FINANCIAL SERVICES REVIEW

Financial Services Review 11 (2002) 65–84

How Quickly Should You Liquidate Your Vested Stock?

Robert Dubil*

Finance Department U-1041F, University of Connecticut, 2100 Hillside Rd., Storrs, CT 06269, USA

Accepted 24 January 2002

Abstract

I model the optimal behavior of an individual or trustee who decides to liquidate a position in an asset. As his holdings are large, his own sales may have adverse permanent and temporary impact on the realized return. To minimize this effect, a wealthy insider may choose to liquidate slowly. At the optimum, he balances the exposure to the return variance against sale-induced price concessions by using a Value-at-Risk-inspired risk-reward function as his decision tool. The prescriptive appeal of the model is demonstrated numerically. For example, I show that a risk-averse individual holding $1 million of his wealth in a stock, subject to 50% volatility and linear temporary impact of $1,900, should optimally take 10 days to liquidate. © 2002 Academy of Financial Services. All rights reserved.

JEL classification: C0; D1; D4; G1; G2

Keywords: Liquidity; Market impact; Trading strategy; Value at risk

1. Introduction

Personal finance researchers devote most of their time to the examination of the most "efficient" ways of accumulating wealth. They examine on one hand the investment vehicles themselves (asset classes, asset allocation strategies, performance measures, laws governing retirement saving) and on the other the investor's personal circumstance (age, tax and estate situation, risk tolerance, etc.). They also consider the best ways of disposing of the wealth (tax efficient income generation, estate management, actuarial life expectancy). They spend

* Tel: +1-860-871-8254; fax: +1-860-486-0634.
E-mail address: rdubil@sba.uconn.edu (R. Dubil).

precious little time on the most efficient strategies of converting the wealth, once it is acquired, into disposable cash! The subject of this paper is one such problem: that of a wealthy insider or an estate trust holding a significant share of a company's stock. In the past, the owners of privately held industrial empires hired managers for cash. Today, more and more technocrat-executives acquire significant stakes in the companies they run through various incentive compensation schemes. Apart from having undiversified market risks, they (or their estates) face liquidity risks when they come to sell their stock. Their behavior may be closely monitored by investors (e.g., SEC filings) or their stock is so thinly traded (e.g., for closely-held smaller public companies) that their sales have significant adverse impact on the realized value of their holdings. To formalize the problem, I draw on recent advances in the modeling of liquidity constraints for institutional traders. Specifically, I build on micro-structure models of dynamic optimization of a trade trajectory. I contribute a discussion of the risk-reward tradeoff of a selling individual. While relying on the fundamental micro-economic utility theory for help, at the same time, I borrow the tools from the banking literature on risk management. In particular, I adapt the Value-at-Risk (VaR) methodology to the specific problem under consideration, as the risk-reward tradeoff in the personal finance context is often more complex than simple expected profit maximization. See, for example, Chhabra and Zaharoff (2001).

Market microstructure research abounds in models explaining the size of the bid-ask spread in financial markets. Glosten and Harris (1988) decompose the spread of NYSE stocks into the inventory and asymmetric information parts. Brennan and Subrahmanyam (1996) estimate variable (proportional) and fixed illiquidity parameters from stock return data sorted by the Kyle measure of market depth λ. The goal of these studies is descriptive and not prescriptive, even though some, e.g., Kyle (1995), solve for the optimal trading trajectory of an agent who comes to market to trade. In the finance literature in general, there are very few theoretical attempts to explicitly focus on the latter. I build on the work of Bertsimas and Lo (1998), and Almgren and Chriss (2000) who consider an optimization problem of an agent liquidating a large block of securities under an exogenously assumed final time (similar to that of a proprietary trader). As the agent trades, he affects the market price. His impact on the market price can be temporary (disappears when the market absorbs the quantity he supplies) and permanent (persists due to its informational content). The models solve for the optimal sales trajectory (a sequence of sales quantities) over the total liquidation time. In this paper, I take the perspective of a wealthy insider or his estate trust. The insider does not face an exogenously imposed deadline by which to dispose of his holdings. In fact, he optimizes his total risk-reward function on the best total time in which to liquidate his position. Given his closely watched holdings, he decides how long to extend his liquidation to minimize the discount at which he sells and the risk of subjecting his holdings to the general market fluctuations. As he possesses no special information about the timing of the market information, he chooses a simple linear sales trajectory, i.e., spreads his trades evenly over the sales horizon. He trades off between the impact of his trades on the realized sale value and the risk of holding the assets for an extended period of time. The focus of this paper is on the concept of endogenous liquidity, as defined by Bangia et al. (1999). A large stakeholder who decides to sell a large position in the market, where "large" is defined as exceeding the standard quote depth, will adversely affect the price at which he transacts if he sells too quickly. This sale discount will be the price he will pay for avoiding the market risk of the position. At the optimal time to liquidate, the change in his marginal utility will be zero.

The choice of the utility function, which ultimately determines the behavior of the optimizing agent, can be subject to great debate. In this paper, the agent's risk-reward preferences are implied by the confidence interval parameter, α, he sets for his VaR function. By setting that parameter, the agent reveals the worst-loss tail probability he is willing to accept. This parameterization is motivated by two factors: simplicity and the notion of asymmetric risk (unlike with standard deviation or beta, here loss risk is worse than gain risk). Some admittedly unappealing aspects of the VaR utility in the general equilibrium context are discussed in Basak and Shapiro (2001). These are of little relevance to the specific case of an individual investor deciding on the best strategy of converting his liquidity-constrained wealth into cash. Only the form, not the essence of the results depends on the choice of the utility function. One could just as easily obtain closed-form solutions for a general HARA preferences class as defined by Huang and Litzenberger (1988).

The paper is organized as follows. First, I review in detail the institutional research examining the optimal block trading behavior. Next, I present a model of the optimal liquidation time for an individual holding a large stake in an asset. The model is recast from the intuitive terms of units and prices into the more general terms of dollar exposures and returns. I consider two parameterizations of the market impact functions: general power functions uncorrelated with the price, for some of which I obtain closed-form solutions, and linear functions correlated with the price process. The latter is in some sense more general, as it allows for the market impacts to change (i.e., increase) during large market dislocations (selling a large stake in a rapidly falling market). For both parameterizations, I show numerical examples and provide intuitions for them. I include the discussion of the alternatives for the risk-reward optimization functions.

