barrier.² In the derivations that follow, we assume that $\sigma_l \geq \sigma_h$ and $\lambda_h^i(Q) > \lambda_l^i(Q)$.

B.1 Sell limit orders

Let V_h^* and V_l^* denote the optimal exercise barriers in the high- and low-liquidity states, respectively. Given our assumptions about the relative magnitudes of the model parameters in the two states, and the fact that a payoff to the option is bounded below by zero, we know that $K \leq V_h^* < V_l^*$. When $V_t > V_l^*$ the value of the sell limit order in both states, $L^{(l,S)}$ and $L^{(h,S)}$, is simply equal to $Q \cdot (V_t - K)$ since immediate exercise is optimal. When $V_t < V_h^*$ the state-dependent limit order value satisfies the following system of equations:

$$\begin{aligned} L_F^{(h,S)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(h,S)} \cdot (\sigma_h F_{Q,t})^2 - rL^{(h,S)} &= \lambda_h^B(Q) \cdot (L^{(h,S)} - 0) + q_h \cdot (L^{(h,S)} - L^{(l,S)}) \\ L_F^{(l,S)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(l,S)} \cdot (\sigma_l F_{Q,t})^2 - rL^{(l,S)} &= \lambda_l^B(Q) \cdot (L^{(l,S)} - 0) + q_l \cdot (L^{(l,S)} - L^{(h,S)}) \end{aligned}$$

and when $V_t \in [V_h^*, V_l^*]$ the state-dependent value functions satisfy the following pair of PDEs:

$$\begin{aligned} L_F^{(l,S)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(l,S)} \cdot (\sigma_l F_{Q,t})^2 - rL^{(l,S)} &= \lambda_l^B(Q) \cdot (L^{(l,S)} - 0) + q_l \cdot (L^{(l,S)} - Q \cdot (V_t - K)) \\ L^{(h,S)} &= Q \cdot (V_t - K) \end{aligned}$$

Our solution methodology follows the approach in Guo and Zhang (2004).

¹The rate matrix yields the following infinitesimal description:

$$\begin{aligned} P(Y_{t+dt} = i | Y_t = i) &= 1 - q_i dt + o(dt) \\ P(Y_{t+dt} = j | Y_t = i) &= q_i dt + o(dt) \quad i \neq j \end{aligned}$$

²A proof of the optimality of this exercise rule is provided by Theorem 3 in Guo and Zhang (2004).

The characteristic function associated with the first set of PDEs is:

$$g_1(\beta) \cdot g_2(\beta) = q_h \cdot q_l \quad (11)$$

where

$$\begin{aligned} g_1(\beta) &= (r + q_h + \lambda_h^B(Q)) - \left(r - \frac{1}{2}\sigma_h^2\right) \cdot \beta - \frac{1}{2}\sigma_h^2\beta^2 \\ g_2(\beta) &= (r + q_l + \lambda_l^B(Q)) - \left(r - \frac{1}{2}\sigma_l^2\right) \cdot \beta - \frac{1}{2}\sigma_l^2\beta^2 \end{aligned}$$

Guo (2001) shows that the characteristic function - a quartic polynomial - has four distinct roots, $\beta_1 < \beta_2 < 0 < \beta_3 < \beta_4$ such that the general form of the solution to the system of PDEs is given by:

$$L^{(h,S)} = \sum_{i=1}^4 A_i F_{Q,t}^{\beta_i} \quad \text{and} \quad L^{(l,S)} = \sum_{i=1}^4 B_i F_{Q,t}^{\beta_i} \quad (12)$$

with $B_i = k_i \cdot A_i$ and $k_i = k(\beta_i) = \frac{g_1(\beta_i)}{g_2(\beta_i)} = \frac{q_l}{g_2(\beta_i)}$. Since the functions $L^{(h,S)}$ and $L^{(l,S)}$ should remain bounded as $F_{Q,t} \rightarrow 0$ we can immediately exclude all the negative powers of $F_{Q,t}$, yielding:

$$\begin{aligned} L^{(h,S)} &= A_3 F_{Q,t}^{\beta_3} + A_4 F_{Q,t}^{\beta_4} \\ L^{(l,S)} &= B_3 F_{Q,t}^{\beta_3} + B_4 F_{Q,t}^{\beta_4} = k_3 A_3 F_{Q,t}^{\beta_3} + k_4 A_4 F_{Q,t}^{\beta_4} \end{aligned}$$

The second equation is an inhomogenous equation whose solution can be written as the sum of the function solving the corresponding homogenous PDE and a particular solution, $\nu_h(F_{q,t})$:

$$L^{(l,S)} = C_1 F_{Q,t}^{\phi_+(\lambda_l^B)} + C_2 F_{Q,t}^{\phi_-(\lambda_l^B)} + \nu_l(F_{q,t}) \quad (13)$$

where $\phi_{\pm}(\lambda_l^B)$ are given by:

$$\phi_{\pm}(\lambda_l^B) = \left(\frac{1}{2} - \frac{r}{\sigma_l^2}\right) \pm \sqrt{\left(\frac{1}{2} - \frac{r}{\sigma_l^2}\right)^2 + \frac{2(r + q_l + \lambda_l^B(Q))}{\sigma_l^2}} \quad (14)$$

and the particular solution takes on the form:

$$\nu_l(F_{q,t}) = \left(\frac{q_l}{q_l + \lambda_l^B(Q)}\right) \cdot F_{Q,t} - \left(\frac{q_l}{r + q_l + \lambda_l^B(Q)}\right) \cdot QK \quad (15)$$

In order to solve for $(A_3, A_4, C_1, C_2, V_h^*, V_l^*)$, we apply value matching and smooth pasting conditions at the two exercise thresholds, obtaining a system of six equations in six unknowns:

$$\left\{ \begin{array}{l} k_3 A_3 (Q \cdot V_h^*)^{\beta_3} + k_4 A_4 (Q \cdot V_h^*)^{\beta_4} = Q \cdot (V_h^* - K) \\ \beta_3 k_3 A_3 (Q \cdot V_h^*)^{\beta_3} + \beta_4 k_4 A_4 (Q \cdot V_h^*)^{\beta_4} = Q \cdot V_h^* \\ A_3 (Q \cdot V_h^*)^{\beta_3} + A_4 (Q \cdot V_h^*)^{\beta_4} = C_1 (Q \cdot V_h^*)^{\phi_+(\lambda_l^B)} + C_2 (Q \cdot V_h^*)^{\phi_-(\lambda_l^B)} + \nu_h(Q \cdot V_h^*) \\ \beta_3 A_3 (Q \cdot V_h^*)^{\beta_3} + \beta_4 A_4 (Q \cdot V_h^*)^{\beta_4} = \phi_+(\lambda_l^B) C_1 (Q \cdot V_h^*)^{\phi_+(\lambda_l^B)} + \phi_-(\lambda_l^B) C_2 (Q \cdot V_h^*)^{\phi_-(\lambda_l^B)} + (Q \cdot V_h^*) \cdot \nu_l'(Q \cdot V_h^*) \\ C_1 (Q \cdot V_l^*)^{\phi_+(\lambda_l^B)} + C_2 (Q \cdot V_l^*)^{\phi_-(\lambda_l^B)} + \nu_l(Q \cdot V_h^*) = Q \cdot (V_l^* - K) \\ \phi_+(\lambda_l^B) C_1 (Q \cdot V_l^*)^{\phi_+(\lambda_l^B)} + \phi_-(\lambda_l^B) C_2 (Q \cdot V_l^*)^{\phi_-(\lambda_l^B)} + (Q \cdot V_l^*) \cdot \nu_l'(Q \cdot V_l^*) = Q \cdot V_l^* \end{array} \right.$$

yielding a state-dependent value function of the form:

$$L^{(l,S)} = \begin{cases} A_3 F_{Q,t}^{\beta_3} + A_4 F_{Q,t}^{\beta_4} & V_t < V_h^* \\ C_1 F_{Q,t}^{\phi_+(\lambda_l^B)} + C_2 F_{Q,t}^{\phi_-(\lambda_l^B)} + \nu_l(F_{q,t}) & V_h^* \leq V_t < V_l^* \\ Q \cdot (V_t - K) & V_t \geq V_l^* \end{cases} \quad (16)$$

$$L^{(h,S)} = \begin{cases} B_3 F_{Q,t}^{\beta_3} + B_4 F_{Q,t}^{\beta_4} & V_t < V_h^* \\ Q \cdot (V_t - K) & V_t \geq V_h^* \end{cases} \quad (17)$$

