

Optimal execution of portfolio transactions

Robert Almgren

Department of Mathematics, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada

Neil Chriss

ICor Brokerage Inc., 500 Fifth Avenue, Suite 415, New York, New York 10110, USA

The authors consider the execution of portfolio transactions with the aim of minimizing a combination of volatility risk and transaction costs arising from permanent and temporary market impact. In the space of time-dependent liquidation strategies, the *efficient frontier* consists of strategies having the lowest expected execution cost for a given level of uncertainty; with a linear cost model this frontier can be explicitly constructed. It is then possible to select particular optimal strategies either by minimizing a quadratic utility function or by minimizing value-at-risk (VaR). The latter choice leads to the concept of liquidity-adjusted VaR, or L-VaR, which explicitly considers the best tradeoff between volatility risk and liquidation costs.

1. INTRODUCTION

This paper concerns the optimal execution of *portfolio transactions*, transactions that move a portfolio from a given starting composition to a specified final composition within a specified period of time. Bertsimas and Lo (1998) have defined best execution as the dynamic trading strategy that provides the minimum expected cost of trading over a fixed period of time, and they show that in a variety of circumstances one can find such a strategy by employing a dynamic optimization procedure; but their approach ignores the volatility of revenues for different trading strategies. This paper works in the more general framework of maximizing the *expected revenue* of trading (or, equivalently, minimizing the costs), with a suitable penalty for the *uncertainty* of revenue (or cost).¹

We study variance of trading cost in optimal execution because it fits with the intuition that a trader's utility should figure in the definition of *optimal* in

¹ This general framework arises in market microstructure theory, but with a different purpose in mind. The *uninformed discretionary trader* trades an exogenous endowment over an exogenously specified amount of time to maximize profits (Admati and Pfleiderer 1988); the *informed strategic trader* trades over multiple periods on information not widely available, again to maximize profits (Kyle 1985). In both cases, the literature focuses on the link between the trader and the market maker and a theory is produced to predict the market clearing price of a security at each period. Thus, a trader's optimal strategy is used as a means to study price formation in markets, not as an object of interest in itself.

“optimal execution”. For example, in trading a highly illiquid volatile security, there are two extreme strategies: trade everything now at a known but high cost, or trade in equal-sized packets over a fixed time at relatively lower cost. The latter strategy has lower expected cost, but this comes at the expense of greater uncertainty in final revenue. How to evaluate this uncertainty is partly subjective and a function of the trader’s tolerance for risk. All we can do is insist that, for a given level of uncertainty, cost be minimized. This idea extends to a complete theory of optimal execution that includes an efficient frontier of optimal execution strategies (see Section 3).

The framework of risk in execution yields several important results consistent with intuition. For example, it is evident that, all else equal, a trader will choose to execute a block of an illiquid security less rapidly than a liquid security. While this seems obvious, we show that a model that ignores risk does not have this property: without enforcing a strictly positive penalty for risk, one cannot produce models that trade differently across the spectrum of liquidity.

The incorporation of risk into the study of optimal execution does not come without cost. First, in order to produce tractable analytic results, we are forced to work largely in the framework of price dynamics that are an arithmetic random walk with independent increments. We obtain our results using *static* optimization procedures which we show lead to globally optimal trading trajectories. That is, optimal trading paths may be determined in advance of trading. Only the composition of the portfolio and the trader’s utility function figure in the trading path. The fact that a static strategy can be optimal even when a trader has the option to dynamically change his trading mid-course is a direct result of the assumptions of independence of returns and symmetry of the penalty function for risk.²

As it is well known that price movements exhibit some serial correlation across various time horizons (Lo and MacKinlay 1988), that market conditions change, and that some participants possess private information (Bertsimas and Lo 1998), one may question the usefulness of results that obtain strictly in an independent-increment framework. Moreover, as trading is known to be a dynamic process, our conclusion that optimal trading strategies can be statically determined calls for critical examination. We consider quantitatively what gains are available to strategies that incorporate all relevant information.