2. Optimal liquidation trajectories in finance literature

As Chan and Lakonishok (1995) showed, a typical institutional investor's trade in the stock market is broken down into smaller packages and executed over a period of 4 days or more. Presumably, such protracted liquidation is designed to minimize the adverse impact on the overall transaction price, but exposes the trader to market risk. It is not hard to imagine that large individual investors follow similar strategies, the only differences being the transaction sizes and the length of the strategy interval.

Bertsimas and Lo (1998) consider an expected trading cost minimization for a trading program designed to acquire a fixed number of shares, X, by the final time, T. While the program is in effect, new information arrives in the market. It is in the form of random shocks to the trading price. However, the price is also affected by each trade executed in the program. The authors write the basic Bellman equation for this dynamic programming optimization and employ the optimal control machinery to solve recursively for the best trading trajectory. A trajectory is defined as a sequence of the amounts purchased in each of the N equally spaced time intervals. They show that the optimal strategies are often linear combinations of a "naïve" strategy, of breaking the total size X into N identical packages of

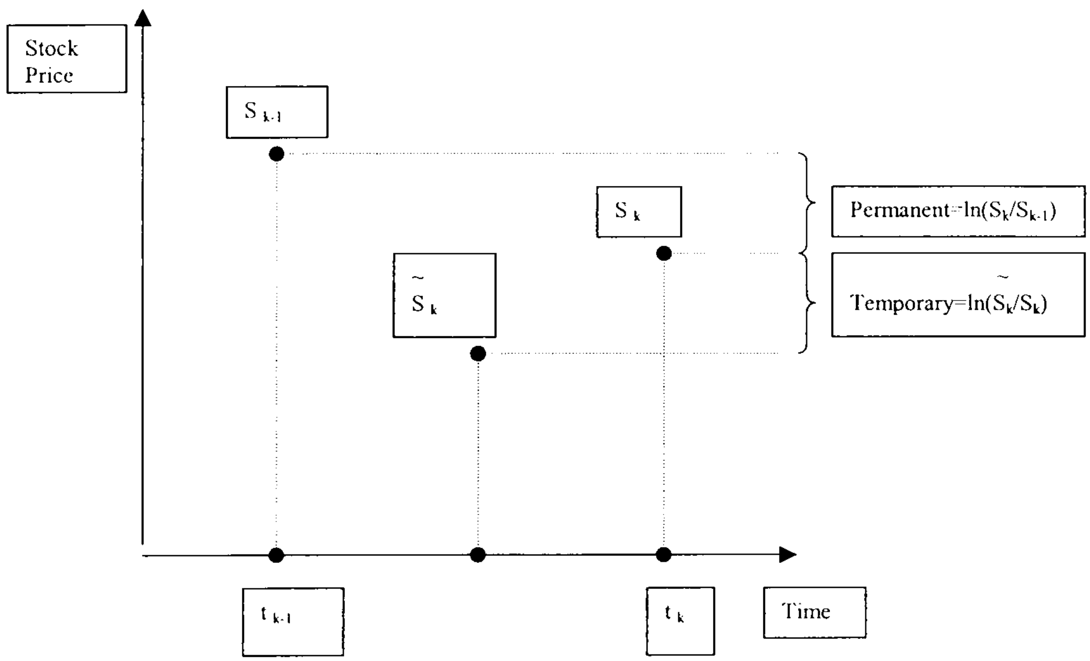

Fig. 1. Temporary and permanent price effects of a seller-initiated block.

size X/N, and a correction portion reflecting the new information. In the absence of private stock-specific information, although the naïve strategy is not optimal in general, it is the best under the assumption that the price follows an arithmetic random walk and the impact of the liquidation strategy is linear.

Almgren and Chriss (2000) follow the intra-period distinction between temporary and permanent impacts, first introduced by Holthausen et al. (1987, 1990) who estimate the impact of large block trades on the prices of NYSE stocks. The temporary price effect is defined as the return earned on the difference of the block transaction price and the equilibrium price prior to the block transaction. The permanent price effect is defined as the return on the difference between the equilibrium prices after and before the block transaction. For a seller-initiated block, these are depicted in Fig. 1, cf. Holthausen et al. (1987).

In Almgren and Chriss, the stock price is subject to both of these effects in each of the N equally spaced intervals (t_{k-1},t_k) of length $\tau = t_k - t_{k-1} = T/N$, $k = 1, \ldots, N$. A trader faces selling X shares over the fixed total time T through a sequence of sales in each of the N intervals. His holdings at the interval end points are $x_0 = X, x_1, x_2, \ldots, x_{N-1}, x_N = 0$, his sales during the intervals are $n_k = x_k - x_{k-1}$, $k = 1, \ldots, N$, and the speed with which he sells in each interval is denoted by $v_k = n_k/\tau$. The equilibrium price, S_k, follows an arithmetic random walk. It has volatility, σ, but no drift. The only drift is that due to a permanent impact effect, $g(v_k)$, resulting from the trader's action. The trading price, $\tilde{S}_k$, the trader faces in each interval, is subject to a temporary impact function $h(v_k)$. Instead of the expected cost minimization, the authors present a mean-variance approach in which the trader cares about the risk of the strategy (variance of the liquidation cost). He does it by minimizing the negative utility associated with the cost of the protracted liquidation over the time interval (0, T). That random cost, C, is defined as the difference between an instantaneous sale of all X units and the sum of the proceeds from a sequence of sales $\{n_k, k = 1, \ldots, N\}$ at prices $\{\tilde{S}_k, k = 1, \ldots, N\}$. It is a function of the impact parameters and the stock volatility. Because the trading price, $\tilde{S}_k$, is normally distributed, the mean, $E\,[C]$, and the variance, $V\,[C]$, of the cost can be derived. The trader minimizes the expected cost given a level of variance. This is equivalent to an unconstrained minimization of

$$\min_{\{x_k:k=1,\ldots,N\}} \{E[C] + \alpha V[C]\} \tag{1}$$

where the Lagrange multiplier α can be interpreted as the relative risk-aversion coefficient. When $\alpha > 0$, a unique solution $\{x_k\}^*$ is guaranteed by the strict concavity of the minimand. For a risk-neutral trader ($\alpha = 0$), the optimal trajectory is a straight line of declining holdings over time defined by the decrement $n_k = X/N$. Risk-averse traders follow convex lines below that line; risk-loving traders follow concave lines above the straight line. Risk-averse traders thus sell relatively more up front, and less later, incurring higher impact costs, to avoid the exposure to the random shocks. The mathematical details of the Almgren and Chriss set-up are contained in Appendix A.