B.2 Buy limit orders

Let V_h^{**} and V_l^{**} denote the optimal exercise barriers in the high- and low-liquidity states, respectively. Given our assumptions about the relative magnitudes of the model parameters in the two states and the fact that a payoff to the option is bounded below by zero, we know that $V_l^* < V_h^* \leq K$. When $V_t < V_l^*$ the value of the sell limit order in both states, $L^{(l,B)}$ and $L^{(h,B)}$, is simply equal to $Q \cdot (K - V_t)$ since immediate exercise is optimal. When $V_t > V_h^*$ the state-dependent limit order

value satisfies the following system of equations:

$$\begin{aligned} L_F^{(h,B)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(h,B)} \cdot (\sigma_h F_{Q,t})^2 - rL^{(h,S)} &= \lambda_h^S(Q) \cdot (L^{(h,B)} - 0) + q_h \cdot (L^{(h,B)} - L^{(l,B)}) \\ L_F^{(l,B)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(l,B)} \cdot (\sigma_l F_{Q,t})^2 - rL^{(l,S)} &= \lambda_l^S(Q) \cdot (L^{(l,B)} - 0) + q_l \cdot (L^{(l,B)} - L^{(h,B)}) \end{aligned}$$

and when $V_t \in [V_l^*, V_h^*]$ the state-dependent value functions satisfy the following pair of PDEs:

$$\begin{aligned} L_F^{(l,B)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(l,B)} \cdot (\sigma_h F_{Q,t})^2 - rL^{(l,B)} &= \lambda_l^S(Q) \cdot (L^{(l,B)} - 0) + q_l \cdot (L^{(l,B)} - Q \cdot (K - V_t)) \\ L^{(h,B)} &= Q \cdot (K - V_t) \end{aligned}$$

Our solution methodology follows the approach in Guo and Zhang (2004).

The characteristic function associated with the first set of PDEs is:

$$g_1(\beta) \cdot g_2(\beta) = q_h \cdot q_l \quad (18)$$

where

$$\begin{aligned} g_1(\beta) &= (r + q_h + \lambda_h^S(Q)) - \left(r - \frac{1}{2} \sigma_h^2 \right) \cdot \beta - \frac{1}{2} \sigma_h^2 \beta^2 \\ g_2(\beta) &= (r + q_l + \lambda_l^S(Q)) - \left(r - \frac{1}{2} \sigma_l^2 \right) \cdot \beta - \frac{1}{2} \sigma_l^2 \beta^2 \end{aligned}$$

Guo (2001) shows that the characteristic function - a quartic polynomial - has four distinct roots, $\beta_1 < \beta_2 < 0 < \beta_3 < \beta_4$ such that the general form of the solution to the system of PDEs is given by:

$$L^{(h,B)} = \sum_{i=1}^4 A_i F_{Q,t}^{\beta_i} \quad \text{and} \quad L^{(l,B)} = \sum_{i=1}^4 B_i F_{Q,t}^{\beta_i} \quad (19)$$

with $B_i = k_i \cdot A_i$ and $k_i = k(\beta_i) = \frac{g_1(\beta_i)}{g_2(\beta_i)} = \frac{q_l}{g_2(\beta_i)}$. Since the functions $L^{(h,B)}$ and $L^{(l,B)}$ should remain bounded as $F_{Q,t} \rightarrow \infty$ we can immediately exclude all the positive powers of $F_{Q,t}$, yielding:

$$\begin{aligned} L^{(h,B)} &= A_1 F_{Q,t}^{\beta_1} + A_2 F_{Q,t}^{\beta_2} \\ L^{(l,B)} &= B_1 F_{Q,t}^{\beta_1} + B_2 F_{Q,t}^{\beta_2} = k_1 A_1 F_{Q,t}^{\beta_1} + k_2 A_2 F_{Q,t}^{\beta_2} \end{aligned}$$

The second equation is an inhomogenous equation whose solution can be written as the sum of the function solving the corresponding homogenous PDE and a particular solution, $\nu_l(F_{q,t})$:

$$L^{(l,B)} = C_1 F_{Q,t}^{\phi_+(\lambda_l^S)} + C_2 F_{Q,t}^{\phi_-(\lambda_l^S)} + \nu_l(F_{q,t}) \quad (20)$$

where $\phi_{\pm}(\lambda_l^S)$ are given by:

$$\phi_{\pm}(\lambda_l^S) = \left(\frac{1}{2} - \frac{r}{\sigma_l^2} \right) \pm \sqrt{\left(\frac{1}{2} - \frac{r}{\sigma_l^2} \right)^2 + \frac{2(r + q_l + \lambda_l^S(Q))}{\sigma_l^2}} \quad (21)$$

and the particular solution takes on the form:

$$\nu_l(F_{q,t}) = \left(\frac{q_l}{r + q_l + \lambda_l^S(Q)} \right) \cdot QK - \left(\frac{q_l}{q_l + \lambda_l^S(Q)} \right) \cdot F_{Q,t} \quad (22)$$

In order to solve for $(A_1, A_2, C_1, C_2, V_h^{**}, V_l^{**})$ we apply value matching and smooth pasting conditions at the two exercise thresholds obtaining a system of six equations in six unknowns:

$$\left\{ \begin{aligned} k_1 A_1 (Q \cdot V_h^{**})^{\beta_1} + k_2 A_2 (Q \cdot V_h^{**})^{\beta_2} &= Q \cdot (K - V_h^{**}) \\ \beta_1 k_1 A_1 (Q \cdot V_h^{**})^{\beta_1} + \beta_2 k_2 A_2 (Q \cdot V_h^{**})^{\beta_2} &= -Q \cdot V_h^{**} \\ A_1 (Q \cdot V_h^{**})^{\beta_1} + A_2 (Q \cdot V_h^{**})^{\beta_2} &= C_1 (Q \cdot V_h^{**})^{\phi_+(\lambda_l^S)} + C_2 (Q \cdot V_h^{**})^{\phi_-(\lambda_l^S)} + \nu_l(Q \cdot V_h^{**}) \\ \beta_1 A_1 (Q \cdot V_h^{**})^{\beta_1} + \beta_2 A_2 (Q \cdot V_h^{**})^{\beta_2} &= \phi_+(\lambda_l^S) C_1 (Q \cdot V_h^{**})^{\phi_+(\lambda_l^S)} + \phi_-(\lambda_l^S) C_2 (Q \cdot V_h^{**})^{\phi_-(\lambda_l^S)} + (Q \cdot V_h^{**}) \cdot \nu_l'(Q \cdot V_h^{**}) \\ C_1 (Q \cdot V_l^{**})^{\phi_+(\lambda_l^S)} + C_2 (Q \cdot V_l^{**})^{\phi_-(\lambda_l^S)} + \nu_l(Q \cdot V_l^{**}) &= Q \cdot (K - V_l^{**}) \\ \phi_+(\lambda_l^S) C_1 (Q \cdot V_l^{**})^{\phi_+(\lambda_l^S)} + \phi_-(\lambda_l^S) C_2 (Q \cdot V_l^{**})^{\phi_-(\lambda_l^S)} + (Q \cdot V_l^{**}) \cdot \nu_l'(Q \cdot V_l^{**}) &= -Q \cdot V_l^{**} \end{aligned} \right.$$

