Optimal Execution with Nonlinear Impact Functions and Trading-Enhanced Risk

Robert F. Almgren*

October 2001

Abstract

We determine optimal trading strategies for liquidation of a large single-asset portfolio to minimize a combination of volatility risk and market impact costs. We take the market impact cost per share to be a power law function of the trading rate, with an arbitrary positive exponent. This includes, for example, the square-root law that has been proposed based on market microstructure theory. In analogy to the linear model, we define a "characteristic time" for optimal trading, which now depends on the initial portfolio size and decreases as execution proceeds. We also consider a model in which uncertainty of the realized price is increased by demanding rapid execution; we show that optimal trajectories are described by a "critical portfolio size" above which this effect is dominant and below which it may be neglected.

Key words: market impact, trading strategy, liquidity modeling.

*University of Toronto, Departments of Mathematics and Computer Science, `almgren@math.toronto.edu`. Research supported by the National Sciences and Engineering Research Council of Canada and by the University of Toronto.

1 Introduction

In the execution of large portfolio transactions, a trading strategy must be determined that balances the risk of delayed execution against the cost of rapid execution; the choice is roughly between an "active" and a "passive" trading strategy (Hasbrouck and Schwartz 1988; Wagner and Banks 1992). Several recent articles (Almgren and Chriss 1999; Almgren and Chriss 2000; Grinold and Kahn 1999; Konishi and Makimoto 2001) have constructed optimal strategies for this problem, under the assumption that liquidity cost per share traded is a linear function of trading rate or of block size, and that the only source of uncertainty in execution is the volatility of the underlying asset.[1]

In practice, linearity of trading costs is an unrealistic assumption. Perold and Salomon (1991) have argued that the liquidity premium per share demanded by the market will be either a convex or a concave function of block size, if the market's perception is that the trader is information-driven or liquidity-driven, respectively. In the Barra Market Impact Model (Barra 1997; Kahn 1993; Grinold and Kahn 1999; Loeb 1983) it is argued, based on a detailed analysis of the risk-reward choices faced by an equity market maker, that the liquidity premium per share should grow as the square root of the block size traded. Electronic trading systems such as Optimark (Rickard and Torre 1999) have been constructed to allow traders to specify precisely what liquidity premium they are willing to pay as a function of block size, and search for clearing opportunities in the mismatch between profiles of different market participants. These effects can be captured by introducing *nonlinear cost functions* into the cost function which is minimized to determine optimal trading strategies.[2]

An additional effect not considered in the theoretical strategies constructed in previous work is that the liquidity premium demanded by the market is not deterministic. In fact, this premium will depend on the presence in the market at that instant of participants who are willing to take the other side of the trade. Since the presence of these counterparties cannot be predicted in advance, it represents an additional source of risk incurred by

[1] There is an extensive literature studying the effect of block trades on prices (Kraus and Stoll 1972; Holthausen, Leftwich, and Mayers 1987; Holthausen, Leftwich, and Mayers 1990; Chan and Lakonishok 1993; Chan and Lakonishok 1995; Keim and Madhavan 1995; Keim and Madhavan 1997; Koski and Michaely 2000).

[2] Although linear models are commonly used in empirical regression analyses for simplicity, nonlinear models can be often emulated by dividing trades into categories by size (Huang and Stoll 1997; Bessembinder and Kaufman 1997). In fact, Chakravarty (2001) argues that medium-sized trades have a disproportionately large effect on prices.

the trading profile. That is, a more complete model should include *trading-enhanced risk,* representing the increased uncertainty in execution price incurred by demanding rapid execution of large blocks.[3]

Thus this paper extends the model of Almgren/Chriss, Grinold/Kahn, and others in two important ways:

- The liquidity premium, expressed as an unfavorable motion of the price per share, may be an increasing nonlinear function of the trading rate or block size (we consider either one to be a proxy for the other). This cost of trading is reduced by trading slowly, but it must be balanced against the volatility risk incurred by holding the initial portfolio longer than necessary. In particular, in Section 3, we provide exact solutions in the case that this function is a power law with an arbitrary positive exponent, which covers the range of behavior outlined above. Whereas in the linear case, optimal trajectories are characterized by a single "characteristic time" independent of portfolio size, in the nonlinear case the characteristic time depends on the initial portfolio size, and scales appropriately as the remaining portfolio is diminished during trading.

- The realized price per share is itself a random variable, whose variance increases with increasing rate of trading. This introduces an additional source of risk in addition to the volatility. In constrast to the effect of volatility, this additional risk is *decreased* by trading slowly, submitting small blocks for execution at each time. In Section 4 we construct nearly explicit optimal solutions including this effect, and use an asymptotic analysis to show that the effect of trading-enhanced risk is most important for large initial portfolios. Indeed, for any given set of parameters there is a characteristic portfolio size, above which the optimal strategy is dominated by the need to reduce trading-enhanced risk, and below which this effect may be ignored.

In the next section we describe the details of our model, and the general method by which we determine optimal solution trajectories. In subsequent sections we present explicit solutions in various particular cases.

[3] This is an implicit feature of the model described in Rickard and Torre (1999). Chordia, Subrahmanyam, and Anshuman (2001) and Hasbrouck and Seppi (2001) argue that liquidity fluctuates due to intrinsic variations in market activity independently of trade size. This effect is included in our model (the constant term $f(0)$), but we are additionally interested in the *increase* in execution price uncertainty due to large block sizes.

2 Model

We follow the general framework of Almgren and Chriss (2000). At time $t = 0$ we hold X shares of an asset, which is to be completely liquidated by time $t = T$. The initial size X is positive for a sell program, and negative for a buy program: in the former case we have long exposure to the market until we have eliminated our holdings, while in the latter we have short exposure until we have completed the purchase to which we have committed ourselves at $t = 0$. In this paper we focus on the case $X > 0$. In the case of a portfolio trading problem X may be a vector, but we consider only the case of a single asset.

We denote by $x(t)$ our share holdings at time t, with $x(0) = X$ and $x(T) = 0$; the problem is to determine the optimal function $x(\cdot)$ so as to minimize a chosen cost functional. Later, we will take the limit $T \to \infty$, in which the natural execution time emerges as a result of the analysis, but for now we consider it as an exogeneously imposed horizon. It is a rather surprising fact that in the absence of serial correlation in the asset price increments, the optimal strategy may be determined "statically" at the start of trading. Unless market parameters change, observation of price motions in the course of trading do not convey any information which would lead us to change the strategy.