The two highlighted papers differ in their choices of the optimized functions. Bertsimas and Lo (1998), assume that their agent is an expected cost minimizer. The expected cost may depend on the variance of the price process, but the agent does not care about the variance of the cost itself. The optimal control methodology allows for the agent to change his strategy during the execution of the program. Almgren and Chriss (2000) explicitly make their agent worry about risk by choosing only those solutions that lie on the mean-variance efficient frontier. Equivalently, they adopt a constant relative risk-aversion utility function. Only if the latter is of a particular form (e.g., log-utility) does the static up-front optimization yield the same solution as the less restrictive dynamic program. A VaR-derived utility does not guarantee that the static solution is dynamically optimal. It does avoid, however, the somewhat counterintuitive result whereby large and small baskets of the same securities are liquidated identically by a trader with a given risk-aversion coefficient. It also produces solutions closer to the straight line.

The use of the VaR-derived utility function in a partial equilibrium setting is examined by Basak and Shapiro (2001). Their agent maximizes the expected utility of his horizon wealth. He is constrained to maintain the probability of his wealth falling below some floor at a pre-specified level α. At the optimum, the VaR-managed agent chooses his portfolio allocation to incur losses in the worst states of the nature up to the probability α and leaves them uninsured. When he incurs large losses, they are greater than those for an unconstrained case. Furthermore, an agent with constant relative risk-aversion preferences engages in a "go-for-broke" behavior in the transitional states near the cut-off point. He invests a large proportion of his wealth in risky equities hoping for a favorable outcome at the horizon at the expense of potentially large losses in the worst states. An adoption of the expected loss conditional on reaching the α tail, as opposed to the loss at the α cut-off, is shown to remedy this pathological behavior. I will return to the discussion of the appropriate utility function in the next section.

Lastly, in order for a liquidity model to be of practical use, the market impact parameters of the temporary and permanent component have to be known. Fortunately, the market microstructure literature abounds in estimation specifications for organized (listed and OTC) stock markets. One such study is Holthausen et al. (1987, 1990). I refer the reader to O'Hara (1997) for other references.

3. The optimal liquidation horizon model

In general, a block seller faces a difficult problem of selecting simultaneously the best total time and the best liquidation sequence. To simplify that choice, in this paper's model, the individual investor solves for the optimal final horizon by adopting a simple strategy to liquidate at a constant rate per unit of time. His strategy trajectory (a plot of his remaining position against time) is a straight line. This is motivated by the fact that he has no special information about the timing or the "lumpiness" of trades coming to the market beyond knowing the market impact of his own actions. He chooses the simplest strategy to sell the same amount in each time interval, but is concerned about the total time of liquidation.

The model can be cast in two different ways. One is to retain the set-up of an agent optimally liquidating X number of shares under the assumption that the equilibrium stock price is locally an arithmetic Brownian motion with no drift. The other is to consider an agent optimally reducing, instead of the number of shares, his dollar exposure to the return on a stock under the assumption that the return, not the price, follows a locally arithmetic Brownian motion. The mathematics and the results for both formulations are essentially the same, although the parameter interpretations are slightly different. As the latter is more general allowing for assets other than stocks, I will adopt that version while retaining the notation of the set-up based on the number of shares. Note that the main differences between the two versions are the stochastic process assumption (normal in price vs. normal in returns) and the form of the impact functions (cost defined as a known function of returns instead of the price).

Within the model, I consider two different forms of the market impact formulations: a general power function, which is uncorrelated with the price process, with some of its special cases (e.g. linear, square root), and a stochastic linear impact function correlated with the price process. The latter case allows for a feedback loop, whereby a significant drop in returns (prices) can cause a deterioration of liquidity. The agent's objective will be to choose optimally the final time by which his holdings will be reduced to zero under the assumption that he sells at a constant speed over that horizon. His sales will cause a permanent market return (price) change at the "end" of each interval $(t - dt, t)$. They will also cause a temporary shock to the realized return (trading price) which will deviate from the general market return (price), during the interval $(t - dt, t)$, but will dissipate completely by the "end" of that interval. I will work in continuous time, but it is easy to show that the results are a natural limit of the discrete version in which the number of trading intervals within the fixed final horizon increases to infinity.

3.1. The set-up of the model

For mathematical convenience, I will adapt the notation to continuous time. Let W_t be the dollar amount of investment in the underlying asset at time t. The initial exposure to the asset return (initial holding) is denoted by W_0, and the final exposure is zero, i.e., $W_T = 0$. Let us normalize the agent's holdings at time t, as the ratio of his remaining dollar exposure to the original investment, i.e., let us define $x_t = \frac{W_t}{W_0}$. Thus the agent starts with holdings $x_0 = X = 1$ at time $t = 0$ and liquidates all of them by the final time T, ending up with $x_T = 0$. In each interval $(t - dt, t)$, the agent sells $n_t = x_{t-dt} - x_t = -dx_t$ of his holdings. The speed of trading in each interval is defined as $v_t = \frac{n_t}{dt} = -\frac{dx_t}{dt}$. The cumulative equilibrium return on the underlying asset R_t, representing all public information in the market, follows an arithmetic Brownian motion process[1]. The only drift is due to the accumulated permanent impact on the price from the sales executed by our agent. The equilibrium return, R_t, at time $t \in [0, T]$ can thus be written:

$$R_t = \sigma z_t - \int_0^t g(v_s)ds \tag{2}$$

where $g(v_t)$ is a function representing the permanent impact of the agent's sales on the equilibrium return, and z_t is the standard Wiener process. The cumulative trading return, $\tilde{R}_t$, realized by the agent through a trade within each interval $(t - dt, t)^2$, is also subject to the temporary impact $h(v_t)$ as a function of the speed of sales within the interval:

$$\tilde{R}_t = R_t - h(v_t) \tag{3}$$

Let us further make an assumption that the agent sells at a constant speed $v_t \equiv v$, so that

$$n_t = -dx_t = vdt \tag{4}$$

The excess profit (which is most likely negative, i.e., a cost), as a fraction of his original wealth, due to the non-instantaneous liquidation is defined as:

$$\prod = \int_0^T \tilde{R}_t(-dx_t) = \int_0^T \tilde{R}_t v dt \tag{5}$$

Note that, in the alternative formulation of the model, with x_t denoting the number of shares, S_t denoting the equilibrium stock price following

$$S_t = S_0 + \sigma z_t - \int_0^t g(v_s)ds \tag{2'}$$

and $\tilde{S}_t$ denoting the realized trading price, related to the equilibrium price through

$$\tilde{S}_t = S_t - h(v_t) \tag{3'}$$

the absolute dollar profit would be defined as

$$\prod = \int_0^T \tilde{S}_t(-dx_t) - XS_0 = \int_0^T \tilde{S}_t v dt - XS_0 \qquad (5')$$

Once an appropriate expression is substituted for $\tilde{S}_t$, the XS_0 term will cancel out, and all of the subsequent derivations will be identical to the return version of the model.