First, we consider short-term serial correlation in price movements. We

² An interesting deviation from the symmetric penalty function was communicated to us by Ferstenberg, Karchmer, and Malamut at ITG, Inc. They argue that opportunity cost is a subjective quantity that is measured differently by different traders. Using a trader-defined cost function g , they define opportunity costs as the expected value of g applied to the average execution price obtained by the trader relative to the benchmark price. They assume that risk-averse traders will use a convex function g that is not symmetric in the sense that there is a strictly greater penalty for under-performance than for the same level of outperformance. They show that, in this setting, the optimal execution strategy relative to g depends not only on the time remaining but also on the performance of the strategy up to the present time and on the present price of the security. In particular, this means that, in their setting, optimal strategies are dynamic.

demonstrate that the marginal improvement available by explicitly incorporating this information into trading strategies is small and, more importantly, independent of portfolio size; as portfolio sizes increase, the percentage gains possible decrease proportionally.³

Second, we examine the impact of scheduled news events on optimal execution strategies. There is ample evidence that anticipated news announcements, depending on their outcome, can have significant temporary impact on the parameters governing price movements.⁴ We work in a simple extension of our static framework by assuming that the security again follows an arithmetic random walk, but that, at a time known at the beginning of trading, an uncorrelated event will cause a material shift in price dynamics (e.g., an increase or decrease in volatility).

In this context, we show that optimal strategies are piecewise static. To be precise, we show that the optimal execution strategy entails following a static strategy up to the moment of the event, followed by another static strategy that can only be determined once the outcome of the event is known. It is interesting to note that the static strategy one follows in the first leg is in general not the same strategy one would follow in the absence of information concerning the event.

Finally, we note that any optimal execution strategy is vulnerable to *unanticipated events*. If such an event occurs during the course of trading and causes a material shift in the parameters of the price dynamics, then indeed a shift in the optimal trading trajectory must also occur. However, if one makes the simplifying assumption that all events are either “scheduled” or “unanticipated”, then one concludes that optimal execution is always a game of static trading punctuated by shifts in trading strategy that adapt to material changes in price dynamics. If the shifts are caused by events that are known ahead of time, then optimal execution benefits from precise knowledge of the possible outcomes of the event. If not, then the best approach is to be actively “watching” the market for such changes, and to react swiftly should they occur. One approximate way to include such completely unexpected uncertainty into our model is to artificially raise the value of the volatility parameter.

³ This is precisely true for a linear transaction cost model, and approximately true for more general models. The results of Bertsimas and Lo (1998) suggest that a trading strategy built to take advantage of serial correlations will essentially be a combination of a “correlation-free” strategy and a “shifting” strategy that moves trades from one period to the next based on information available in last period’s return. Therefore we argue that, by ignoring serial correlation, we (a) preserve the main interesting features of the analysis, and (b) introduce virtually no bias away from the “truly optimal” solutions.

⁴ For a theoretical treatment, see Brown, Harlow, and Tinic (1988), Easterwood and Nutt (1999), Kim and Verrecchia (1991), and Ramaswami (1999). For empirical studies concerning earnings announcements, see Patell and Wolfson (1984) for changes in mean and variance of intraday prices, and Krinsky and Lee (1996) and Lee, Mucklow, and Ready (1993) for changes in the bid–ask spread. For additional studies concerning anticipated news announcements, see Charest (1978), Kalay and Loewentstein (1985), and Morse (1981).

Having indicated why we work in the framework we have chosen, we now outline some of our results. First, we obtain optimal solutions in closed form for any level of risk aversion. We show that this leads to an efficient frontier of optimal strategies, where an element of the frontier is represented by a strategy with the minimal level of cost for its level of variance of cost. The structure of the frontier is of some interest. It is a smooth convex function, differentiable at its minimal point. The minimal point is what Bertsimas and Lo (1998) call the “naïve” strategy because it corresponds to trading in equal-sized packets, using all available trading time equally. The differentiability of the frontier at its minimum point indicates that one can obtain a first-order reduction in variance of trading cost at the expense of only a second-order increase in cost by trading a strategy slightly away from the globally minimal strategy. The curvature of the frontier at its minimum point is a measure of liquidity of the security.

Another ramification of our study is that, for all levels of risk aversion except risk neutrality, optimal execution trades have a “half-life” which falls out of our calculations. A trade’s half-life is independent of the actual specified time to liquidation, and is a function of the security’s liquidity and volatility and the trader’s level of risk aversion. As such, we regard the half-life as an idealized time for execution, and perhaps a guide to the proper amount of time over which to execute a transaction. If the specified time to liquidation is short relative to the trade’s half-life, then one can expect the cost of trading to be dominated by transaction costs. If the time to trade is long relative to its half-life, then one can expect most of the liquidation to take place well in advance of the limiting time.