We begin by construcing a discrete time model. Thus, for a given trading interval $\tau > 0$, write $t_k = k\tau$ for $k = 0, \ldots, N$ with $N = T/\tau$, and let x_k be our holdings at time t_k, with $x_0 = X$ and $x_N = 0$. Our sales between time t_{k-1} and t_k are $n_k = x_{k-1} - x_k$ corresponding to velocity (shares per unit time) $v_k = n_k/\tau$. Thus

$$x_k \;=\; X - \sum_{j=1}^{k} n_j \;=\; \sum_{j=k+1}^{N} n_j, \qquad k = 0, \ldots, N. \tag{1}$$

In the discrete-time model we do not assume that shares are traded at a uniform rate *within* each interval. Rather, we assume that a trader achieves the optimal execution possible, subject to the constraint that n_k shares are to be traded in the next time interval τ. The functions introduced below are a model to describe the results of the trader's best efforts.

In a standard manner (Stoll (1989)), we divide impact into a permanent and a temporary component. Thus we denote by S_k the price per share of the asset that is publically available in the market. This price satisfies the arithmetic random walk

$$\begin{aligned} S_k &= S_{k-1} + \sigma\tau^{1/2}\xi_k - \tau\, g\Big(\frac{n_k}{\tau}\Big) \\ &= S_0 + \sigma\tau^{1/2}\sum_{j=1}^{k}\xi_j - \tau\sum_{j=1}^{k} g\big(v_j\big) \end{aligned} \tag{2}$$

where the ξ_j are independent random variables of zero mean and unit variance, σ is *absolute* (not percentage) volatility, and $g(v)$ is a "permanent impact function," representing the effect on share price of the information conveyed by our trade. This effect is generally small, and below we will take $g(v)$ to be a linear function, in which case it will have no effect on determining the optimal strategy.

The price that we actually get on the kth trade is

$$\tilde{S}_k = S_{k-1} - h\Big(\frac{n_k}{\tau}\Big) + \tau^{-1/2} f\Big(\frac{n_k}{\tau}\Big)\,\tilde{\xi}_k, \qquad k = 1,\ldots,N. \tag{3}$$

Here $h(v)$ is a nonlinear "temporary impact function," representing the expected price concession we must accept in order to trade $v\tau$ shares in time τ. The random variables $\tilde{\xi}_j$ are independent of each other and of the ξ_j, with zero mean and unit variance. The new function $f(v)$ represents the uncertainty of trade execution as a function of block size (see Figure 1).

The factor $\tau^{-1/2}$ in the last term of (3) simply represents a scaling of the parameters, if τ is fixed and finite. When τ varies, for example, as the continuous-time limit $\tau \to 0$ is taken, this factor is necessary in order to preserve the effect of trading-enhanced risk. If it were not present, then breaking a block into several smaller blocks would diversify away the risk due to the uncertainty on each one, regardless of the form of this risk.

The "capture" of the trade program is the total cash received:

$$\begin{aligned} \sum_{k=1}^{N} n_k\tilde{S}_k = XS_0 + \sigma\tau^{1/2}\sum_{k=1}^{N-1} x_k\,\xi_k - \tau\sum_{k=1}^{N} x_k\, g\big(v_k\big) \\ - \tau\sum_{k=1}^{N} v_k\, h\big(v_k\big) + \tau^{1/2}\sum_{k=1}^{N} v_k\, f\big(v_k\big)\,\tilde{\xi}_k. \end{aligned}$$

We ignore discounting since we assume that the trade horizon is short. The "implementation cost" is $XS_0 - \sum n_k\tilde{S}_k$, a random variable due to the uncertainties in price movements and in realized prices.[4] Its expectation

[4]Note that implementation cost includes both the costs of finite liquidity and the price uncertainty due to delayed execution. This is the "implementation shortfall" of Perold (1988); also see Jones and Lipson (1999).

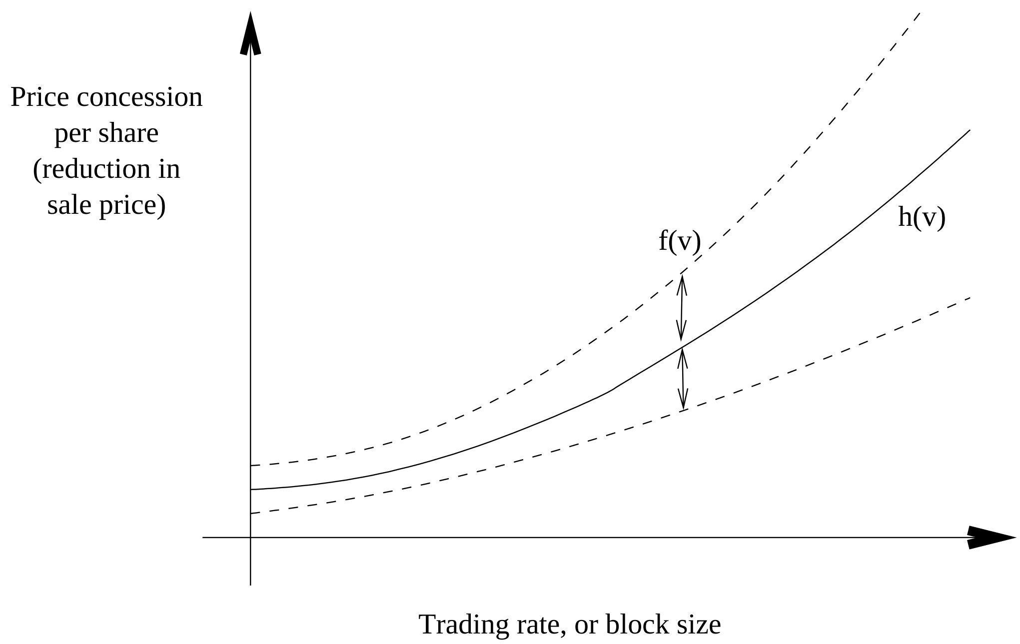

Figure 1: Temporary market impact cost, as a function of trading rate v in a continuous-time model, or of block size $v\tau$ in a discrete-time model. The solid curve is $h(v)$, the expected cost per share of demanding liquidity at the specified rate. The expected cost increases with trading rate, possibly according to a nonlinear model. The dashed curves are $h(v) \pm f(v)$ indicating one standard deviation on either side of the mean. These curves are an indication of the *uncertainty* in the realized liquidity premium, which also increases with liquidity demands. This uncertainty must be taken into account along with the volatility risk of holding the asset.

and variance at $t = 0$ depend on the free parameters $x_1, \ldots, x_{N-1}$ of the trade strategy:

$$\begin{aligned} E\big(x_1, \ldots, x_{N-1}\big) \;&=\; \sum_{k=1}^{N} x_k\, g(v_k)\, \tau \;+\; \sum_{k=1}^{N} v_k\, h(v_k)\, \tau \\ V\big(x_1, \ldots, x_{N-1}\big) \;&=\; \sum_{k=1}^{N} \sigma^2 x_k^2\, \tau \;+\; \sum_{k=1}^{N} v_k^2\, f(v_k)^2 \tau. \end{aligned}$$

A rational trader will construct his or her trade strategy in order to minimize some combination of E and V. As t advances, the values of E and V change, but if E and V are combined using a classic mean-variance approach, the optimal strategy continues to be the one determined initially (Almgren and Chriss 2000; Huberman and Stanzl 2001).