3.2. The VaR utility function

The next step is to choose the appropriate prism through which the individual evaluates the risk-return tradeoff. An expected profit maximizer faced with self-induced market impact would choose to liquidate very slowly (over an infinite horizon) in order to minimize the total return lost due to sales concessions. He would not care about risk. The only way for the model to introduce risk into the decision process is by adopting a concave utility function. In asset allocation decisions of personal finance, this is typically done by selecting the appropriate performance measure like the Sharpe ratio of expected return to the standard deviation, or a ratio involving only diversifiable risk (beta). These approaches have sound theoretical backing in the form of one-period asset pricing models like the CAPM. Here the liquidity risk is self induced and thus undiversifiable. The decision does not involve a selection of an investment alternative in an efficient market setting, but rather dealing with the inefficiency of excess inventory absorption. There are many candidates for the optimization function. One could be HARA utility, which would result in a mean-variance efficient frontier of the "best" combinations of the expectation and the variance of the realized return of the liquidation. The individual would then be required to have well-specified preferences over that mean-variance space. Although this is a viable option for general market risk, it is not an obvious alternative for dealing with liquidity risk.

Here, I adopt the VaR as a suitable risk-reward tradeoff framework. It has several appealing features from a perspective of an individual whose general wealth accumulation-disposal objective function is quite complex. It is expressed in dollars of potential loss as opposed to in statistical terms of variances, betas, etc. It also takes into account the fact that with undiversifiable risks, investors typically do not view upside volatility the same way as downside volatility. A wealthy individual who has already decided to sell his concentrated holdings worries a lot more about losing his accumulated wealth than about potential additional market gains. The VaR requires only one parameter. An individual investor liquidating his accumulated stock position would express his risk preferences by choosing the confidence level for a given level of loss he is willing to tolerate. For example, he could state: "With 90% probability (confidence level), I want to lose no more than $200,000 during the liquidation process". This would imply that he worries about the 10% left tail probability of a loss. The smaller the tail region about which he worries (i.e. the higher the confidence level), the more risk averse the individual is, as he is concerned with very low probability loss events. It is reasonable to assume that individuals would typically be more risk averse than institutions and would choose higher confidence levels. Regulated banks publish 95% and 99% VaR numbers[3]. It is fair to say that for most banks these numbers are not "binding" and do not reflect the banks' true risk aversion; they simply reflect the regulators' worry about catastrophic situations leading to systemic meltdowns. Internally, banks use specific position limits to govern their traders' risk-taking behavior. An individual or estate trustee who does not have agency benefits of a bank manager (plays with his own money or that of his beneficiaries) would most likely choose a very high VaR confidence (i.e., "worry") level. At his chosen level, a decision maker would then seek to minimize the corresponding loss cutoff by choosing the best sale strategy. Let us examine his optimal behavior using a formal definition.

The VaR of a set of positions is defined as a critical, typically negative, profit value, Π^*, for which the probability of the profit, Π, falling below that critical value, Π^*, is equal to a given quantity $\Phi(\alpha)$. The latter is equal to one minus the chosen confidence interval. In a typical institutional setting, the VaR is defined only implicitly through a multivariate numerical simulation. However, for a normally distributed single asset-dependent profit function, the VaR, Π^*, can be defined explicitly as:

$$\prod{}^* = E[\prod] - \alpha\sqrt{V[\prod]} \tag{6}$$

where α is a known number taken from the standard normal table corresponding to a chosen confidence level, e.g., equal to 1.645 for a 5% left-tail probability, and $E[\Pi]$ and $V[\Pi]$ denote the expected value and the variance of the random profit. The confidence level-related parameter α embodies an implicit set of risk preferences chosen by the agent. His objective is thus to choose an optimal final liquidation time to maximize the VaR of his profit:

$$\max_T \prod{}^* = \max_T \{E[\prod] - \alpha\sqrt{V[\prod]}\} \tag{7}$$

Note that Eq. (7) is analogous to Eq. (1) because $\Pi = -C$, but with a substitution of the VaR-compliant square-root term for the variance, and is equivalent to the minimization of the VaR of the agent's excess cost due to the non-instantaneous liquidation. The interpretation of the agent's "utility" function is straightforward and quite appealing. He maximizes the expected profit of the liquidation (sale revenue net of the market impact), but assigns a penalty function to the risk defined as the standard deviation of the profit. The penalty parameter, α, depends directly on his "worry" level. Another argument in favor of choosing Eq. (7) as the optimization specification is that the VaR, as defined in Eq. (6), could be viewed as a cost determinant rather than a risk measure. The liquidating individual would regard the dollar VaR times his borrowing rate as the amount of capital he should set aside to cover potential cost of liquidation or "carry cost." This would be analogous to the definition of the market risk capital for banks with trading portfolios. It is also related to the concept of the tracking error for benchmarked portfolios, as pointed out in Fischer (2001).

The individual's marginal or capital gains tax rate does not enter the decision process. This is because both the expectation and the standard deviation of the liquidation profit, which is a normal random variable, are expressed in dollars. If one minus the tax rate pre-multiplied profit, as defined in Eqs. (5) or (5′), then it would also pre-multiply the entire Eq. (7) and would drop out of the subsequent first-order conditions (13) and (25).

Although the exact solutions presented in the next two sections depend on the choice of the maximand in Eq. (7), their essence does not. Once expressions for the expectation and the variance of the profit function are evaluated, the first order conditions could be set up for any mean-variance utility specification. The liquidation model stands on its own even without the adoption of the VaR risk-return tradeoff.