yielding a state-dependent value function of the form:

$$L^{(l,B)} = \begin{cases} A_1 F_{Q,t}^{\beta_1} + A_2 F_{Q,t}^{\beta_2} & V_h^{**} < V_t \\ C_1 F_{Q,t}^{\phi_+ (\lambda_1^S)} + C_2 F_{Q,t}^{\phi_- (\lambda_1^S)} + \nu_l(F_{q,t}) & V_l^{**} < V_t \leq V_h^{**} \\ Q \cdot (K - V_t) & V_t \leq V_l^{**} \end{cases} \quad (23)$$

$$L^{(h,B)} = \begin{cases} B_1 F_{Q,t}^{\beta_1} + B_2 F_{Q,t}^{\beta_2} & V_t > V_h^{**} \\ Q \cdot (K - V_t) & V_t \leq V_h^{**} \end{cases} \quad (24)$$

C Finite Duration Limit Orders

The baseline model assumes that the trader submits limit order of infinite duration, in other words, the limit order writer never cancels a submitted limit order. To examine the robustness of the model with respect to this assumption we examine the pricing of limit orders with a finite maturity. To do so, we will assume that the limit order writer will cancel the order at the N -th arrival time of a Poisson process with intensity, η . Under this auxiliary assumption, the order cancellation time, τ , will be Erlang distributed:

$$Pr\{\tau \in dt\} = \frac{\eta^N}{(N-1)!} t^{N-1} e^{-\eta t} dt \quad (25)$$

and the expectation and variance of the cancellation time, τ , will be given by:

$$E[\tau] = \frac{N}{\eta} \quad Var[\tau] = \frac{N}{\eta^2} \quad (26)$$

In order to ensure that the limit order is canceled at time, T , with probability one, it is sufficient to set $\eta = \frac{N}{T}$, and allow $N \rightarrow \infty$. In this case, $E[\tau] \rightarrow T$ and $Var[\tau] \rightarrow 0$. This mathematical device, known as the randomization of the expiry date, was first introduced in the context of option valuation by Carr (1998).

Under the proposed order cancellation scheme, the value of a limit order will depend not only on the fundamental value of the underlying, $F_{Q,t}$, and the arrival intensity of the opposing, price-insensitive order flow, $\lambda^i(Q)$, but also, on the number of periods left to the termination of the option, n . Specifically, we assume that when the limit order is issued $n = N$, and the limit order is canceled when $n = 0$, with the counter n decrementing at the jump times of the $Poisson(\eta)$ variable. Given these assumptions, the value of the limit order will be given by the solution to the following system of N ordinary differential equations:

$$L_F^{(n)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(n)} \cdot (\sigma F_{Q,t})^2 - r \cdot L^{(n)} = \lambda^i(Q) \cdot (L^{(n)} - 0) + \eta \cdot (L^{(n)} - L^{(n-1)}) \quad \text{for } n = N \dots 1 \quad (27)$$

with $L^{(0)} = 0$ (i.e. the limit order becomes worthless upon cancelation). The terms on the left hand side of the equality represent the evolution of the limit order value in the absence of jumps, while the terms on the right hand side represent the probability weighted losses from order exercise by oncoming order flow and the elapsing of time (as measured by the jumps in the $Poisson(\eta)$ variable). To solve this system of ODEs we proceed by backwards recursion, starting with state $n = 1$. As before, the limit order price, corresponding to the option strike price, is denoted by K .

C.1 Sell limit orders

C.1.1 Base case ($n = 1$)

Setting $n = 1$ we immediately see that (27) specializes to:

$$L_F^{(1)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(1)} \cdot (\sigma F_{Q,t})^2 - (r + \eta + \lambda^B(Q)) \cdot L^{(1)} = 0 \quad (28)$$

Given the time-homogeneity of the model, the optimal exercise rule for the option will be barrier rule, $V_{n=1}^*$. Using the boundary conditions:

$$\begin{aligned} \lim_{V_t \downarrow 0} L^{(1)} &= 0 \\ \lim_{V_t \uparrow V_1^*} L^{(1)} &= Q \cdot (V_1^* - K) \\ \lim_{V_t \uparrow V_1^*} L_F^{(1)} &= 1 \end{aligned}$$

Following the solution methodology of Appendix A it is simple to show that the value of the Q -share sell limit order is given by:

$$\begin{cases} L^{(1,S)}(V_t, Q, K, t) = \frac{Q \cdot K}{\phi_+(\lambda^B) - 1} \cdot \left(\frac{\phi_+(\lambda^B) - 1}{\phi_+(\lambda^B)} \cdot \frac{1}{Q \cdot K} \right)^{\phi_+(\lambda^B)} \cdot F_{Q,t}^{\phi_+(\lambda^B)} & V_t < V_1^* \\ L^{(1,S)}(V_t, Q, K, t) = Q \cdot (V_{n=1}^* - K) & V_t \geq V_1^* \end{cases} \quad (29)$$

where $F_{Q,t} = Q \cdot V_t$ and:

$$\phi_+(\lambda^B) = \left(\frac{1}{2} - \frac{r}{\sigma^2} \right) + \sqrt{\left(\frac{1}{2} - \frac{r}{\sigma^2} \right)^2 + \frac{2(r + \eta + \lambda^B(Q))}{\sigma^2}}$$

The optimal exercise threshold is given by:

$$V_1^* = \left(\frac{\phi_+(\lambda^B)}{\phi_+(\lambda^B) - 1} \right) \cdot K$$

Intuitively, the above solution provides a description of the value and optimal early exercise rules for an limit order with an exponentially distributed cancelation time. Because the limit order can be either canceled by the order writer or executed by the oncoming order flow, its maturity date will have an exponential distribution with parameter $\eta + \lambda^B(Q)$. Moreover, when $\eta = 0$, it is trivial to verify that the solution collapses to one corresponding to the perpetual limit order.