Now, for analytical convenience, we take the continuous-time limit $\tau \to 0$. The trade strategy x_k becomes a continuous path $x(t)$, and we assume that the block sizes n_k are well behaved so that $v_k \to v(k\tau)$, with $v(t) = -\dot{x}(t)$. The above expressions have finite limits

$$E[x] \;=\; \int_0^T \Big(x(t)\, g\big(v(t)\big) \;+\; v(t)\, h\big(v(t)\big) \Big)\, dt \tag{4}$$

$$V[x] \;=\; \int_0^T \Big(\sigma^2 x(t)^2 \;+\; v(t)^2 f\big(v(t)\big)^2 \Big)\, dt \tag{5}$$

where the square brackets indicate that these are "functionals" of the entire continuous-time path $x(t)$. We emphasize that the continuous limit is simply an analytical device for obtaining solutions when τ is reasonably small; in reality the discreteness of the trading intervals must be taken into account in order to correctly describe trading-enhanced risk (see the example in Section 4.3).

Introducing a risk-aversion parameter λ, we minimize the combined quantity

$$U[x] \;=\; E[x] \;+\; \lambda\, V[x].$$

Whether or not mean-variance optimization is appropriate in a particular case, λ may be considered a Lagrange multiplier for the constrained problem of minimizing E for a given V, and used to construct an efficient frontier in the space of trading trajectories (Almgren and Chriss 2000; Konishi and Makimoto 2001). More general weightings of risk, including Value-at-Risk, present thorny conceptual problems for time dependent strategies (Artzner, Delbaen, Eber, and Heath 1999; Basak and Shapiro 2001).

Minimizing $U[x]$ is a standard problem in the calculus of variations:

$$\min_{x(t)} \int_0^T F\big(x(t), -\dot{x}(t)\big)\, dt$$

with

$$F(x,v) \;=\; x\,g(v) \;+\; v\,h(v) \;+\; \lambda\sigma^2\, x^2 \;+\; \lambda\, v^2\, f(v)^2.$$

Stationarity to small perturbations requires that the optimal $x(t)$ solve the Euler-Lagrange equation

$$\begin{aligned} 0 \;&=\; F_x(x,-\dot{x}) + \frac{d}{dt} F_v(x,-\dot{x}) \\ &=\; F_x(x,-\dot{x}) + \dot{x} F_{xv}(x,-\dot{x}) - \ddot{x} F_{vv}(x,-\dot{x}), \end{aligned}$$

a second-order ordinary differential equation to be solved subject to the given endpoint values $x(0)$ and $x(T)$. Since F does not depend explicitly on t, we can multiply by $\dot{x}$ and integrate to obtain the first-order equation

$$F(x,-\dot{x}) \;+\; \dot{x} F_v(x,-\dot{x}) \;=\; \text{constant}.$$

In our case, we obtain

$$P(-\dot{x}) \;-\; P(v_0) \;=\; x\big(g(-\dot{x}) + \dot{x} g'(-\dot{x})\big) \;+\; \lambda\sigma^2 x^2, \tag{6}$$

with

$$P(v) \;=\; v^2 h'(v) + \lambda v^2 \Big(\, f(v)^2 \;+\; 2v f(v) f'(v) \Big) \;=\; v^2 \frac{d}{dv}\Big(h(v) + \lambda v f(v)^2\Big).$$

The constant of integration $v_0 = -\dot{x}|_{x=0}$ is the velocity with which $x(t)$ hits $x = 0$. For a sell program with $X > 0$, we have $v_0 \geq 0$, and conversely for a buy program. Note that $P(0) = 0$; we shall further assume that $P(v)$ is always an *increasing* function of v and hence invertible.

We now make two simplifying assumptions to obtain explicit solutions:

Permanent impact is linear in trading rate. A linear cost function $g(v) = \gamma v$ gives a total cost γX independent of the path $x(t)$. The first term on the right side of (6) vanishes, and then since $\dot{x}$ appears only on the left side and x itself appears only on the right, we can write the general solution in quadrature form as

$$\int_{x(t)}^{X} \frac{dx}{P^{-1}\big(\lambda\sigma^2 x^2 + P(v_0)\big)} \;=\; t. \tag{7}$$

The constant v_0 is to be chosen so that $x = 0$ corresponds to $t = T$. Note also that any constant in h disappears; the bid-ask spread doesn't affect our optimal strategy.

The imposed time horizon is infinite. Since $P(\cdot)$ is an increasing function, so is $P^{-1}(\cdot)$. It is thus clear that as v_0 decreases towards zero, the liquidation time T increases. If no time horizon is exogeneously imposed, then we obtain the longest possible liquidation time by setting $v_0 = 0$, which leads to the quadrature problem

$$\int_{x(t)}^{X} \frac{dx}{P^{-1}(\lambda\sigma^2 x^2)} \;=\; t.$$

We can often find analytic solutions to this problem when (7) with $v_0 \neq 0$ would be too intractable. These solutions will still give nearly complete liquidation in a finite time determined by market parameters.

3 Nonlinear cost functions

Restricting our attention to a sell program, with $v \geq 0$, we take the temporary impact functions

$$h(v) \;=\; \eta\, v^k \qquad f(v) \;=\; 0.$$

with $k > 0$ (for a buy program we would change signs in an obvious way). The linear case corresponds to $k = 1$. As noted above, we have neglected a possible constant in h corresponding to the bid-ask spread. Then

$$P(v) \;=\; \eta k\, v^{k+1}$$

which, for the general case of a finite time horizon with $v_0 \geq 0$, leads to the quadrature problem

$$\int_{x(t)}^{X} \left(\frac{\lambda\sigma^2}{k\eta}\, x^2 \;+\; v_0^{k+1} \right)^{-\frac{1}{k+1}} dx \;=\; t.$$

Taking $v_0 = 0$, we can solve explicitly for the longest optimal trajectories:

$$\frac{x(t)}{X} \;=\; \begin{cases} \left(1 + \dfrac{1-k}{1+k}\,\dfrac{t}{T_*}\right)^{-(1+k)/(1-k)} & \text{if } 0 < k < 1, \\ \exp\!\left(-\dfrac{t}{T_*}\right) & \text{if } k = 1, \\ \left(1 - \dfrac{k-1}{k+1}\,\dfrac{t}{T_*}\right)^{(k+1)/(k-1)} & \text{if } k > 1. \end{cases} \tag{8}$$

in which the "characteristic time" is

$$T_* \;=\; \left(\frac{k\eta X^{k-1}}{\lambda\sigma^2}\right)^{1/(k+1)}. \tag{9}$$

This is the analog of the "half-life" in the linear case. Only in the linear case $k=1$ is T_* independent of the initial portfolio size X. For $k \neq 1$, the characteristic time depends on the initial size as $T_* \sim X^{(k-1)/(k+1)}$:

- For $k<1$, rapid trading is *under*-penalized relative to the linear case (see Figure 3). As the portfolio size increases, volatility risk dominates trading cost, and the optimal trading time *decreases* since the exponent is negative.