4. A general power function market impact

Let us assume that the market impact functions are of the following general form:

$$g(v_t) = \gamma v_t^G$$
$$h(v_t) = \varepsilon \ \mathrm{sgn}(n_t) + \eta v_t^H \tag{8}$$

where G and H are given constants. In a pure sale strategy, $\mathrm{sgn}(n_t) = 1$ for all t. Almgren (2001) considers a power form for the temporary impact function[4]. The trading return in the interval $(t - dt, t)$ is equal to:

$$\tilde{R}_t = \sigma z_t - \varepsilon - \eta v_t^H - \gamma \int_0^t v_s^G ds \tag{9}$$

The last three terms in Eq. (9) represent a total "liquidity discount." Using the definition (5), we can write the profit function for the agent liquidating at a constant speed, v, as[5]:

$$\prod = v\sigma \int_0^T z_t dt - \varepsilon v T - \eta v^{H+1} T - \frac{1}{2} \gamma v^{G+1} T^2 \tag{10}$$

Applying the expectation operators and using the fact that $X = vT$, where X denotes the initial holdings, the mean and the variance of the profit can be derived as[6]:

$$E[\prod] = -\varepsilon X - \eta X^{H+1} T^{-H} - \frac{1}{2} \gamma X^{G+1} T^{-G+1}$$

$$V[\prod] = \frac{1}{3} T X^2 \sigma^2 \tag{11}$$

Now we can set up the agent's optimization problem as the trade-off between the total "liquidity discount," affecting the mean of the profit, and the "market risk" of returns, represented in the total variance of the profit, multiplied by the penalty parameter α, representing the agent's risk tolerance:

$$\max_T \left\{ -\varepsilon X - \eta X^{H+1} T^{-H} - \frac{1}{2} \gamma X^{G+1} T^{-G+1} - \alpha \sqrt{\frac{1}{3} T X^2 \sigma^2} \right\} \tag{12}$$

The first-order condition for this optimization is the following equation in T:

$$H\eta X^{H+1}T^{-H-1} + \frac{1}{2}(G-1)\gamma X^{G+1}T^{-G} - \frac{1}{2\sqrt{3}}\alpha\sigma X T^{-\frac{1}{2}} = 0 \tag{13}$$

Note that $X = 1$ in the return version of the model. It is carried here to allow Eqns. (11–25) to generalize to the price-based version where it is equal to the initial number of shares.

In general, Eq. (13) has to be solved by numerical methods. However, we can obtain closed-form solutions for some special cases.

4.1. Linear market impact functions

If $G = 1$, the permanent impact function becomes linear in the speed of sales, i.e.:

$$g(v_t) = \gamma v_t$$
$$h(v_t) = \varepsilon + \eta v_t^H \tag{8'}$$

and we can solve Eq. (13) explicitly for T:

$$T = \left(\frac{2\sqrt{3}H\eta X^H}{\alpha\sigma}\right)^{\frac{1}{H+\frac{1}{2}}} \tag{14}$$

If we further assume that $G = H = 1$ so that

$$g(v_t) = \gamma v_t$$
$$h(v_t) = \varepsilon + \eta v_t \tag{8''}$$

then the optimal liquidation time is given explicitly by:

$$T = \left(\frac{2\sqrt{3}\eta X}{\alpha\sigma}\right)^{\frac{2}{3}} \tag{15}$$

Let us consider a few numerical examples for the linear market impact specification (8″). Let $\alpha = 1.645$, $\sigma = 0.5$, $\eta = 0.001899$. This corresponds to a 95% confidence interval VaR utility, an annualized return volatility of 15%, and a temporary impact of 18.99 bp when selling at a speed of X holdings per year (imagine an individual owning \$1 million (normalize $X = 1$) worth of stock whose volatility is 50% estimating to lose \$1,899 when selling X). The optimal liquidation time is equal to $T = 0.04 = 10/250$, which is equivalent to 10 days. From Eq. (11), the expected cost of liquidation at the optimal time is 4.75% assuming no permanent impact γ and additive temporary term ε. If he sells half of his holdings ($X = ½$), then his total impact will be lower and he can sell faster in $T = 0.025194$ or 6.298 business days and at an expected cost of liquidation of only 1.88%. If the annualized volatility is changed to 25% ($\alpha = 1.645$, $\sigma = 0.25$, $\eta = 0.001899$), then the agent will optimally sell over 15.87 days as the risk of staying in the market is reduced. This will allow him to lower the expected cost to 2.99%. If instead he is more risk averse and sets his confidence level to 99%, i.e. ($\alpha = 2.33$, $\sigma = 0.50$, $\eta = 0.001899$), he will reduce his optimal liquidation time to about 7.93 days, incur a higher total impact and face the expected cost of 5.99%. In all the numerical examples, the value of the temporary impact parameter η was normalized for a given level of holdings to be liquidated. This would happen naturally in the estimation process of all inputs for a given stock.

Let us turn to a "bond" example, i.e., a security whose annualized volatility is much lower than that of the stock. Let $\alpha = 1.645$, $\sigma = 0.15$, $\eta = 0.001714$. The optimal liquidation time is equal to 1 month (20.8 business days). At an even lower volatility of 10%, i.e., $\alpha = 1.645$, $\sigma = 0.1$, $\eta = 0.001714$, it is extended to 27.3 business days. And at the 99% confidence level, it is 16.5 business days. The individual takes a longer time to liquidate a less risky security in an attempt to "save" on the market impact concession.

Note that the solutions (14–15) do not depend on the permanent impact coefficient, γ, as, in both cases, it affects the trading return only through the cumulative sales total and not through the sale amount in each interval $(t - dt, t)$. They are also independent of the mean temporary impact coefficient, ε, as it is a constant subtraction from the market return no matter what the speed of sales is.

4.2. Other special cases

Let us consider the square root case for the permanent impact function. We let $G = \frac{1}{2}$ and H be general. The impact functions are of the form:

$$g(v_t) = \gamma\sqrt{v_t}$$
$$h(v_t) = \varepsilon + \eta v_t^H \tag{8'''}$$

and the closed-form solution is

$$T = \left(\frac{2\sqrt{3}H\eta X^H}{\alpha\sigma + \frac{\sqrt{3}}{2}\gamma\sqrt{X}}\right)^{\frac{1}{H+\frac{1}{2}}} \tag{16}$$

If, in addition, the temporary impact is of a square root form, with $H = \frac{1}{2}$ and the impact functions

$$g(v_t) = \gamma\sqrt{v_t}$$
$$h(v_t) = \varepsilon + \eta\sqrt{v_t} \tag{8''''}$$

then the closed-form solution is:

$$T = \frac{\sqrt{3}\eta}{\alpha\sigma + \frac{\sqrt{3}}{2}\gamma} \tag{17}$$

Of course, the choice of the exponent in the power function is subject to empirical research. The exponent of less than 1, and the square root in particular, is quite intuitive, as it exhibits diminishing returns in the relevant range. As one increases the speed of liquidation, initially the market impact grows rapidly, but as one reaches a high level of sales, the growth in the market impact decreases. Note that in the square root case the solution is linear in the temporary impact parameter η and is independent of the initial size of the holdings to be liquidated, X.