C.1.2 Recursive case ($n > 1$)

To solve the general option valuation problem it is necessary to first note that in any state n the agent will perceive two possible continuation values, $L^{(n,S)}$, depending on the prevailing value of the fundamental value, V_t . If $V_t \geq V_{n-1,t}^*$, the agent anticipates the exercise of the option in the state following the update of the *Poisson*(η) counter. Conversely, if $V_t < V_{n-1,t}^*$ the agent will be made best off by leaving the option unexercised.³ For an arbitrary state n the option value must satisfy the following pair of PDEs:

$$\begin{cases} L_F^{(n,S)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(n,S)} \cdot (\sigma F_{Q,t})^2 - r \cdot L^{(n,S)} = \lambda^B(Q) \cdot (L^{(n,S)} - 0) + \eta \cdot (L^{(n,S)} - Q \cdot (V_t - K)) & V_t \geq V_{n-1}^* \\ L_F^{(n,S)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(n,S)} \cdot (\sigma F_{Q,t})^2 - r \cdot L^{(n,S)} = \lambda^B(Q) \cdot (L^{(n,S)} - 0) + \eta \cdot (L^{(n,S)} - L^{(n-1,S)}) & V_t < V_{n-1}^* \end{cases} \quad (30)$$

We first solve the case in which early exercise is optimal subsequent to a decrement in the *Poisson*(η) counter. Because this ODE for, $L^{(n,S)}$, is non-homogenous in this region, a general solution will take the form of a linear combination of the solution to the corresponding homogenous PDE and a particular solution satisfying the non-homogeneity (in this case given by the last term):

$$L^{(n,S)}(V_t \geq V_{n-1}^*) = L_{hom}^{(n,S)}(V_t \geq V_{n-1}^*) + L_p^{(n,S)}(V_t \geq V_{n-1}^*)$$

We begin by deriving a solution for the homogenous counterpart to the first PDE in (30):

$$L_F^{(n,S)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(n,S)} \cdot (\sigma F_{Q,t})^2 - (r + \eta + \lambda^B(Q)) \cdot L^{(n,S)} = 0 \quad (31)$$

Recognizing the equidimensional structure of the PDE it is immediate that the solution will be a linear combination of power functions. Moreover, since the corresponding characteristic function is not dependent on the state, n , the exponents will be given by $\phi_+(\lambda^B)$ and $\phi_-(\lambda^B)$. The solution to this PDE is therefore:

$$L_{hom}^{(n,S)}(V_t \geq V_{n-1}^*) = \alpha_{0,n} \cdot F_{Q,t}^{\phi_+(\lambda^B)} + \alpha_{1,n} \cdot F_{Q,t}^{\phi_-(\lambda^B)} \quad (32)$$

A particular solution to the non-homogeneous PDE can be identified by inspection:

$$L_p^{(n,S)}(V_t \geq V_{n-1}^*) = \left(\frac{\eta}{\eta + \lambda^B(Q)} \right) \cdot F_{Q,t} - \left(\frac{\eta}{r + \eta + \lambda^B(Q)} \right) \cdot QK \quad (33)$$

Now consider identifying the general form of the solution to the second PDE in (30). Depending on the structure of the value function, $L^{(n-1,S)}$, this equation can either end up being homogeneous or not. To maintain the greatest level of generality we will assume the latter and additionally guess that the general structure of $L^{(n-1,S)}$ will be of the following form:

$$L^{(n-1,S)} = \sum_{i=1}^k \epsilon_{i,n-1} F_{Q,t}^{\delta_i} - \epsilon_{0,n-1} \cdot (QK) \quad (34)$$

³Because the option value is increasing in time, intuition suggests that the optimal exercise boundaries $V_{n,t}^*$ will be monotonically increasing with $V_{0,t}^* = K$.

Under this auxiliary assumption, it is easy to verify that the particular solution to the second PDE in (30) will take the following form:

$$L_p^{(n,S)}(V_t < V_{n-1}^*) = \sum_{i=1}^k \pi_{i,n} F_{Q,t}^{\delta_i} - \pi_{0,n} \cdot (QK) \quad (35)$$

with:

$$\begin{aligned} \pi_{0,n} &= \frac{\eta}{r + \lambda^B(Q) + \eta} \cdot \epsilon_{0,n-1} \\ \pi_{i,n} &= -\frac{\eta}{\delta_i \cdot \left(r + (\delta_i - 1) \cdot \frac{\sigma^2}{2} \right) - (r + \lambda^B(Q) + \eta)} \cdot \epsilon_{i,n-1} \quad i = 1, \dots, k \end{aligned}$$

Lastly, because the homogenous counterparts to the first and second PDEs in (30) are identical we immediately have:

$$L_{hom}^{(n,S)}(V_t < V_{n-1}^*) = \beta_{0,n} \cdot F_{Q,t}^{\phi_+(\lambda^B)} + \beta_{1,n} \cdot F_{Q,t}^{\phi_-(\lambda^B)} \quad (36)$$

Moreover, because $\lim_{V_t \rightarrow 0} L^{(n,S)} = 0$ we can immediately exclude the negative exponent, $\phi_-(\lambda^B)$, by setting $\beta_{1,n} = 0$.

To complete the derivation we apply the smooth pasting and value matching conditions between $L^{(n,S)}(V_t \geq V_{n-1}^*)$ and $L^{(n,S)}(V_t < V_{n-1}^*)$ at V_{n-1}^* , and between $L^{(n,S)}(V_t \geq V_{n-1}^*)$ and $Q \cdot (V_t - K)$ at V_n^* . These four conditions will be used to pin down the values of $(\alpha_{0,n}, \alpha_{1,n}, \beta_{0,n}, V_n^*)$. The two pairs of conditions are given by:

$$\begin{aligned} \beta_{0,n} \cdot (QV_{n-1}^*)^{\phi_+(\lambda^B)} + \sum_{i=1}^k \pi_{i,n} (QV_{n-1}^*)^{\delta_i} - \pi_{0,n} \cdot (QK) &= \alpha_{0,n} \cdot (QV_{n-1}^*)^{\phi_+(\lambda^B)} + \alpha_{1,n} \cdot V_{n-1}^{\phi_-(\lambda^B)} + \\ &+ \left(\frac{\eta}{\eta + \lambda^B(Q)} \right) \cdot (QV_{n-1}^*) - \left(\frac{\eta}{r + \eta + \lambda^B(Q)} \right) \cdot QK \quad (37) \end{aligned}$$

$$\begin{aligned} \beta_{0,n} \cdot \phi_+(\lambda^B) \cdot (QV_{n-1}^*)^{\phi_+(\lambda^B)} + \sum_{i=1}^k \delta_i \pi_{i,n} \cdot (QV_{n-1}^*)^{\delta_i} &= \alpha_{0,n} \cdot \phi_+(\lambda^B) \cdot (QV_{n-1}^*)^{\phi_+(\lambda^B)} + \alpha_{1,n} \cdot \phi_-(\lambda^B) \cdot (QV_{n-1}^*)^{\phi_-(\lambda^B)} + \\ &+ \left(\frac{\eta}{\eta + \lambda^B(Q)} \right) \cdot (QV_{n-1}^*) \quad (38) \end{aligned}$$

and:

$$\alpha_{0,n} \cdot (QV_n^*)^{\phi_+(\lambda^B)} + \alpha_{1,n} \cdot (QV_n^*)^{\phi_-(\lambda^B)} + \left(\frac{\eta}{\eta + \lambda^B(Q)} \right) \cdot (QV_n^*) - \left(\frac{\eta}{r + \eta + \lambda^B(Q)} \right) \cdot QK = Q \cdot (V_n^* - K) \quad (39)$$

$$\alpha_{0,n} \cdot \phi_+(\lambda^B) \cdot (QV_n^*)^{\phi_+(\lambda^B)} + \alpha_{1,n} \cdot \phi_-(\lambda^B) \cdot (QV_n^*)^{\phi_-(\lambda^B)} + \left(\frac{\eta}{\eta + \lambda^B(Q)} \right) \cdot (QV_n^*) = QV_n^* \quad (40)$$

C.2 Buy limit orders

The case of buy limit order subject to stochastic cancelation is largely symmetric to that of sell limit orders, so the derivations are presented in a somewhat more compact manner.