- For $k>1$, rapid trading is *over*-penalized relative to the linear case. As the portfolio size increases, trading cost dominates volatility risk, and the optimal trading time *increases* since the exponent is positive. For example, if $k=3$ then $T_* \sim \sqrt{X}$.

As the portfolio size $x(t)$ decreases to zero, recalculation of the optimal trajectory would use a different starting value X, and hence give a different time T_*. The meaning of T_* is thus a little less fundamental than in the linear case. However, T_* scales in exactly the right way to make $x(t)$ still a static solution.

For more intuition, we note that the initial rate of selling is $-\dot{x}(0) = X/T_*$, and T_* is the solution to the relation

$$\lambda\,\sigma^2 X^2 T \;=\; k\,\eta\left(\frac{X}{T}\right)^k X.$$

The left side is the risk penalty associated with holding X shares for time T, and the right side, up to a factor of k, is $X\,h(X/T)$, the impact cost associated with selling X shares within time T (without the constant term representing the bid-ask spread, which does not affect the optimal solution).

For $k>1$, the trajectory reaches $x=0$ with $v=0$ at a finite time

$$T_{\max} \;=\; \frac{k+1}{k-1}\,T_*.$$

Thus these trajectories are the solution for finite imposed time T, if $T > T_{\max}$: the trajectory reaches zero at $T_{\max}$ and stays there until T.

Figure 2 shows the optimal trajectories of (8). The form of the solution is independent of the particular choice of time scale T_* and initial portfolio size X; these solutions may easily be scaled to any particular case.

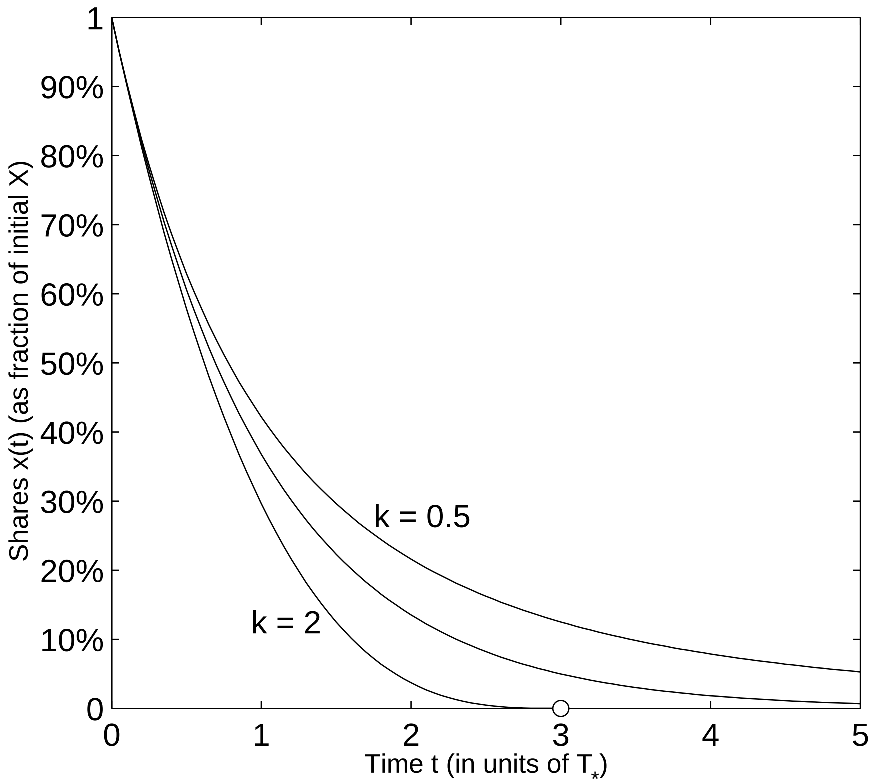

Figure 2: Optimal solution trajectories $x(t)$, for $k = \frac{1}{2}, 1, 2$. For each value of k, the solution has a "universal" form, which may be scaled as necessary for different choice of time scales T_* and initial portfolio size X. The disk shows $T_{\max} = 3T_*$ for $k = 2$.

A sense of the differences between these solutions may be gained by noting that for short times, all the optimal trajectories are fairly close to each other; but the "tail" of the trajectory is extended for small values of k, which strongly penalize trading at slow rates. For example, at $t = T_*$, the optimal trajectories reduce the holdings to 30%, 37%, and 42% of the initial portfolio, for $k = 2, 1, \frac{1}{2}$ respectively. At $t/T_* = 3$, the trajectory for $k = 2$ has reached $x = 0$ and remains there, the trajectory for $k = 1$ retains 5% of its initial holdings, and the trajectory for $k = 2$ retains 12.5% of the initial. The relative differences become even more pronounced as time continues.

3.1 Objective function

We can explicitly compute $E[x]$ and $V[x]$ for these solutions from (4,5), and hence draw the frontier. In doing this, we neglect the contribution from $g(v)$ and the term ϵX in $E[x]$. Then for general k, we have

$$E(\lambda) = \frac{k+1}{3k+1} \cdot \eta \left(\frac{X}{T_*}\right)^{k+1} T_* = \frac{k+1}{3k+1} \left(\frac{\eta\sigma^{2k} X^{3k+1}}{k^k}\,\lambda^k\right)^{1/(k+1)} \qquad (10)$$

$$V(\lambda) = \frac{k+1}{3k+1} \cdot \sigma^2 T_* X^2 = \frac{k+1}{3k+1} \left(\frac{k\eta\sigma^{2k} X^{3k+1}}{\lambda}\right)^{1/(k+1)} \qquad (11)$$

As λ varies, (V, E) moves along the hyperboloid-like curve

$$E \cdot V^k = \left(\frac{k+1}{3k+1}\right)^{k+1} \eta\,\sigma^{2k} X^{3k+1}.$$

For any positive λ there is a unique solution. As $\lambda \to 0$, we have $T_* \to \infty$, $E \to 0$, and $V \to \infty$; optimizing expected cost without regard to variance leads us to use all available time. As $\lambda \to \infty$, we have $T_* \to 0$, $E \to \infty$, and $V \to 0$; we minimize uncertainty regardless of the cost.