What stands out in the results (16) and (17) is that the different components of the "liquidity discount" do not have the same effect on the optimal solution. The higher the temporary impact parameter η, the longer the optimal final time. At a given liquidation speed, an increase in the temporary impact parameter would reduce the agent's realized return for each sale. To compensate for that, the agent would decide to slow down his sales and thus extend his selling horizon. This can be seen in Eq. (12) with $G = H = ½$. The part of the agent's maximand related to the temporary impact is an increasing function of T. In contrast, the objective function is a decreasing function of the final time T. The permanent impact parameter γ also reduces the realized return in each interval, but only via an expression more closely related to the cumulative amount transacted. [In Eqs. (2) and (9) the last integral is equal to the time elapsed t times a constant "determined" by the total position up-front and the final time]. It thus acts like a deterministic time-proportional "drag" on the expected returns and the profit, not unlike the market risk penalty function on the agent's total utility (the last term in Eq. (12)). Were this "drag" per unit of time to increase, the agent would be forced to speed up his sales to minimize its negative effect on total profit. In the linear case (8″), this "drag" was simply proportional to the total amount for sale and independent of time. With $G = 1$, it dropped out of the Eq. (12). As such, it operated similarly to a constant bid-ask spread and did not affect the solution at all.

Let us now consider the case with no permanent impact, i.e., $\gamma = 0$. This case is also of special interest as it is frequently difficult to distinguish between the general information-driven equilibrium price movement in the market and one's own permanent impact. In that case it is convenient to assume no permanent impact to ease the estimation of the temporary effects due to one's inventory problems. Here we can allow a general power form of the temporary impact function:

$$g(v_t) = 0$$
$$h(v_t) = \varepsilon + \eta v_t^H \qquad (8'''')$$

and still get a closed-form solution

$$T = \left(\frac{2\sqrt{3}H\eta X^H}{\alpha\sigma}\right)^{\frac{1}{H+\frac{1}{2}}} \qquad (18)$$

The solution (18) is identical to that in the linear case where the permanent impact affected the equilibrium return through the total of sales, and not in each interval. Here that reduction is zero.

5. Stochastic market impact correlated with returns

The specification in this section is designed to attempt to capture a systemic feedback loop where, as prices and realized returns in the market drop, liquidity deteriorates, leading to further temporary depression of the returns. An individual coming to market to liquidate his position would face a compound effect of market risk and evaporating liquidity. This can be accomplished by assuming

$$g(v_t) = \gamma v_t$$
$$h(v_t) = \varepsilon + (\eta + \theta \tilde{z}_t) v_t \quad (19)$$

where $\tilde{z}_t$ is another Wiener process correlated with that driving the returns in (2), i.e.

$$E[dz_s d\tilde{z}_t] = \rho\, dt, \quad if\ s = t$$
$$= 0, \qquad s \neq t \quad (20)$$

and η and θ are two constants. The trading return is a random variable dependent on the realizations of two Brownian motions:

$$\tilde{R}_t = \sigma z_t - \varepsilon - (\eta + \theta \tilde{z}_t) v_t - \gamma \int_0^t v_s ds \quad (21)$$

Under the assumption of the constant liquidation speed the profit function can be written as[7]:

$$\prod = v\sigma \int_0^T z_t dt - \varepsilon v T - \eta v^2 T - \theta v^2 \int_0^T \tilde{z}_t dt - \frac{1}{2} \gamma v^2 T^2 \quad (22)$$

To set up the optimization, we derive, using $X = vT$, the expected profit and the variance of the profit[8] as

$$E[\prod] = -\varepsilon X - \frac{\eta X^2}{T} - \frac{1}{2} \gamma X^2$$
$$V[\prod] = \frac{1}{3} X^2 \left(\sigma^2 T + \frac{\theta^2 X^2}{T} - 2\rho\theta\sigma X \right) \quad (23)$$

The agent maximizes the VaR of his profit:

$$\max_T \left\{ -\varepsilon X - \frac{\eta X^2}{T} - \frac{1}{2} \gamma X^2 - \alpha \sqrt{\frac{1}{3} X^2 \left(\sigma^2 T + \frac{\theta^2 X^2}{T} - 2\rho\theta\sigma X \right)} \right\} \quad (24)$$

He faces the following first-order condition:

$$\frac{\eta X^2}{T^2} - \frac{1}{2} \alpha \left(\frac{1}{3} X^2 \right) \left(\sigma^2 - \frac{\theta^2 X^2}{T^2} \right) \left[\frac{1}{3} X^2 \left(\sigma^2 T + \frac{\theta^2 X^2}{T} - 2\rho\theta\sigma X \right) \right]^{-\frac{1}{2}} = 0 \quad (25)$$

This expression can be rearranged to get a fifth degree polynomial in T that can be easily solved numerically. Table 1 shows the results for the numerical 'bond' example of the previous section, where we let $\alpha = 1.645$, $\sigma = 0.15$, $\eta = 0.001714$. This corresponds to a

95% confidence interval VaR utility, an annual volatility of 15%, and a temporary impact of 17.14 bp when selling at a pace of X holdings over a year. Recall that the optimal liquidation time without the correlated term was $T = 0.0833$, which was equivalent to 1 month.

Here the case where $\theta = 0$ is equivalent to the non-stochastic linear case. One common feature of all the results is that, for all correlations levels, the optimal liquidation time attains a minimum at some level of θ[9] and increases rapidly as we move away from that level in either direction. That is, the larger the shocks in absolute value, the more the agent compensates for them by staying longer in the market. Of interest are the combinations $\theta > 0$, $\rho < 0$ (and $\theta < 0$, $\rho > 0$). In those cases, the liquidity deteriorates when the returns in the markets drop. The agent's profit has high variability as the combined shocks to his trading returns are strengthened. It is optimal for him to extend the liquidation period in order to reduce the temporary impact on the realized return. In Table 1, when the "reinforcement" coefficient is low, $\theta = 0.001$, and the correlation of the market impact and the fundamental returns processes is negative, $\rho = -0.50$, the agent liquidates quickly within a little over a month, $T^* = 0.08596$. But when the feedback loop is amplified, as the "reinforcement" coefficient increases to $\theta = 0.050$, the agent's profits suffer greater variability, $Std(\Pi) = 0.0867 \gg 0.0264$. He compensates by extending his sales over time to $T^* = 0.36772$ in order to minimize the utility penalty for that. The VaR of his profit would have declined to $VaR(\Pi) = -0.2063$[10], had he stayed at the no longer optimal $T = 0.08596$. Instead it drops only to $VaR(\Pi) = -0.1474$ at the optimal $T^* = 0.36772$ as shown in Table 1.