In Section 2, we present our model for market impact. In Section 3, we construct optimal trajectories and present the efficient frontier, including the “half-life of optimal trading”. In Section 4, we consider various ways that risk can be balanced against certain costs, both using a concave utility function and value-at-risk, and we present a concrete numerical example. In Section 5, we consider the value of possible additional information on future stock price motion: a nonzero drift in the random walk, serial autocorrelation, and parameter changes or “regime shifts”. Finally, in the Appendix, we extend our analysis to multiple-asset portfolios and again produce explicit solutions for optimal trading paths. In that case, not surprisingly, the correlation between assets has a strong influence on optimal trading behavior.

2. THE TRADING MODEL

This section defines a trading strategy, and lays out the price dynamics we will study. We start with a formal definition of a trading strategy for execution of a sell program consisting of liquidating a single security. The definition and results for a buy program are completely analogous.

2.1 The Definition of a Trading Strategy

Suppose we hold a block of X units of a security⁵ that we want to completely liquidate before time T . We divide T into N intervals of length $\tau = T/N$, and define the discrete times $t_k = k\tau$ for $k = 0, \dots, N$. We define a *trading trajectory* to be a list x_0, \dots, x_N , where x_k is the number of units that we plan to hold at time t_k . Our initial holding is $x_0 = X$, and liquidation at time T requires $x_N = 0$.⁶

We may equivalently specify a strategy by the “trade list” n_1, \dots, n_N , where $n_k = x_{k-1} - x_k$ is the number of units that we will sell between times t_{k-1} and t_k . Clearly, x_k and n_k are related by

$$x_k = X - \sum_{j=1}^k n_j = \sum_{j=k+1}^N n_j \quad \text{for } k = 0, \dots, N.$$

We consider more general programs of simultaneously buying and selling several securities in the Appendix. For notational simplicity, we have taken all the time intervals to be of equal length τ , but this restriction is not essential. Although we shall not discuss it, in all our results it is easy to take the continuous-time limit $N \rightarrow \infty$, $\tau \rightarrow 0$.

We define a “trading strategy” to be a rule for determining n_k in terms of information available at time t_{k-1} . Broadly speaking we distinguish two types of trading strategies: dynamic and static. Static strategies are determined in advance of trading, i.e., the rule for determining each n_k depends only on information available at time t_0 . Dynamic strategies, conversely, depend on all information up to and including time t_{k-1} .

2.2 Price Dynamics

Suppose that the initial security price is S_0 , so that the initial market value of our position is XS_0 . The security’s price evolves according to two exogenous factors, volatility and drift, and one endogenous factor, market impact. Volatility and drift are assumed to be the result of market forces that occur randomly and independently of our trading.

As market participants begin to detect the volume we are selling (buying) they naturally adjust their bids (offers) downward (upward).⁷ We distinguish two kinds of market impact. *Temporary* impact refers to temporary imbalances in supply and demand caused by our trading leading to temporary price movements away from equilibrium. *Permanent* impact means changes in the

⁵ To keep the discussion general we will speak of *units* of a security. Specifically we have in mind shares of stock, futures contracts, and units of foreign currency.

⁶ A trading trajectory may be thought of as either the *ex-post* realized trades resulting from some process, or as a plan concerning how to trade a block of securities. In either case, we may also consider *rebalancing* trajectories by requiring $x_0 = X$ (initial position) and $x_1 = Y$ (new position), but this is formally equivalent to studying trajectories of the form $x_0 = X - Y$ and $x_N = 0$.

⁷ Our discussion largely reflects the work of Kraus and Stoll (1972), and the subsequent work of Holthausen, Leftwich, and Mayers (1987, 1990) and Chan and Lakonishok (1993, 1995). See also Keim and Madhavan (1995, 1997).

“equilibrium” price due to our trading, which remain at least for the life of our liquidation.