C.2.1 Base case ($n = 1$)

Setting $n = 1$ we immediately see that (27) specializes to:

$$L_F^{(1)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(1)} \cdot (\sigma F_{Q,t})^2 - (r + \eta + \lambda^S(Q)) \cdot L^{(1)} = 0 \quad (41)$$

Given the time-homogeneity of the model, the optimal exercise rule for the option will be barrier rule, $V_{n=1}^{**}$. Using the boundary conditions:

$$\begin{aligned} \lim_{V_t \uparrow \infty} L^{(1)} &= 0 \\ \lim_{V_t \downarrow V_1^{**}} L^{(1)} &= Q \cdot (K - V_1^{**}) \\ \lim_{V_t \downarrow V_1^{**}} L_F^{(1)} &= 1 \end{aligned}$$

Using the same solution methodology as in Appendix A it is simple to show that the value of the Q -share sell limit order is given by:

$$L^{(1,B)}(V_t, Q, K, t) = \frac{Q \cdot K}{\phi_-(\lambda^S) - 1} \cdot \left(\frac{\phi_-(\lambda^S) - 1}{\phi_-(\lambda^S)} \cdot \frac{1}{Q \cdot K} \right)^{\phi_-(\lambda^S)} \cdot F_{Q,t}^{\phi_+(\lambda^S)} \quad (42)$$

where $F_{Q,t} = Q \cdot V_t$ and:

$$\phi_-(\lambda^S) = \left(\frac{1}{2} - \frac{r}{\sigma^2} \right) - \sqrt{\left(\frac{1}{2} - \frac{r}{\sigma^2} \right)^2 + \frac{2(r + \eta + \lambda^S(Q))}{\sigma^2}}$$

As before, $F_{Q,t}$, represents the fundamental value of the underlying block of shares. The fundamental value at which it is optimal to exercise the buy order early is given by:

$$V_1^* = \left(\frac{\phi_-(\lambda^S)}{\phi_-(\lambda^S) - 1} \right) \cdot K$$

As for the case of the sell limit order, the above expressions provide a complete characterization of the value and optimal exercise rules for a buy limit order which is subject to stochastic cancelation by the limit order writer at an instantaneous rate of η .

C.2.2 Recursive case ($n > 1$)

The system of PDEs describing the value of a buy limit order subject to cancelation at the N -th jump time of a Poisson process with intensity η , with n jumps to go, is essentially identical to (30). The sole modification concerns the ordering of the optimal exercise thresholds, V_n^{**} , which will not be *decreasing* in n . The system of PDEs is given by:

$$\begin{cases} L_F^{(n,B)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(n,B)} \cdot (\sigma F_{Q,t})^2 - r \cdot L^{(n,B)} = \lambda^S(Q) \cdot (L^{(n,B)} - 0) + \eta \cdot (L^{(n,B)} - Q \cdot (V_t - K)) & V_t \leq V_{n-1}^{**} \\ L_F^{(n,B)} \cdot (rF_{Q,t}) + \frac{1}{2} L_{FF}^{(n,B)} \cdot (\sigma F_{Q,t})^2 - r \cdot L^{(n,B)} = \lambda^S(Q) \cdot (L^{(n,B)} - 0) + \eta \cdot (L^{(n,B)} - L^{(n-1,B)}) & V_t > V_{n-1}^{**} \end{cases} \quad (43)$$

Proceeding as before the solution to the first, non-homogeneous equation will be given by:

$$L_{hom}^{(n,B)}(V_t \leq V_{n-1}^{**}) = \alpha_{0,n} \cdot F_{Q,t}^{\phi_+(\lambda^S)} + \alpha_{1,n} \cdot F_{Q,t}^{\phi_-(\lambda^S)} + L_p^{(n,B)}(V_t \leq V_{n-1}^{**}) \quad (44)$$

Moreover, since $\lim_{V_t \rightarrow \infty} L^{(n,B)} = 0$ it is immediate that the coefficient in front of the power function with exponent, $\phi_+(\lambda^B)$, will have to be equal to zero. That is, for the buy limit order, $\alpha_{0,n} = 0$. The form of the particular solution can be identified by inspection to be:

$$L_p^{(n,B)}(V_t \leq V_{n-1}^{**}) = \left(\frac{\eta}{r + \eta + \lambda^S(Q)} \right) \cdot QK - \left(\frac{\eta}{\eta + \lambda^S(Q)} \right) \cdot F_{Q,t} \quad (45)$$

To solve the second PDE in the system, (43), we propose and verify a general form for the value continuation value, $L^{(n-1,B)}$:

$$L^{(n-1,B)} = \epsilon_{0,n-1} \cdot (QK) + \sum_{i=1}^k \epsilon_{i,n-1} F_{Q,t}^{\delta_i} \quad (46)$$

Under this assumption, it is easy to verify that the particular solution to the second PDE will take the following form:

$$L_p^{(n,B)}(V_t > V_{n-1}^{**}) = \pi_{0,n} \cdot (QK) + \sum_{i=1}^k \pi_{i,n} F_{Q,t}^{\delta_i} \quad (47)$$

where $\pi_{0,n}$ and $\pi_{i,n}$ are given by the expressions derived in the section on the pricing of sell limit orders. Lastly, the homogenous solution to the second PDE is given by:

$$L_{hom}^{(n,B)}(V_t > V_{n-1}^{**}) = \beta_{0,n} \cdot F_{Q,t}^{\phi_+(\lambda^S)} + \beta_{1,n} \cdot F_{Q,t}^{\phi_-(\lambda^S)} \quad (48)$$

To complete the solution we apply the smooth-pasting and value-matching conditions at V_{n-1}^{**} and V_n^{**} . At V_{n-1}^{**} we require that the values and first derivatives of $L^{(n,B)}(V_t > V_{n-1}^{**})$ and $L^{(n,B)}(V_t \leq V_{n-1}^{**})$ be equal. And at V_n^{**} we require that the values and first derivatives of $L^{(n,B)}(V_t \leq V_{n-1}^{**})$ and $Q \cdot (K - V_t)$ be equal. This yields a system of four equations for the four

unknowns $(\alpha_{1,n}, \beta_{0,n}, \beta_{1,n}, V_n^{**})$. The two pairs of conditions are given by:

$$\alpha_{1,n} \cdot (QV_{n-1}^{**})^{\phi_-(\lambda^S)} + \left(\frac{\eta}{r + \eta + \lambda^S(Q)} \right) \cdot QK - \left(\frac{\eta}{\eta + \lambda^S(Q)} \right) \cdot (QV_{n-1}^{**}) = \beta_{0,n} \cdot (QV_{n-1}^{**})^{\phi_+(\lambda^S)} + \beta_{1,n} \cdot (QV_{n-1}^{**})^{\phi_-(\lambda^S)} + \pi_{0,n} \cdot (QK) + \sum_{i=1}^k \pi_{i,n} (QV_{n-1}^{**})^{\delta_i} \quad (49)$$

$$\alpha_{1,n} \cdot \phi_-(\lambda^S) \cdot (QV_{n-1}^{**})^{\phi_-(\lambda^S)} - \left(\frac{\eta}{\eta + \lambda^S(Q)} \right) \cdot (QV_{n-1}^{**}) = \beta_{0,n} \cdot \phi_+(\lambda^S) \cdot (QV_{n-1}^{**})^{\phi_+(\lambda^S)} + \beta_{1,n} \cdot \phi_-(\lambda^S) \cdot (QV_{n-1}^{**})^{\phi_-(\lambda^S)} + \pi_{0,n} \cdot (QK) + \sum_{i=1}^k \pi_{i,n} (QV_{n-1}^{**})^{\delta_i} \quad (50)$$

and:

$$\beta_{0,n} \cdot (QV_n^{**})^{\phi_+(\lambda^S)} + \beta_{1,n} \cdot (QV_n^{**})^{\phi_-(\lambda^S)} + \pi_{0,n} \cdot (QK) + \sum_{i=1}^k \pi_{i,n} (QV_n^{**})^{\delta_i} = Q \cdot (K - V_n^{**}) \quad (51)$$

$$\beta_{0,n} \cdot \phi_+(\lambda^S) \cdot (QV_n^{**})^{\phi_+(\lambda^S)} + \beta_{1,n} \cdot \phi_-(\lambda^S) \cdot (QV_n^{**})^{\phi_-(\lambda^S)} + \sum_{i=1}^k \pi_{i,n} \delta_i \cdot (QV_n^{**})^{\delta_i} = -QV_n^{**} \quad (52)$$