The interesting feature of the results is that in a rapidly falling market (falling for reasons other than the individual's sales), it may actually be optimal to wait longer to liquidate. This is true not because the individual hopes for the returns to improve, but because he tries to prevent a further deterioration of returns by 'dumping' his large holding into the market.

6. Conclusion

This paper's endogenous liquidity model extends the institutional research on optimal execution to a case of a wealthy individual or an estate trust with large illiquid holdings of a stock. The conversion of wealth into cash, subject to sale "frictions," has not been heretofore examined in the personal finance literature. The individual's case is different from the institutional one for two reasons. First, the individual balances the desire to limit the risk of exposure to random market factors against the negative impact his sales have on the realized return by choosing the length of time in which to liquidate his holdings as opposed to a tick-by-tick trajectory. The total time is not exogenously imposed on him by some profit cycle. Second, the liquidation strategy is only a part of his lifetime wealth accumulation-disposition goal. Because of that, his risk-reward tradeoff takes into account the desire to ensure the highest level of the realized value of his position at some confidence level, given the market risk of the asset and the impact of his sales. In a way, liquidation is a "random tax" on his accumulated wealth which the individual tries to minimize. Despite its mathematical complexity, the model is quite easy to use as it offers closed-form solutions for certain parameterizations. The inputs can be easily estimated from general stock market data.

Table 1
Optimal liquidation times, the expected value, variance and VaR of the profit as functions of θ and ρ

θ	Optimal T					E[Π]			Std[Π]			VaR[Π]		
	$\rho = -0.99$	$\rho = -0.5$	$\rho = 0$	$\rho = 0.5$	$\rho = 0.99$	$\rho = -0.99$	$\rho = -0.5$	$\rho = 0$	$\rho = -0.99$	$\rho = -0.5$	$\rho = 0$	$\rho = -0.99$	$\rho = -0.5$	$\rho = 0$
−0.050	0.33627	0.35359	0.36165	0.36772	0.37266	−0.0051	−0.0048	−0.0047	0.0071	0.0501	0.0708	−0.0168	−0.0872	−0.1213
−0.030	0.20380	0.22547	0.23520	0.24239	0.24818	−0.0084	−0.0076	−0.0073	0.0055	0.0390	0.0551	−0.0175	−0.0718	−0.0980
−0.010	0.07462	0.10636	0.11805	0.12624	0.13260	−0.0230	−0.0161	−0.0145	0.0040	0.0247	0.0342	−0.0296	−0.0568	−0.0707
−0.003	0.07093	0.08122	0.08783	0.09296	0.09714	−0.0242	−0.0211	−0.0195	0.0167	0.0223	0.0263	−0.0516	−0.0577	−0.0628
−0.002	0.07486	0.08080	0.08541	0.08920	0.09246	−0.0229	−0.0212	−0.0201	0.0195	0.0229	0.0256	−0.0550	−0.0588	−0.0622
−0.001	0.07902	0.08164	0.08390	0.08596	0.08782	−0.0217	−0.0210	−0.0204	0.0223	0.0238	0.0252	−0.0584	−0.0601	−0.0618
0.000	0.08338	0.08338	0.08338	0.08338	0.08338	−0.0206	−0.0206	−0.0206	0.0250	0.0250	0.0250	−0.0617	−0.0617	−0.0617
0.001	0.08782	0.08596	0.08390	0.08164	0.07902	−0.0195	−0.0199	−0.0204	0.0276	0.0264	0.0252	−0.0649	−0.0634	−0.0618
0.002	0.09246	0.08920	0.08541	0.08080	0.07486	−0.0185	−0.0192	−0.0201	0.0301	0.0280	0.0256	−0.0680	−0.0653	−0.0622
0.003	0.09714	0.09296	0.08783	0.08122	0.07093	−0.0176	−0.0184	−0.0195	0.0325	0.0297	0.0263	−0.0711	−0.0672	−0.0628
0.010	0.13260	0.12624	0.11805	0.10636	0.07462	−0.0129	−0.0136	−0.0145	0.0473	0.0414	0.0342	−0.0907	−0.0816	−0.0707
0.030	0.24818	0.24239	0.23520	0.22547	0.20380	−0.0069	−0.0071	−0.0073	0.0777	0.0675	0.0551	−0.1348	−0.1181	−0.0980
0.050	0.37266	0.36772	0.36165	0.35359	0.33627	−0.0046	−0.0047	−0.0047	0.0999	0.0867	0.0708	−0.1689	−0.1474	−0.1213

For closely held stocks, they can be assessed with significant accuracy with a help of an investment banker specializing in private placements.

Notes

1. The cumulative return, R_t, is computed over the entire interval (0, t). In discrete time, the cumulative return through period k, would be equivalently defined by $1 + R_k = \Pi_{i=1}^{k}(1 + r_i)$, where r_i is the return on the asset in period i.
2. Note that both R_t and $\tilde{R}_t$ are random. The tilde symbol merely distinguishes the 'intra-period' return realized by our agent from the equilibrium return in the market.
3. See References for the main regulatory pronouncements.
4. In a concurrent paper, Almgren extends beyond his original linear cases for temporary impact.
5. First apply definition (5) to write:

$$\prod = v\sigma\int_0^T z_t dt - \varepsilon v\int_0^T dt - \eta v^{H+1}\int_0^T dt - \gamma v^{G+1}\int_0^T\left(\int_0^t ds\right)dt$$

then evaluate the integrals.

6. First apply the expectations operator to Eq. (10) to get:

$$E[\prod] = -\varepsilon vT - \eta v^{H+1}T - \frac{1}{2}\gamma v^{G+1}T^2.$$

Use integration by parts to show that

$$E\left[\int_0^T z_t dt\right] = E\left[tz_t\Big|_0^T\right] - E\left[\int_0^T tdz_t\right] = 0.$$

To derive the variance, apply the expectations operators to Eq. (10). First show that

$$\int_0^T z_t dt = tz_t|_0^T - \int_0^T tdz_t = Tz_T - \int_0^T tdz_t = T\int_0^T dz_t - \int_0^T tdz_t = \int_0^T (T-t)dz_t.$$

Then show that

$$V\left[\int_0^T z_t dt\right] = E\left[\int_0^T (T-t)dz_t\right]^2 = E\left[\int_0^T (T-t)^2 dz_t^2\right] = E\left[\int_0^T (T-t)^2 dt\right]$$

$$= \int_0^T (T-t)^2 dt = \frac{1}{3}T^3$$

and that

$$V[\prod] = \frac{1}{3}\, T^3\sigma^2 v^2.$$

7. First apply Definition (5) to get:

$$\prod = v\sigma\int_0^T z_t dt - \varepsilon v\int_0^T dt - \eta v^2\int_0^T dt - \theta v^2\int_0^T \tilde{z}_t dt - \gamma v^2\int_0^T\left(\int_0^t ds\right)dt$$

and then simply evaluate the non-stochastic integrals.