We assume that the security price evolves according to the discrete arithmetic random walk

$$S_k = S_{k-1} + \sigma\tau^{1/2}\xi_k - \tau g\left(\frac{n_k}{\tau}\right) \quad \text{for } k = 1, \dots, N. \quad (1)$$

Here σ represents the volatility of the asset, the ξ_j are draws from independent random variables each with zero mean and unit variance, and the permanent impact $g(v)$ is a function of the *average rate* of trading $v = n_k/\tau$ during the interval t_{k-1} to t_k . In equation (1), there is no drift term. We interpret this as the assumption that we have no information about the direction of future price movements.⁸

2.3 Temporary Market Impact

The intuition behind temporary market impact is that a trader plans to sell a certain number of units n_k between times t_{k-1} and t_k , but may work the order in several smaller slices to locate optimal points of liquidity. If the total number of units n_k is sufficiently large, the execution price may steadily decrease between t_{k-1} and t_k , in part due to exhausting the supply of liquidity at each successive price level. We assume that this effect is short lived and, in particular, that liquidity returns after each period and a new equilibrium price established.

We model this effect by introducing a temporary price impact function $h(v)$, the temporary drop in average price per share caused by trading at average rate v during one time interval. Given this, the actual price per share received on sale k is

$$\tilde{S}_k = S_{k-1} - h\left(\frac{n_k}{\tau}\right), \quad (2)$$

but the effect of $h(v)$ does not appear in the next “market” price S_k .⁹

The functions $g(v)$ in (1) and $h(v)$ in (2) may be chosen to reflect any preferred model of market microstructure, subject only to certain natural convexity conditions.

2.4 Capture and Cost of Trading Trajectories

We now discuss the profits resulting from trading along a certain trajectory. We define the *capture* of a trajectory to be the full trading revenue upon completion

⁸ Over long-term “investment” time scales or in extremely volatile markets, it is important to consider *geometric* rather than arithmetic Brownian motion; this corresponds to letting σ in (1) scale with S . But over the short-term “trading” time horizons of interest to us, the total fractional price changes are small, and the difference between arithmetic and geometric Brownian motions is negligible.

⁹ This model is, in effect, completely general and we claim no special insight into the workings of market impact in this section.

of all trades.¹⁰ This is the sum of the product of the number of units n_k that we sell in each time interval times the effective price per share \tilde{S}_k received on that sale. We readily compute

$$\sum_{k=1}^N n_k \tilde{S}_k = XS_0 + \sum_{k=1}^N \left[\sigma \tau^{1/2} \xi_k - \tau g\left(\frac{n_k}{\tau}\right) \right] x_k - \sum_{k=1}^N n_k h\left(\frac{n_k}{\tau}\right). \quad (3)$$

The first term on the right-hand side of (3) is the initial market value of our position; each additional term represents a gain or a loss due to a specific market factor.

The first term of this type is $\sum \sigma \tau^{1/2} \xi_k x_k$, representing the total effect of volatility. The permanent market impact term $-\sum \tau x_k g(n_k/\tau)$ represents the loss in value of our total position, caused by the permanent price drop associated with selling a small piece of the position. And the temporary market impact term, $\sum n_k h(n_k/\tau)$, is the price drop due to our selling, acting only on the units that we sell during the k th period.

The *total cost of trading* is the difference $XS_0 - \sum n_k \tilde{S}_k$ between the initial book value and the capture. This is the standard *ex-post* measure of transaction costs used in performance evaluations, and is essentially what Perold (1988) calls *implementation shortfall*.

In this model, prior to trading, implementational shortfall is a random variable. We write $E(x)$ for the expected shortfall and $V(x)$ for the variance of the shortfall. Given the simple nature of our price dynamics, we readily compute

$$E(x) = \sum_{k=1}^N \tau x_k g\left(\frac{n_k}{\tau}\right) + \sum_{k=1}^N n_k h\left(\frac{n_k}{\tau}\right), \quad (4)$$

$$V(x) = \sigma^2 \sum_{k=1}^N \tau x_k^2. \quad (5)$$

The units of E are dollars; the units of V are dollars squared.

The distribution of shortfall is exactly Gaussian if the ξ_k are Gaussian; in any case, if N is large, it is very nearly Gaussian.

The rest of this paper is devoted to finding trading trajectories that minimize $E(x) + \lambda V(x)$ for various values of λ . We will show that, for each value of λ , there corresponds a unique trading trajectory x such that $E(x) + \lambda V(x)$ is minimal.

2.5 Linear Impact Functions

Although our formulation does not require it, computing optimal trajectories is significantly easier if we take the permanent and temporary impact functions to be *linear* in the rate of trading.