3.2 Example

We estimate the parameters in the following way. We choose a representative level of trading rate v_{ref}. If a specific time period τ is chosen, then v_{ref} is equivalent to a certain block size $n_{\mathrm{ref}} = \tau v_{\mathrm{ref}}$ traded in that time period; it may be interpreted as the market "depth" in the sense of Kyle (1985) or Bondarenko (2001). For our examples, we will consider a stock which trades one million shares per day, and we will take v_{ref} to be 10% of that rate, or v_{ref} =100,000 shares/day. For time period $\tau = 1$ hour, with 6.5 periods per day, this rate is equivalent to trading a block of approximately 15,300 shares in each hour.

Next, we choose the price impact h_{ref} which would be incurred by steady trading at the reference rate v_{ref}. In our example, we shall assume the share price is \$50/share, and we assume that trading v_{ref} =100,000 shares/day incurs a price impact of 1%, or \$0.50 /share.

Finally, we choose a value for the exponent k which best fits our belief about how the price impact would depend on trading rate for rates larger or smaller than v_{ref}. The choice $k = 1$ corresponds to linear dependence of price impact on rate; $k > 1$ means that *large* trading rates or block sizes have a disproportionately *large* effect on price, while $k < 1$ means that large trading rates or block sizes have a relatively *smaller* impact.

The impact model is then written

$$h(v) \;=\; h_{\mathrm{ref}} \left(\frac{v}{v_{\mathrm{ref}}} \right)^{k}, \qquad \text{or} \quad \eta \;=\; \frac{h_{\mathrm{ref}}}{v_{\mathrm{ref}}^{k}}. \tag{12}$$

Figure 3 illustrates this model.

We also suppose that our stock has annual volatility of 32%, for expected daily price changes of $\sigma = \$1/\text{share} \cdot \sqrt{\text{day}}$. We shall consider a portfolio of initial size $X = 100,000$, equal to one-tenth of the daily volume. We have then specified enough information to construct the efficient frontier (10,11) from $\big(E(\lambda), V(\lambda)\big)$ for any chosen k, describing the family of optimal solutions as the risk-aversion parameter λ ranges over all possible values $0 < \lambda < \infty$. To construct particular optimal solutions, we need to specify a value for λ.

The results are shown numerically in Table 1. For any value of k, the natural liquidation time T_* increases with the "risk-tolerance" parameter $1/\lambda$; as both increase, the expected cost decreases and the variance increases.

The dependence on exponent k, for fixed h_{ref} and v_{ref}, is shown in Figure 4. For large values of λ, the optimal trajectories all execute rapidly, to reduce the volatility risk associated with holding the portfolio. When the trading rate is larger than v_{ref}, costs *increase* with increasing k, so larger k leads to slightly slower trading.

Conversely, for small λ, trading generally proceeds *more slowly* than v_{ref} in order to minimize total expected cost. In this regime, *smaller* k is more expensive, and leads to relatively slower trading. In the intermediate parameter regime, the trajectories cross over from one behavior to the other: larger k suggests slower trading at the beginning when the rate is large, then relatively more rapid trading in the tail.

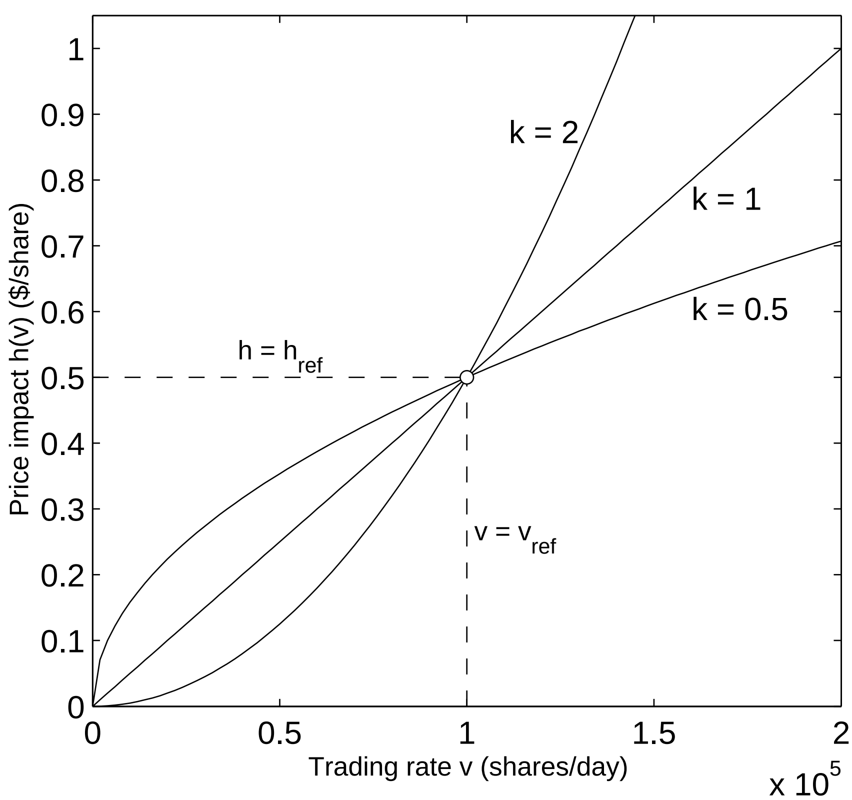

Figure 3: Nonlinear temporary impact model $h(v)$. A representative trading rate v_{ref} and a representative price impact h_{ref} are chosen, and these values are extended to general v using an exponent k.

Note that we may rewrite (9) as

$$T_* \;=\; \left(\frac{k h_{\mathrm{ref}}}{\lambda\sigma^2 X}\right)^{1/(k+1)} \left(\frac{X}{v_{\mathrm{ref}}}\right)^{k/(k+1)}$$

from which it is clear that $T_* \to X/v_{\mathrm{ref}}$ as $k \to \infty$, regardless of the values of the other parameters. In this limit, trading more rapidly than the reference rate is very strongly penalized, while trading more slowly is almost without cost, so the optimal strategy is always to trade exactly at the critical rate.

Finally, since λ is a difficult parameter to select in practice, let us note that it may be estimated if a time scale T_* is chosen: from (9,12) we find

$$\lambda = k \cdot \frac{h_{\mathrm{ref}} \left(\dfrac{X/T_*}{v_{\mathrm{ref}}}\right)^k \cdot X}{\sigma^2 T_* X^2}.$$

The numerator is the price concession per share for trading at a constant rate X/T_*, multiplied by the total number of shares X to get total expected cost; the denominator is the variance that would be incurred by holding X shares for time T_*. We multiply this ratio by k to correct for the nonlinearities which are ignored in this simple description.

4 Trading-enhanced risk

Now we take (for a sell program with $v \ge 0$)

$$h(v) \;=\; \eta\, v, \qquad f(v) \;=\; \alpha \;+\; \beta\, v.$$

The deterministic part of the temporary impact is the linear case $k = 1$ of the previous section.