Figure 1: **Quantity structure of immediacy prices.** The cost of immediacy is computed as the fraction of the value of the underlying asset which has to be forgone to induce the market maker to execute the limit order instantaneously. The cost of immediacy is given by $\frac{K_j - V_t}{V_t}$. It is plotted against the limit order quantity assuming that the expected arrival time for a Q -share order is given by $\frac{Q}{\lambda^i(1)}$. The graph is calculated assuming a 5% annualized riskless rate and an expected arrival time of one second for a single share ($\lambda^i(1) = 1$).

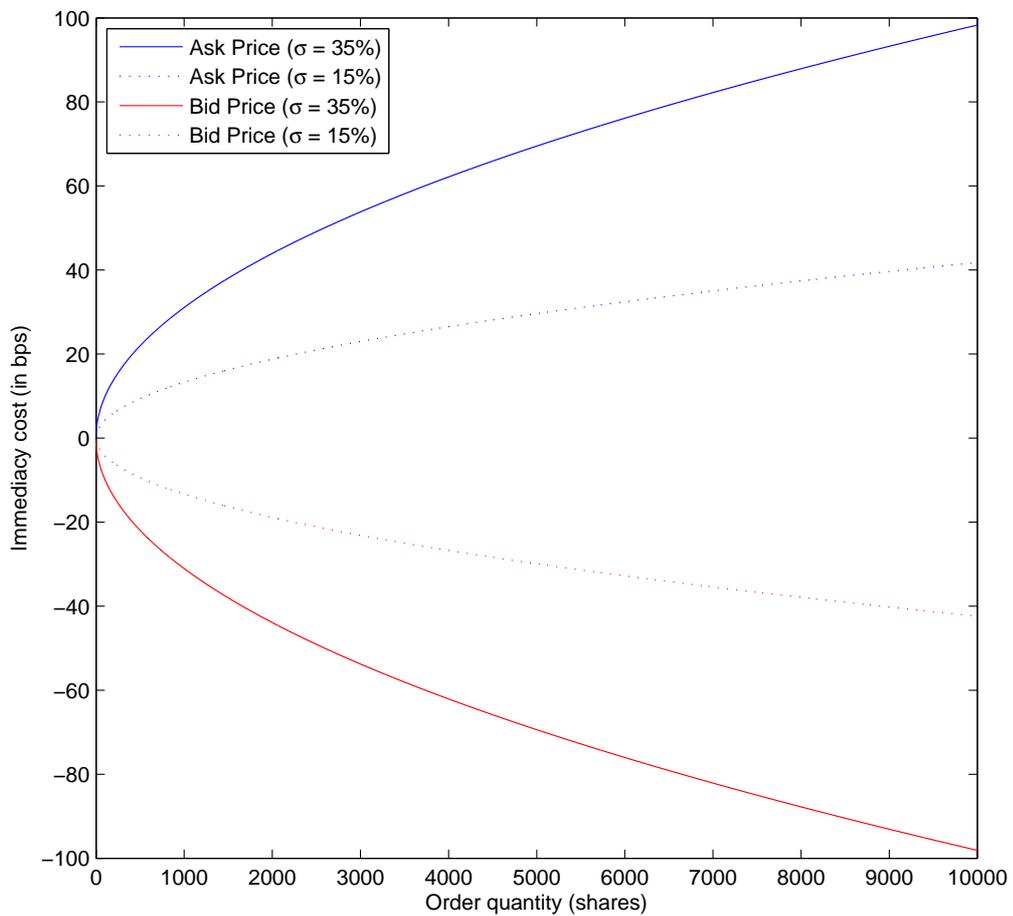


Figure 2: **Immediacy prices as a function of market liquidity.** The figure depicts the percentage cost of obtaining immediacy for a buy (sell) transaction - as a function of the order arrival rate of the price-insensitive order flow. The x-axis plots the base 10 logarithm of the mean share interarrival time, which is given by $\lambda^S(1)^{-1}$ ($\lambda^B(1)^{-1}$) for sells (buys). The fundamental volatility, σ , equals 15% or 35% *per annum*, and the riskless rate is fixed at 5% *per annum*

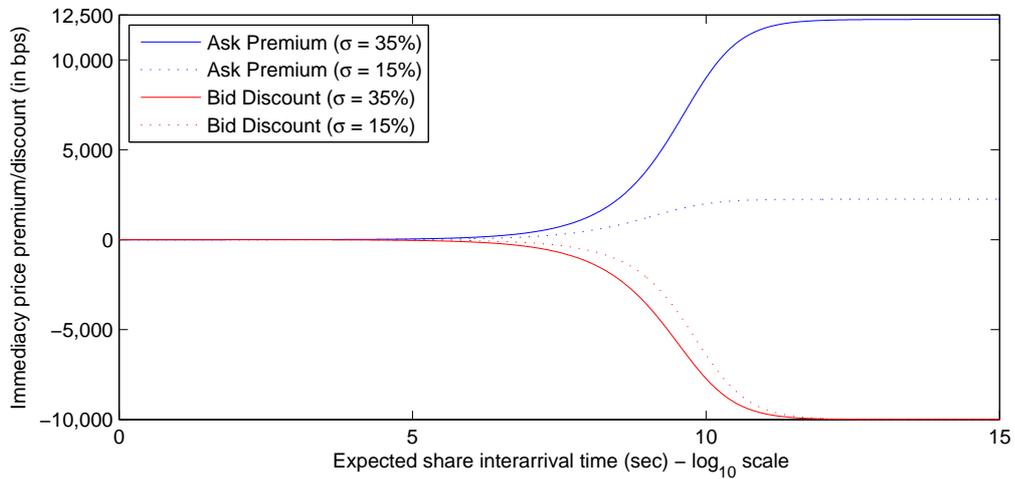
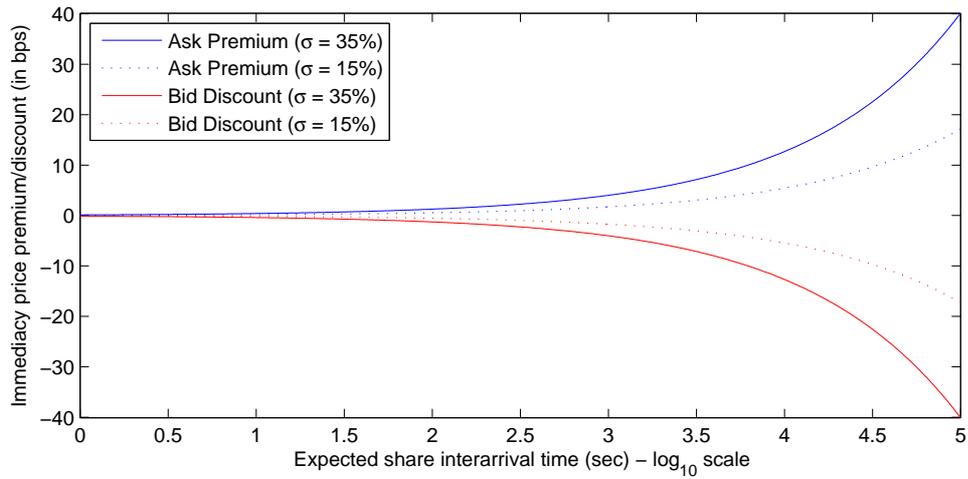