¹⁰ Due to the short time horizons we consider, we do not include any notion of carry or time value of money in this discussion.

For linear permanent impact, we consider $g(v)$ with the form

$$g(v) = \gamma v, \tag{6}$$

in which the constant γ has units of (\$/share)/share. With this form, each n units that we sell depresses the price per share by γn , regardless of the time we take to sell the n units; equation (1) readily yields

$$S_k = S_0 + \sigma \sum_{j=1}^k \tau^{1/2} \xi_j - \gamma(X - x_k).$$

Summing by parts, we then obtain the permanent impact term in (4) as

$$\begin{aligned} \sum_{k=1}^N \tau x_k g\left(\frac{n_k}{\tau}\right) &= \gamma \sum_{k=1}^N x_k n_k = \gamma \sum_{k=1}^N x_k (x_{k-1} - x_k) \\ &= \frac{1}{2} \gamma \sum_{k=1}^N [x_{k-1}^2 - x_k^2 - (x_k - x_{k-1})^2] = \frac{1}{2} \gamma X^2 - \frac{1}{2} \gamma \sum_{k=1}^N n_k^2. \end{aligned}$$

Similarly, for the temporary impact, we take

$$h\left(\frac{n_k}{\tau}\right) = \epsilon \operatorname{sgn}(n_k) + \frac{\eta}{\tau} n_k, \tag{7}$$

where sgn is the sign function.

The units of ϵ are \$/share, and those of η are (\$/share)/(share/time). A reasonable estimate for ϵ is the fixed costs of selling, such as half the bid–ask spread plus fees. It is more difficult to estimate η since it depends on internal and transient aspects of the market microstructure. It is in this term that we would expect nonlinear effects to be most important, and the approximation (7) to be most doubtful.

The linear model (7) is often called a *quadratic* cost model because the *total* cost incurred by buying or selling n units in a single unit of time is

$$nh\left(\frac{n}{\tau}\right) = \epsilon |n| + \frac{\eta}{\tau} n^2.$$

With both linear cost models (6) and (7), the expectation of impact costs (4) becomes

$$E(x) = \frac{1}{2} \gamma X^2 + \epsilon \sum_{k=1}^N |n_k| + \frac{\tilde{\eta}}{\tau} \sum_{k=1}^N n_k^2, \tag{8}$$

in which

$$\tilde{\eta} = \eta - \frac{1}{2} \gamma \tau.$$

Clearly, E is a strictly convex function as long as $\tilde{\eta} > 0$. Note that if the n_k all

have the same sign, as would typically be the case for a pure sell program or a pure buy program, then $\sum |n_k| = |X|$.

To illustrate, let us compute E and V for linear impact functions for the two most extreme trajectories: sell at a constant rate, and sell to minimize variance without regard to transaction costs.

Minimum impact The most obvious trajectory is to sell at a constant rate over the whole liquidation period. Thus, we take each

$$n_k = X/N \quad \text{and} \quad x_k = (N - k)X/N \quad \text{for } k = 1, \dots, N. \quad (9)$$

From (4) and (8), we have

$$\begin{aligned} E &= \frac{1}{2}XTg\left(\frac{X}{T}\right)\left(1 - \frac{1}{N}\right) + Xh\left(\frac{X}{T}\right) \\ &= \frac{1}{2}\gamma X^2 + \epsilon X + (\eta - \frac{1}{2}\gamma\tau)\frac{X^2}{T}, \end{aligned} \quad (10)$$

and, from (5),

$$V = \frac{1}{3}\sigma^2 X^2 T \left(1 - \frac{1}{N}\right) \left(1 - \frac{1}{2N}\right). \quad (11)$$

This trajectory minimizes total expected costs, but the variance may be large if the period T is long. As the number of trading periods $N \rightarrow \infty$, $v = X/T$ remains finite, and E and V have finite limits.

Minimum variance The other extreme is to sell our entire position in the first time step. We then take

$$n_1 = X, \quad n_2 = \dots = n_N = 0, \quad x_1 = \dots = x_N = 0, \quad (12)$$

which give

$$E = Xh\left(\frac{X}{\tau}\right) = \epsilon X + \eta \frac{X^2}{\tau}, \quad V = 0. \quad (13)$$

This trajectory has the smallest possible variance, equal to zero because of the way that we have discretized time in our model. If N is large and hence τ is short, then, on our full initial portfolio, we take a price hit which can be arbitrarily large.