The constant term in $f(v)$, with coefficient α, represents a constant uncertainty in the realized sale price, independent of our rate of selling and of the underlying price process. We minimize the total risk associated with this term by splitting our sale into as many equal parts as possible; thus this term pushes us toward the linear trajectory. The linear term, with coefficient β, represents the increase in variance caused by nonzero amounts of selling. This term even more strongly pushes us toward the linear trajectory.

Then (with $\dot{x} = -v \le 0$) the ODE (6) has

$$P(v) \;=\; \left(\eta + \lambda\alpha^2\right) v^2 \;+\; 4\lambda\alpha\beta\, v^3 \;+\; 3\lambda\beta^2\, v^4.$$

		$k = \frac{1}{2}$	$k = 1$	$k = 2$
	$1/\lambda = 1$	0.02	0.07	0.22
	10	0.09	0.22	0.46
T_*	100	0.40	0.71	1.00
	1,000	1.84	2.24	2.15
	10,000	8.55	7.07	4.64
	$1/\lambda = 1$	221	354	462
	10	103	112	99
E	100	48	35	21
	1,000	22	11	5
	10,000	10	4	1
	$1/\lambda = 1$	11	19	30
	10	23	33	45
$\sqrt{V}$	100	49	59	65
	1,000	105	106	96
	10,000	226	188	141

Table 1: Optimal time scale T_*, expected cost E, and standard deviation of cost $\sqrt{V}$, as functions of risk-tolerance parameter $1/\lambda$ and temporary impact exponent k. Market and portfolio parameters are as given in the text (initial portfolio value is \$5 million). As k is varied, the reference values h_{ref} and v_{ref} are held constant; thus the coefficient η varies as in (12). Time T_* is measured in days; $1/\lambda$, E, and $\sqrt{V}$ are in thousands of dollars.

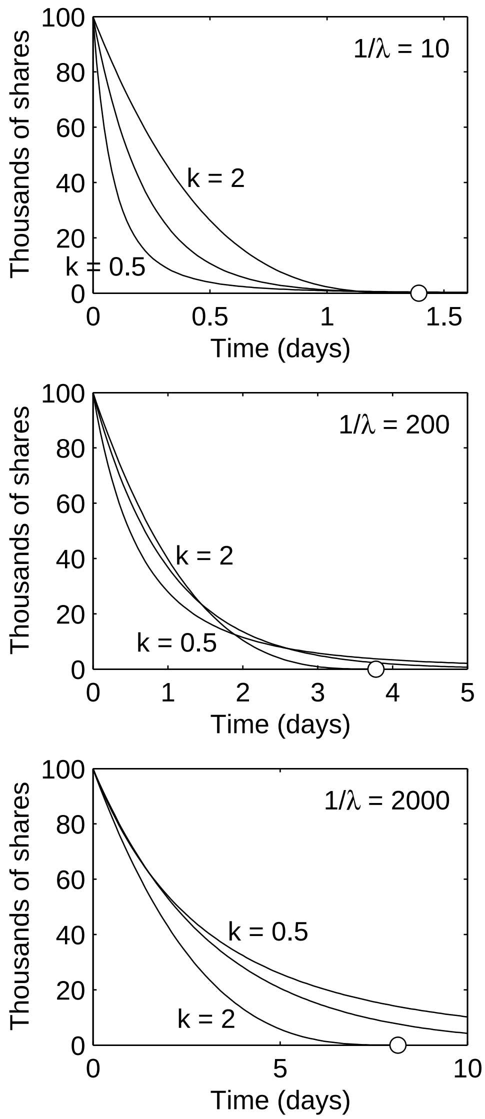

Figure 4: Optimal trajectories for different values of parameters as shown in Table 1 and in the text. For each value of the risk-tolerance parameter $1/\lambda$, we vary the exponent k for fixed values of h_{ref} and v_{ref}. The direction in which the time scale changes with k depends on the value of $1/\lambda$.

The polynomial $P(v)$ has $P(0) = 0$ and is increasing for $v \geq 0$, so the graph of the trajectory is always convex, and the inverse function P^{-1} is well defined. (For a buy program, with $\dot{x} \geq 0$, the sign of the odd term in $P(v)$ is reversed.) Since $P(v) \sim \mathcal{O}(v^2)$ for v near zero, the integrand appearing in the quadrature formulation (7) behaves like $\mathcal{O}(x^{-1})$ as $x \to 0$ for $v_0 = 0$, and there is no "hard" maximum time as we found above for $k > 1$.

4.1 Constant enhanced risk

To obtain analytical solutions, we consider two special cases. The first is $\beta = 0$; with this assumption, the price uncertainty on each trade is independent of the size of the trade. We then find the solution for $v_0 = 0$

$$x(t) \;=\; X \,\exp\!\left(-\frac{t}{T_*}\right), \qquad T_* \;=\; \sqrt{\frac{\eta + \lambda\alpha^2}{\lambda\sigma^2}}.$$

This is a pure exponential solution, except that the time constant has been increased by adding the additional variance per transaction to the impact coefficient: $\eta \mapsto \eta + \lambda\alpha^2$. The value functions are

$$\begin{aligned} E(\lambda) &= \frac{1}{2}\eta\,\frac{X^2}{T_*} && = \frac{1}{2}X^2\sqrt{\frac{\lambda\eta^2\sigma^2}{\eta+\lambda\alpha^2}} \\ V(\lambda) &= \frac{1}{2}X^2\sigma^2 T_*\left(1 + \frac{\alpha^2}{\sigma^2 T_*^2}\right) && = \frac{1}{2}X^2\,\frac{\sigma}{\sqrt{\lambda}}\cdot\frac{\eta + 2\lambda\alpha^2}{\sqrt{\eta+\lambda\alpha^2}}. \end{aligned}$$

The optimal value functions change in a more complicated way than the trajectory. As $\lambda \to 0$ we find the same behavior as in Section 3.1: $E \to 0$ and $V \to \infty$, since we do not care about the enhanced risk. In contrast, as $\lambda \to \infty$, all quantities have finite limits: $T_* \to \alpha/\sigma$, $E \to \frac{1}{2}\eta X^2/T_*$, and $V \to \alpha\sigma X^2$. Since trading itself introduces variance, risk-aversion and cost reduction both encourage you to spread the trading over several periods; the minimum-variance solution takes finite time and has finite cost.