Figure 3: **Immediacy prices during liquidity events.** The figure depicts the percentage cost of obtaining immediacy for a buy (sell) transaction - as a function of the demanded quantity - during a liquidity event. We consider two possible liquidity event scenarios. In scenario 1, the intensity of sell (buy) arrivals, $\lambda^S(\cdot)$ ($\lambda^B(\cdot)$), increases (decreases) fivefold. In scenario 2, the change in the order arrival intensities is accompanied by an increase in the fundamental volatility from 15% to 35%. In the baseline formulation, buy and sell orders are assumed to arrive at a rate of one share per second; the riskless rate is fixed at 5%.

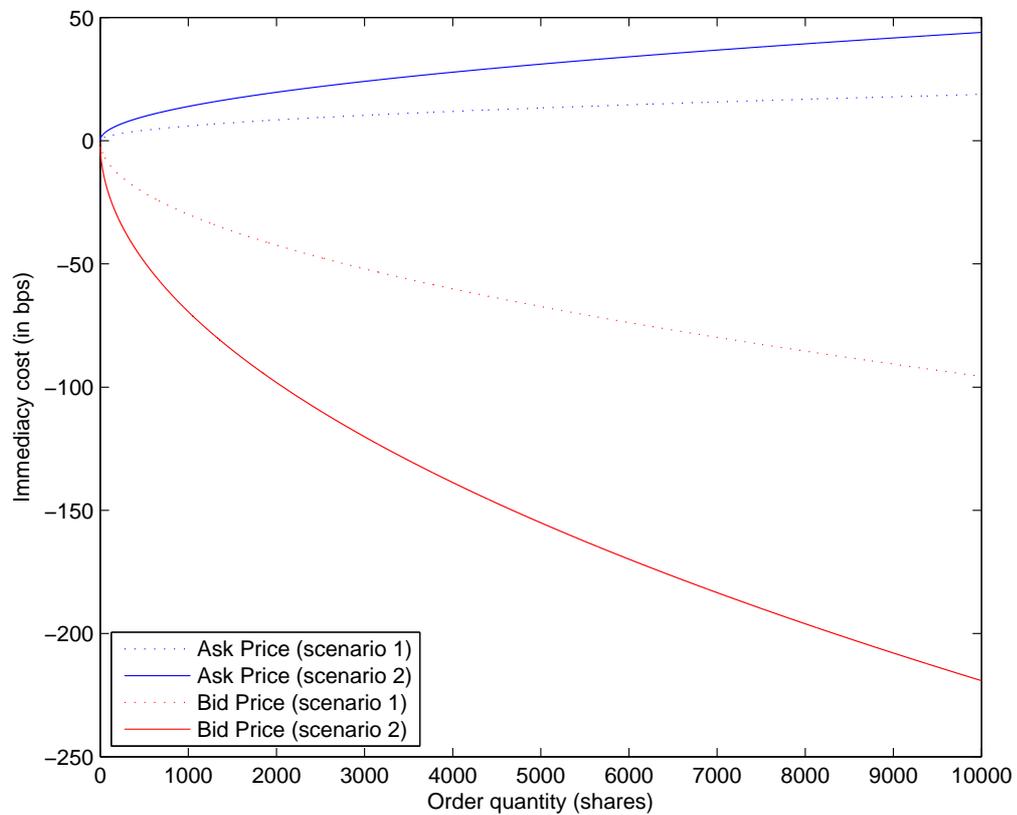


Figure 4: **Empirical calibration of $\lambda^i(Q)$ for Altria.** This figure displays the empirically observed mean waiting times for cumulative buy (sell) order flow to reach various quantities, Q . The plots are generated using transaction level (TAQ) data from 2004 for Altria (ticker: MO). Each plot overlays the OLS fitted relationship between the expected arrival time, $\lambda^i(Q)^{-1}$, and the order quantity, Q , under the simple proportionality model, as well as the fitted relationship implied by the naive estimator of the order arrival rate obtained by dividing total signed volume by total elapsed time. The left (right) panel corresponds to the buy (sell) order flow arrivals; the top (bottom) panel is plotted in the natural (log) scale.