The purpose of this paper is to show how to compute optimal trajectories that lie between these two extremes.

3. THE EFFICIENT FRONTIER OF OPTIMAL EXECUTION

In this section, we define and compute optimal execution trajectories and go on in Section 4 to demonstrate a precise relationship between risk aversion and the

definition of optimality. In particular, we show that for each level of risk aversion there is a uniquely determined optimal execution strategy.

3.1 The Definition of the Frontier

A rational trader will always seek to minimize the expectation of shortfall for a given level of variance of shortfall. Naturally, a trader will prefer a strategy that provides the minimum error in its estimate of expected cost. Thus we define a trading strategy to be *efficient* or *optimal* if there is no strategy which has lower variance for the same or lower level of expected transaction costs, or, equivalently, no strategy which has a lower level of expected transaction costs for the same or lower level of variance.¹¹

We may construct efficient strategies by solving the constrained optimization problem

$$\min_{x: V(x) \leq V_*} E(x). \quad (14)$$

That is, for a given maximum level of variance $V_* \geq 0$, we find a strategy that has minimum expected level of transaction costs. Since $V(x)$ is convex, the set $\{V(x) \leq V_*\}$ is convex (it is a sphere), and since $E(x)$ is strictly convex, there is a unique minimizer $x_*(V_*)$.

Regardless of our preferred balance of risk and return, every other solution x which has $V(x) \leq V_*$ has higher expected costs than $x_*(V_*)$ for the same or lower variance, and can never be efficient. Thus, the family of all possible efficient (optimal) strategies is parametrized by the single variable V_* , representing all possible maximum levels of variance in transaction costs. We call this family *the efficient frontier of optimal trading strategies*.

We solve the constrained optimization problem (14) by introducing a Lagrange multiplier λ , solving the unconstrained problem

$$\min_x [E(x) + \lambda V(x)]. \quad (15)$$

If $\lambda > 0$, then $E + \lambda V$ is strictly convex, and (15) has a unique solution $x^*(\lambda)$. As λ varies, $x^*(\lambda)$ sweeps out the same one-parameter family, and thus traces out the efficient frontier.

The parameter λ has a direct financial interpretation. It is already apparent from (15) that λ is a measure of risk aversion, i.e., how much we penalize variance relative to expected cost. In fact, λ is the curvature (second derivative) of a smooth utility function, as we shall make precise in Section 4.

For given values of the parameters, problem (15) can be solved by various numerical techniques depending on the functional forms chosen for $g(v)$ and $h(v)$. In the special case that these are *linear* functions, we may write the solution explicitly and gain a great deal of insight into trading strategies.

¹¹ This definition of optimality of a strategy is the same whether the strategy is dynamic or static. Later we will establish that, under this definition and the price dynamics already stated, optimal strategies are in fact static.

3.2 Explicit Construction of Optimal Strategies

With $E(x)$ from (8) and $V(x)$ from (5), and assuming that n_j does not change sign, the combination $U(x) = E(x) + \lambda V(x)$ is a quadratic function of the control parameters x_1, \dots, x_{N-1} ; it is strictly convex for $\lambda \geq 0$. Therefore we determine the unique global minimum by setting its partial derivatives to zero. We readily calculate

$$\frac{\partial U}{\partial x_j} = 2\tau \left(\lambda \sigma^2 x_j - \tilde{\eta} \frac{x_{j-1} - 2x_j + x_{j+1}}{\tau^2} \right) \quad \text{for } j = 1, \dots, N-1.$$

Then $\partial U / \partial x_j = 0$ is equivalent to

$$\frac{1}{\tau^2} (x_{j-1} - 2x_j + x_{j+1}) = \tilde{\kappa}^2 x_j, \quad (16)$$

with

$$\tilde{\kappa}^2 = \frac{\lambda \sigma^2}{\tilde{\eta}} = \frac{\lambda \sigma^2}{\eta(1 - \gamma\tau/2\eta)}.$$