8. To derive the variance we need to evaluate the following expression:

$$I = E\left[\int_0^T z_t dt \cdot \int_0^T \tilde{z}_t\, dt\right]$$

$$= E\left[\int_0^T\int_0^T z_u\tilde{z}_w\, du dw\right].$$

$$= \int_0^T\int_0^T E[z_u\tilde{z}_w] du dw$$

We can split the inside integral w.r.t. du in two regions $u \in (0, w)$ and $u \in (w, T)$. Noting that $E[z_u\, \tilde{z}_w] = \rho \cdot \min[u, w]$, the expression I evaluates to:

$$I = \int_0^T\int_0^w \rho u du dw + \int_0^T\int_w^T \rho w du dw = \frac{1}{3}\rho T^3.$$

9. That level depends on all the other parameters through the first-order condition (25). In the example, it is slightly negative as can be seen from Table 1.
10. Not shown in the Table. Plug in the original T = 0.08596 to the VaR Equation (24).

Acknowledgment

Special thanks for inspiration and helpful comments to Shantaram Hegde, my colleagues in the Corporate Risk Management group at Merrill Lynch, and the Doctoral Colloquium participants at the 2001 FMA meetings in Toronto.

Appendix A

In Almgren and Chriss (2000), the equation of motion for the equilibrium price is:

$$S_k = S_{k-1} + \sigma\sqrt{\tau}\Delta\tilde{z}_k - \tau g(v_k)$$

where $g(v_k)$ represents the permanent impact function, σ is the annualized normal volatility of the stock price, and $\Delta\tilde{z}k$ is a standard normal deviate, with $E[\Delta\tilde{z}_k\,\Delta\tilde{z}_{k'}] = 0, k \neq k'$. Note that $g(v_k)$ is pre-multiplied by τ to emphasize that the total permanent impact effect depends more on the total number of shares, τv_k, sold in each interval, than on the pure intensity parameter v_k. The trading price, $\tilde{S}k$, is equal to:

$$\tilde{S}_k = S_{k-1} - h(v_k)$$

The cost of liquidation is defined as:

$$C = XS_0 - \sum_{k=1}^{N} n_k\tilde{S}_k$$

It is easy to show that the cost reduces to:

$$C = -\sum_{k=1}^{N}\sigma\sqrt{\tau}\Delta\tilde{z}_k x_k + \sum_{k=1}^{N}\tau g(v_k)x_k + \sum_{k=1}^{N}n_k h(v_k)$$

The cost is a random variable as it is subject to the sequence of the random shocks $\sigma\sqrt{\tau}\,\Delta\,\tilde{z}_k$. Because the shocks are independent, it follows straightforwardly that the mean and the variance of the cost function are:

$$E[C] = \sum_{k=1}^{N}\tau\, g(v_k)\, x_k + \sum_{k=1}^{N}n_k\, h(v_k)$$

$$V[C] = \sigma^2\tau\sum_{k=1}^{N}x_k^2$$

Both impact functions are assumed to be linear in the sales intensity parameter v_k (and thus the sales quantity parameter n_k). The permanent impact function is $g(v_k) = \gamma v_k$, and the temporary impact function is $h(v_k) = \varepsilon\ \mathrm{sgn}(n_k) + \eta v_k$, with $\mathrm{sgn}(n_k) = 1$ for a sale strategy. Because $h(v_k)$ is denominated in \$/share, ε can be thought of as half the bid-ask spread denominated in \$/share, and η has a peculiar denomination of (\$/share)/(share/time).

The trader minimizes the expected cost given a level of variance. This is equivalent to an unconstrained minimization of

$$\min_{\{x_k:k=1,\ldots,N\}}\{E[C] + \alpha V[C]\}$$

where the Lagrange multiplier α can be interpreted as the relative risk aversion coefficient.

References

Almgren, R., & Chriss N. (2000/2001, Winter). Optimal execution of portfolio transactions. *Journal of Risk*, *3*, 5–39.

Almgren, R. Optimal execution with nonlinear impact functions and trading-enhanced risk, University of Toronto Working Paper.

Bangia, A., Diebold, F.X., Schuermann, T., & Stroughair, J.D. (1999). Modeling liquidity risk, with implications for traditional market risk measurement and management, Working Paper 99-06, Financial Institutions Center, The Wharton School.

Bank of International Settlements. (1996). Amendment to the capital accord to incorporate market risks, January 1996, updated April 1998.

Basak, S., & Shapiro A. (2001). Value-at-risk-based risk management: Optimal policies and asset management. *Review of Financial Studies* 14 (2), 371–406.

Bertsimas, D., Lo, A.W. (1998). Optimal control of execution costs. *Journal of Financial Markets*, 1, 1–50.

Board of Governors of the Federal Reserve System, February 1998. Trading and capital-markets activities manual.

Brennan, M.J., & Subramanjam, A. (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics*, 41, 441–64.

Chan, L.K.C., & Lakonishok J. (1995). The behavior of stock prices around institutional trades. *Journal of Finance*, 50, 1147–74.

Chhabra, A.B., & Zaharoff L. (2001). Setting an asset allocation strategy by balancing personal and market risks. *Journal of Wealth Management*, 4, 30–33.

Dubil, R. (2001). How to include liquidity in a market VaR statistic?, Working Paper, University of Connecticut.

Financial Services Authority. (1998). Board Notice 486 of the Securities and Futures Authority, 9 September 1998.

Fischer, B. (2001). Understanding tracking error and its relationship with VaR. *Journal of Investing*, 10, 54–60.

Glosten, L.R., & Harris L.E. (1988). Estimating the components of the bid-ask spread. *Journal of Financial Economics*, 21, 123–142.

Holthausen, R.W., Leftwich R.W., & Mayers, D. (1987). The effect of large block transactions on security prices. *Journal of Financial Economics*, 19, 237–267.

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

Huang, C.-F., & Litzenberger, R. (1988). Foundations for financial economics. Amsterdam: North Holland.

Kyle, A.S. (1985). Continuous auctions and insider trading. *Econometrica* 53 (6), 1315–1335.

O'Hara, M. (1997). Market microstructure theory. Cambridge, MA: Blackwell Publishers, Inc.