4.2 Linear enhanced risk

Also, let us consider the special case $\alpha = 0$, so that $P(v) = \eta v^2 + 3\lambda\beta^2 v^4$, and we readily find

$$P^{-1}(w) \;=\; \sqrt{\frac{\sqrt{\eta^2 + 12\lambda\beta^2 w} \;-\; \eta}{6\lambda\beta^2}}.$$

We can then integrate to obtain

$$\frac{t}{T_*} = F\left(\frac{X}{X_*}\right) - F\left(\frac{x}{X_*}\right) \tag{13}$$

in which the characteristic time and share level are

$$T_* = \sqrt{\frac{\eta}{\lambda\sigma^2}}, \qquad X_* = \frac{1}{\sqrt{3}}\,\frac{\eta}{\lambda\sigma\beta} = \frac{1}{\sqrt{3}}\,\frac{\sigma T_*^2}{\beta},$$

and the nonlinear function is

$$F(u) = 2z - \coth^{-1} z, \qquad z = \sqrt{\frac{1}{2}\left(1 + \sqrt{1+4u^2}\right)}.$$

The characteristic time is the same as in Section 3 for $k = 1$, and does not depend on the new coefficient β. To understand the characteristic price level X_*, note that

$$\sqrt{3}\,\beta\,\frac{X_*}{T_*}\cdot T_*^{-1/2} = \sigma\sqrt{T_*}.$$

In this expression, the left side is the trading-induced variance in share price given by the model (3), if an initial portfolio of size X_* were sold in a single period of length T_* . The right side is the variance in share price due to volatility on the same time interval; at the characteristic share level, these two quantities are of comparable size.

To compare with previous results, we note that

$$F(u) \sim \log u + \text{constant} + \mathcal{O}(u^2), \qquad u \to 0. \tag{14}$$

If this limit is attained by taking a limit of the *parameters* so that $X_*/X \to \infty$, then (14) is valid uniformly over x (since $0 \le x \le X$) and we have the pure exponential solution

$$\frac{t}{T_*} \sim \log\frac{X}{x(t)} + \mathcal{O}\left(\left(\frac{\lambda\sigma\beta}{\eta}\right)^2\right), \qquad \frac{\lambda\sigma\beta}{\eta} \to 0, \tag{15}$$

which in particular recovers the result of Section 3 with $k = 1$ in the limit $\beta \to 0$. And for any fixed values of the parameters, (14) describes the tail of the solution as $x \to 0$; the time constant of the decay is not affected by the addition of β.

For $x \gg X_*$, *i.e.* the initial behavior when $X \gg X_*$, we use the expansion

$$F(u) \sim 2\sqrt{u} - \mathcal{O}(u^{-1/2}), \qquad u \to \infty,$$

which gives

$$x(t) \sim X_* \left(C - \frac{1}{2}\frac{t}{T_*} \right)^2, \qquad x \gg X_*, \tag{16}$$

with $C = \frac{1}{2}F(1)$. This is the same solution constructed in Section 3 for $k = 3$, with $\eta = \lambda\beta^2$.

The solution (13), together with the asymptotic expressions (15,16) are shown in Figure 5.

Thus the optimal strategy is as follows. Assuming $X > X_*$, trade initially using the trajectories of Section 3 with $k = 3$ and $\eta = \lambda\beta^2$. That is, volatility due to trading completely dominates the intrinsic volatility σ. As $x(t)$ reaches the level X_*, switch to the optimal solution in the linear case $k = 1$, with other parameters taking their market values. In the tail, trading-enhanced risk is a negligible effect compared to volatility.

4.3 Example

We focus on the case in the previous section, in which $\alpha = 0$ and $\beta \neq 0$ so that trading-enhanced risk increases linearly with block size with no constant term. To estimate the coefficients, we must return to our discrete-time model. With $h(v) = \eta v$ and $f(v) = \beta v$, the price model (3) becomes

$$\tilde{S}_k = S_{k-1} - \eta\, n_k\, \tau^{-1} + \beta\, n_k\, \tau^{-3/2}\, \tilde{\xi}_k.$$

Let us assume that for a particular choice of the trading interval τ, the standard deviation of price concession associated with trading-enhanced risk is a fraction ρ of the deterministic impact, since both of these quantities are linearly proportional to the block size. That is, we assume

$$\beta\, n_k\, \tau^{-3/2} = \rho \cdot \eta\, n_k\, \tau^{-1}$$

or

$$\beta = \rho\, \tau^{1/2} \eta$$

which gives

$$X_* = \frac{1}{\sqrt{3}\,\rho} \cdot \frac{1}{\lambda\, \sigma \tau^{1/2}}.$$

At this portfolio size, the volatility risk of holding the portfolio roughly balances the trading-induced risk of selling along the optimal trajectory. Although both of these quantities are risks, the expression for X_* involves λ through its influence on the trading time T_*.

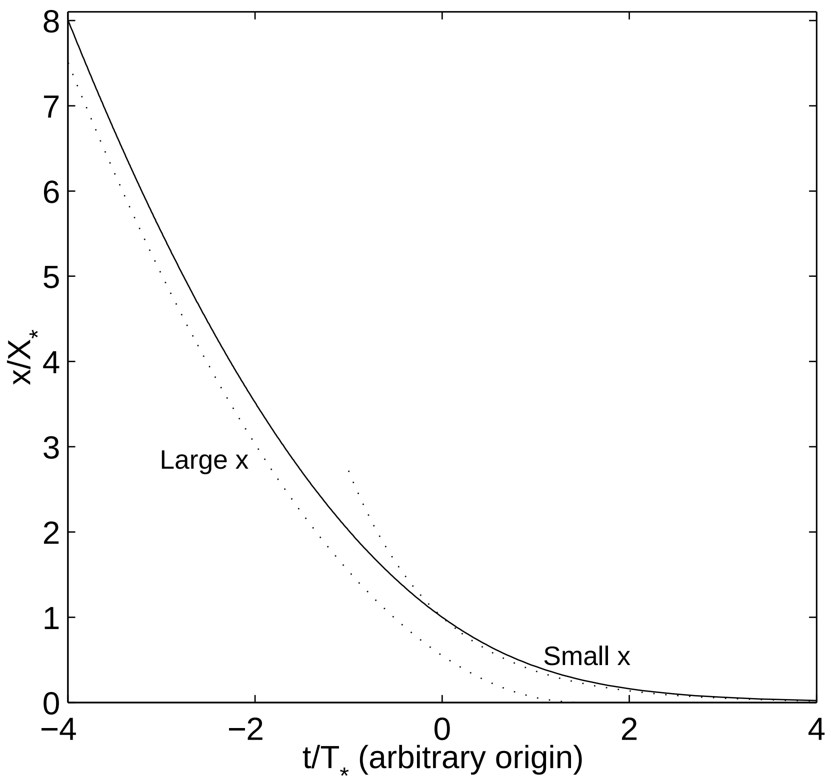

Figure 5: The universal optimal trajectory for linear trading-enhanced risk, scaled by the characteristic time T_* and share level X_*, together with its asymptotic behavior for large and small x. The solution for any desired parameters may be constructed by scaling this picture and choosing an appropriate time origin. When x is large compared with X_*, the solution is essentially the nonlinear solution of Section 3 with exponent $k = 3$; this happens for the initial phase of trading if X is large compared with X_*. When x is small compared with X_*, the solution is essentially the linear solution of Section 3 with $k = 1$; this applies to the tail of the trajectory for any X, or to the entire trading trajectory if X is smaller than X_*.