Note that equation (16) is a linear difference equation whose solution may be written as a combination of the exponentials $\exp(\pm \kappa t_j)$, where κ satisfies

$$\frac{2}{\tau^2} [\cosh(\kappa\tau) - 1] = \tilde{\kappa}^2.$$

The tildes on $\tilde{\eta}$ and $\tilde{\kappa}$ denote an $O(\tau)$ correction; as $\tau \rightarrow 0$, we have $\tilde{\eta} \rightarrow \eta$ and $\tilde{\kappa} \rightarrow \kappa$. The specific solution with $x_0 = X$ and $x_N = 0$ is a trading trajectory of the form

$$x_j = \frac{\sinh[\kappa(T - t_j)]}{\sinh(\kappa T)} X \quad \text{for } j = 0, \dots, N, \quad (17)$$

and the associated trade list

$$n_j = \frac{2 \sinh(\frac{1}{2}\kappa\tau)}{\sinh(\kappa T)} \cosh[\kappa(T - t_{j-\frac{1}{2}})] X \quad \text{for } j = 1, \dots, N, \quad (18)$$

where \sinh and \cosh are the hyperbolic sine and cosine functions, and $t_{j-\frac{1}{2}} = (j - \frac{1}{2})\tau$. These solutions (though not the efficient frontier) have been constructed previously by Grinold and Kahn (1999).

We have $n_j > 0$ for each j as long as $X > 0$. Thus, for a program of selling a large initial long position, the solution decreases *monotonically* from its initial value to zero at a rate determined by the parameter κ . For example, the optimal execution of a sell program never involves the buying of securities.¹²

¹² This can cease to be true if there is drift or serial correlation in the price movements.

For small time step τ , we have the approximate expression

$$\kappa \sim \tilde{\kappa} + O(\tau^2) \sim \sqrt{\frac{\lambda\sigma^2}{\eta}} + O(\tau) \quad \text{as } \tau \rightarrow 0. \tag{19}$$

Thus, if our trading intervals are short, κ^2 is essentially the ratio of the product of volatility and our risk intolerance to the temporary transaction cost parameter.

The expectation and variance of the optimal strategy, for a given initial portfolio size X , are then

$$\begin{aligned} E(X) &= \frac{1}{2}\gamma X^2 + \epsilon X + \tilde{\eta} X^2 \frac{\tanh(\frac{1}{2}\kappa\tau)[\tau \sinh(2\kappa T) + 2T \sinh(\kappa\tau)]}{2\tau^2 \sinh^2(\kappa T)}, \\ V(X) &= \frac{1}{2}\sigma^2 X^2 \frac{\tau \sinh(\kappa T) \cosh[\kappa(T - \tau)] - T \sinh(\kappa\tau)}{\sinh^2(\kappa T) \sinh(\kappa\tau)}, \end{aligned} \tag{20}$$

which reduce to (10)–(13) in the limits $\kappa \rightarrow 0, \infty$.

3.3 The Half-life of a Trade

We pause for a moment to discuss the meaning of the coefficient κ . We call

$$\theta = 1/\kappa$$

the trade’s “half-life”. From the discussion above, we see that the larger the value of κ and the smaller the time θ , the more rapidly the trade list will be depleted. The value θ is exactly the amount of time it takes to deplete the portfolio by a factor of e .

The definition of θ is independent of the exogenously specified execution time T ; it is determined only by the security price dynamics and the market impact factors. If the risk aversion λ is greater than zero, i.e., if the trader is risk averse, then θ is finite and independent of T . Thus, in the absence of any external time constraint ($T \rightarrow \infty$), the trader will still liquidate his position on a time scale θ . The half-life θ is the intrinsic time scale of the trade.

For given T , the ratio $\kappa T = T/\theta$ tells us what factors constrain the trade. If $T \gg \theta$, then the intrinsic half-life θ of the trade is small compared with the imposed time T : this happens because temporary costs are very small, because volatility is extremely large, or because we are very risk averse. In this case, the bulk of trading will be done well in advance of time T . Viewed on time scale T , the trajectory will look like the minimum-variance solution (12).

Conversely, if $T \ll \theta$, then the trade is highly constrained, and is dominated by temporary market impact costs. In the limit $T/\theta \rightarrow 0$, we approach the straight-line minimum-cost strategy (9).

A consequence of this analysis is that different-sized baskets of the same securities will be liquidated in exactly the same fashion, on the same time scale, provided the risk-aversion parameter λ is held constant. This may seem contrary to our intuition that large baskets are effectively less liquid, and hence should be

liquidated less rapidly than small baskets. This is a consequence of our linear market impact assumption which has the *mathematical* consequence that both variance and market impact scale quadratically with respect to portfolio size.