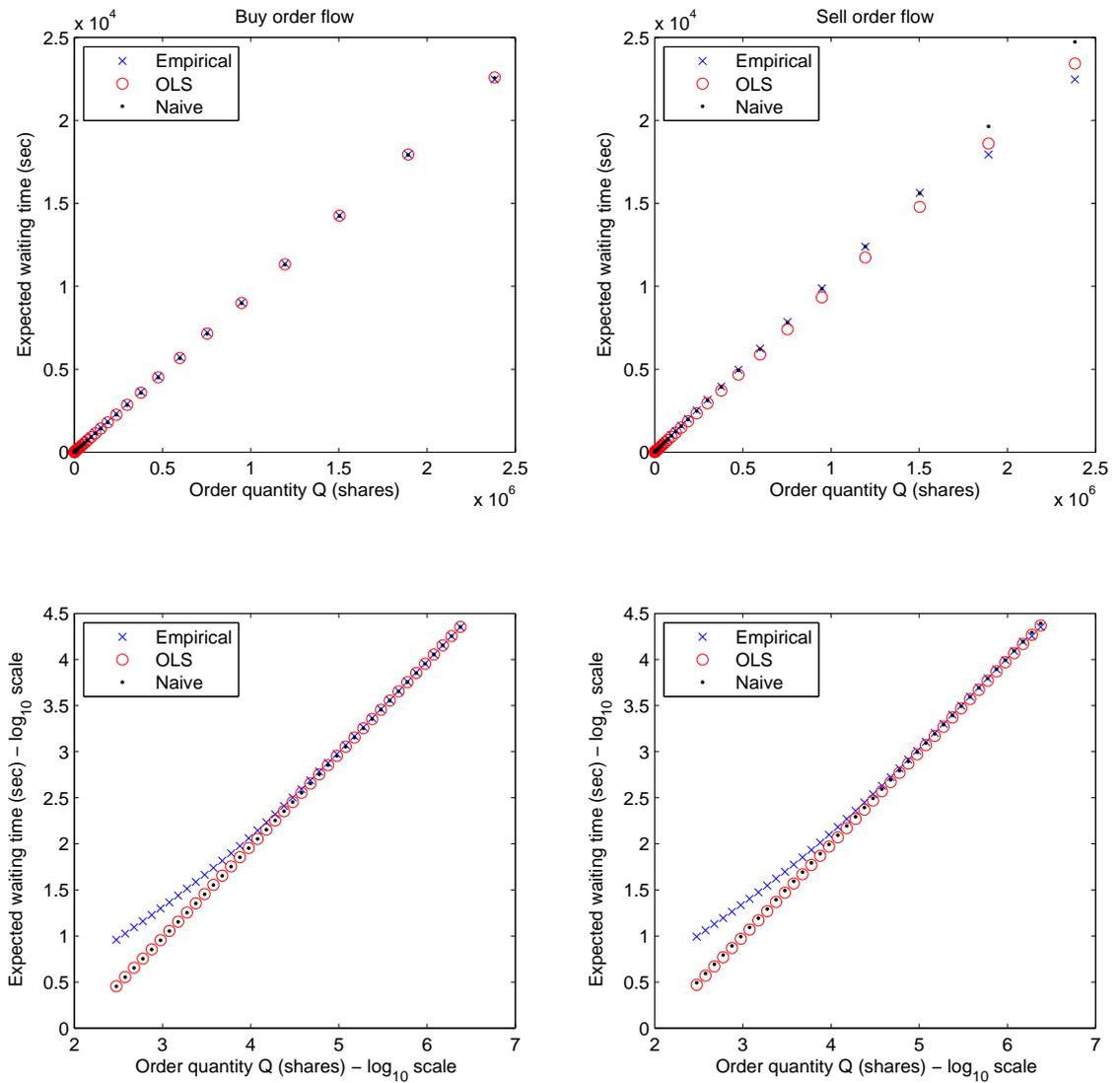
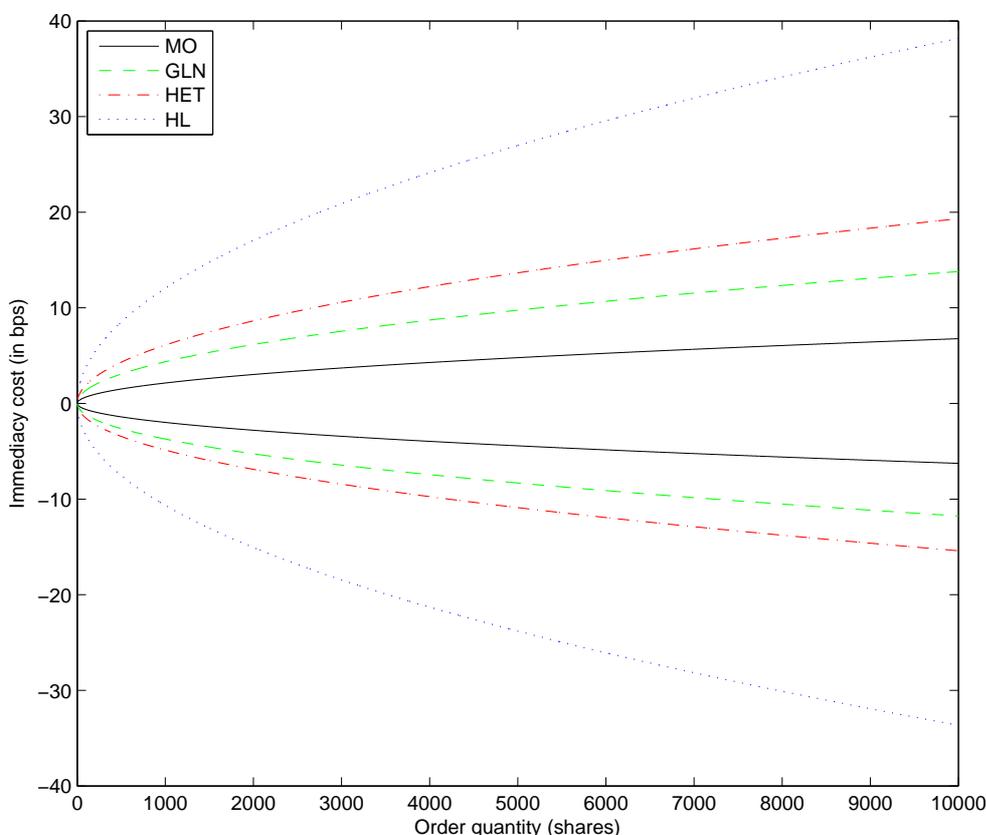


Figure 5: **Empirical quantity structure of immediacy prices.** This figure displays the model implied bid-ask schedules for four NYSE securities, calibrated to empirical data from 2004. The volatility of fundamental value, σ , is computed from the daily return series. The per second arrival rate for buys (sells) is estimated from an OLS regression of mean order interarrival times on order quantity computed using signed TAQ data. The four companies are Altria Group (ticker: MO), Corning (ticker: GLN), Harrahs Entertainment (ticker: HET), and Hecla Mining (ticker: HL). The table below the graph reports the parameter estimates, share prices and the market capitalizations of the firms (in MM\$). The remaining columns tabulate the model implied percentage costs for a 1,000 share, 10,000 share, \$10,000 and \$1,000,000 buy transaction. The column *Takeover* reports the premium for acquiring all of the shares outstanding. The interest rate in the calibration is fixed at 1.27% *per annum*, reflecting the average yield on a 30-day T-bill in 2004. Price and market capitalization data are as of 2004 year-end.



Ticker	$\hat{\sigma}$	$\hat{\lambda}^B(1)$	$\hat{\lambda}^S(1)$	Prc (\$)	Cap (MM\$)	1,000	10,000	\$100k	\$1M	Takeover
MO	22.1%	105.9	90.5	61.10	125,412	0.02%	0.07%	0.03%	0.09%	24.29%
GLN	46.8%	134.4	97.9	11.77	16,504	0.04%	0.14%	0.13%	0.40%	55.64%
HET	22.9%	18.7	11.9	66.89	7,496	0.06%	0.19%	0.07%	0.24%	17.99%
HL	50.2%	18.8	14.7	5.83	689	0.12%	0.38%	0.50%	1.59%	45.25%

Table 1: **Summary of estimated immediacy prices for NYSE stocks by size decile (2004)**. This table reports the sample means of order flow characteristics, volatilities, and estimated transaction costs for the universe of NYSE stocks by size decile. The buy and sell order arrival intensities are obtained from signed daily transaction level (TAQ) data for 2004. Return volatilities are computed using daily total return data from CRSP for the 2004 calendar year. Prices and market capitalizations reflect year-end values constructed using CRSP data.

Size Decile	λ^B	λ^S	σ	Market		Stock price	Cost for			Cost for \$1M
				Cap. (MM\$)	1,000 shares		10,000 shares	\$100K		
1	1.5	1.2	0.373	214	0.43%	15.15	0.43%	1.34%	1.21%	3.82%
2	3.4	2.6	0.369	527	0.25%	22.94	0.25%	0.80%	0.57%	1.81%
3	5.8	4.3	0.336	804	0.18%	27.61	0.18%	0.57%	0.36%	1.15%
4	7.1	5.3	0.317	1,142	0.16%	31.40	0.16%	0.50%	0.30%	0.95%
5	9.0	6.8	0.309	1,618	0.13%	35.11	0.13%	0.41%	0.24%	0.75%
6	9.8	7.5	0.269	2,276	0.11%	39.71	0.11%	0.34%	0.18%	0.58%
7	14.3	10.9	0.265	3,399	0.09%	41.79	0.09%	0.27%	0.14%	0.44%
8	23.7	17.6	0.258	5,581	0.06%	41.84	0.06%	0.20%	0.11%	0.33%
9	31.6	23.9	0.228	11,551	0.05%	57.13	0.05%	0.16%	0.07%	0.22%
10	75.3	60.6	0.211	52,875	0.03%	51.30	0.03%	0.09%	0.04%	0.14%