Let us take market parameters as in Section 3.2, with $1/\lambda =$\$10,000, corresponding to the first panel of Figure 4 with liquidation in less than one day. Let us divide trading into one-hour time intervals, so $\tau = (2/13)$ day, and let us take $\rho = 1/2$. We then obtain $\beta = 10^{-6}\,\$ \cdot \mathrm{day}^{3/2}/\mathrm{share}^2$, and $X_* = 30{,}000$ shares corresponding to a portfolio size of \$1.5M. A liquidation problem with initial value greater than this will begin in the large-x regime where trading-enhanced risk is dominant, and end in the small-x regime where it is negligible.

5 Conclusions

We have obtained explicit analytical solutions for certain special cases of the impact model. First, we neglected the effect of trading-enhanced risk, and took the impact function to be a simple power law. The solutions in this case are a straightforward nonlinear extension of the results in Almgren and Chriss (2000); the exponential solutions obtained there are a particular dividing case of these power-law solutions.

With trading-enhanced risk, we considered two particular cases with linear impact functions. If the price uncertainty per transaction is independent of transaction size, then the optimal trajectories are given by the previous results, simply augmenting the impact coefficient by the additional variance. A risk-averse trader lengthens his trade program, diversifying away some variance by spreading the execution over more different transactions at the expense of slightly higher volatility risk.

If the price uncertainty per transaction is linearly proportional to transaction size, then a characteristic portfolio size emerges, above which reduction of this added variance is the dominant effect. In this regime, trade trajectories are equivalent to the previous power-law solutions with exponent equal to three. For portfolios smaller than this size, the new effect may be neglected relative to deterministic impact costs and ordinary volatility.

Throughout this paper, we have focused on obtaining explicit solutions for the sake of analytical insight. Numerical solution would be quite straightforward, and would allow lifting the restrictions described above and the consideration of a more general class of models.

Portfolios of multiple assets are an interesting extension. Already in the linear case (Almgren and Chriss 2000), to obtain explicit solutions it is necessary to make simplifying assumptions about cross-impacts, for example, that trading in each asset affects only the price of that asset. Even with that assumption, the nonlinear formulation opens a wide class of possible models: for example, should the exponent be the same for each asset? Determination and characterization of the optimal trajectories in this case is a topic for future work.

References

Almgren, R. and N. Chriss (1999). Value under liquidation. *Risk 12*(12), 61–63.

Almgren, R. and N. Chriss (2000). Optimal execution of portfolio transactions. *J. Risk 3*(2), 5–39.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Math. Finance 9*, 203–228.

Barra (1997). *Market Impact Model Handbook*.

Basak, S. and A. Shapiro (2001). Value-at-Risk-based risk management: Optimal policies and asset prices. *Rev. Financial Studies 14*, 371–405.

Bessembinder, H. and H. M. Kaufman (1997). A comparison of trade execution costs for NYSE and NASDAQ-listed stocks. *J. Fin. Quant. Anal. 32*, 287–310.

Bondarenko, O. (2001). Competing market makers, liquidity provision, and bid-ask spreads. *J. Financial Markets 4*(3), 269–308.

Chakravarty, S. (2001). Steath-trading: Which traders' trades move prices? *J. Financial Econ. 61*, 289–307.

Chan, L. K. C. and J. Lakonishok (1993). Institutional trades and intraday stock price behavior. *J. Financial Econ. 33*, 173–199.

Chan, L. K. C. and J. Lakonishok (1995). The behavior of stock prices around institutional trades. *J. Finance 50*, 1147–1174.

Chordia, T., A. Subrahmanyam, and V. R. Anshuman (2001). Trading activity and expected stock returns. *J. Financial Econ. 59*, 3–32.

Grinold, R. C. and R. N. Kahn (1999). *Active Portfolio Management* (2nd ed.), Chapter 16, pp. 473–475. McGraw-Hill.

Hasbrouck, J. and R. A. Schwartz (1988). Liquidity and execution costs in equity markets. *J. Portfolio Management 14*(Spring), 10–16.

Hasbrouck, J. and D. J. Seppi (2001). Common factors in prices, order flows, and liquidity. *J. Financial Econ. 59*, 383–411.

Holthausen, R. W., R. W. Leftwich, and D. Mayers (1987). The effect of large block transactions on security prices: A cross-sectional analysis. *J. Financial Econ. 19*, 237–267.

Holthausen, R. W., R. W. Leftwich, and D. Mayers (1990). Large-block transactions, the speed of response, and temporary and permanent stock-price effects. *J. Financial Econ. 26*, 71–95.

Huang, R. D. and H. R. Stoll (1997). The components of the bid-ask spread: A general approach. *Rev. Financial Studies 10*(4), 995–1034.

Huberman, G. and W. Stanzl (2001). Optimal liquidity trading. Preprint.

Jones, C. M. and M. L. Lipson (1999). Execution costs of institutional equity orders. *J. Financial Intermediation 8*, 123–140.

Kahn, R. N. (1993). How the execution of trades is best operationalized. In K. F. Sherrerd (Ed.), *Execution Techniques, True Trading Costs, and the Microstructure of Markets*. AIMR.

Keim, D. B. and A. Madhavan (1995). Anatomy of the trading process: Empirical evidence on the behavior of institutional traders. *J. Financial Econ. 37*, 371–398.

Keim, D. B. and A. Madhavan (1997). Transactions costs and investment style: An inter-exchange analysis of institutional equity trades. *J. Financial Econ. 46*, 265–292.

Konishi, H. and N. Makimoto (2001). Optimal slice of a block trade. Preprint.

Koski, J. L. and R. Michaely (2000). Prices, liquidity, and the information content of trades. *Rev. Financial Studies 13*, 659–696.

Kraus, A. and H. R. Stoll (1972). Price impacts of block trading on the New York Stock Exchange. *J. Finance 27*, 569–588.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*, 1315–1336.

Loeb, T. F. (1983). Trading costs: The critical link between investment information and results. *Financial Analysts Journal 39*, 39–44.

Perold, A. F. (1988). The implementation shortfall: Paper versus reality. *J. Portfolio Management 14*(Spring), 4–9.

Perold, A. F. and R. S. Salomon, Jr. (1991). The right amount of assets under management. *Financial Analysts J. 47*(May-June), 31–39.

Rickard, J. T. and N. G. Torre (1999). Information systems for optimal transaction implementation. *J. Management Information Systems 16*, 47–62.

Stoll, H. R. (1989). Inferring the components of the bid-ask spread: Theory and empirical tests. *J. Finance 44*, 115–134.

Wagner, W. H. and M. Banks (1992). Increasing portfolio effectiveness via transaction cost management. *J. Portfolio Management 19*, 6–11.