For large portfolios, it may be more reasonable to suppose that the temporary impact cost function has higher-order terms, so that such costs increase *super-linearly* with trade size. With nonlinear impact functions, the general framework used here still applies, but we do not obtain explicit exponential solutions as in the linear impact case. A simple practical solution to this problem is to choose different values of η (the temporary impact parameter) depending on the overall size of the problem being considered, recognizing that the model is at best only approximate.

3.4 Structure of the Frontier

An example of the efficient frontier is shown in Figure 1. The plot was produced using parameters chosen as in Section 4.4. Each point of the frontier represents a distinct strategy for optimally liquidating the same basket. The tangent line indicates the optimal solution for risk parameter $\lambda = 10^{-6}$. The trajectories corresponding to the indicated points on the frontier are shown in Figure 2.

Trajectory A has $\lambda = 2 \times 10^{-6}$; it would be chosen by a risk-averse trader who wishes to sell quickly to reduce exposure to volatility risk, despite the trading costs incurred in doing so. Trajectory B has $\lambda = 0$. We call this the naïve strategy, since it represents the optimal strategy corresponding to simply minimizing expected transaction costs without regard to variance. For a security

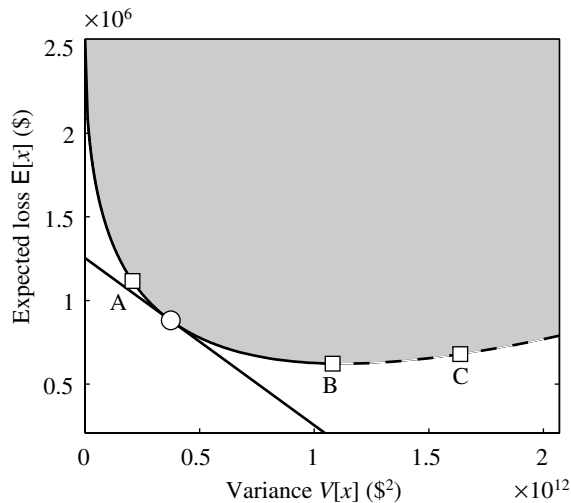


FIGURE 1. The efficient frontier. The parameters are as in Table 1. The shaded region is the set of variances and expectations attainable by some time-dependent strategy. The solid curve is the efficient frontier; the dashed curve is strategies that have higher variance for the same expected costs. Point B is the “naïve” strategy, minimizing expected cost without regard to variance. The straight line illustrates selection of a specific optimal strategy for $\lambda = 10^{-6}$. Points A, B, C are strategies illustrated in Figure 2.

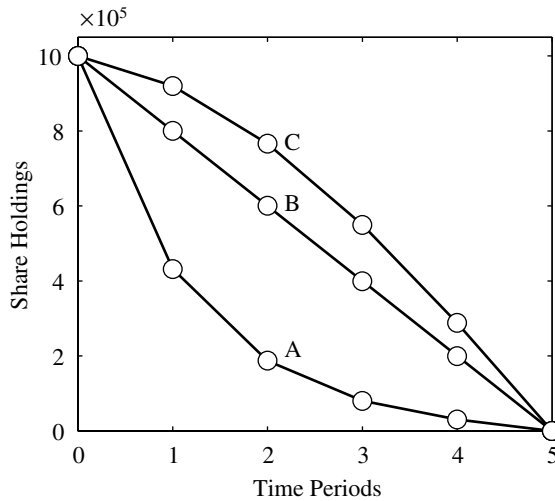


FIGURE 2. Optimal trajectories. The trajectories corresponding to the points shown in Figure 1. A: $\lambda = 2 \times 10^{-6}$, B: $\lambda = 0$, C: $\lambda = -2 \times 10^{-7}$.

with zero drift and linear transaction costs as defined above, it corresponds to a simple linear reduction of holdings over the trading period. Since drift is generally not significant over short trading horizons, the naïve strategy is very close to the linear strategy, as in Figure 2. We demonstrate below that, in a certain sense, this is *never* an optimal strategy, because one can obtain substantial reductions in variance for a relatively small increase in transaction costs. Trajectory C has $\lambda = -2 \times 10^{-7}$; it would be chosen only by a trader who likes risk. He postpones selling, thus incurring both higher expected trading costs due to his rapid sales at the end, and higher variance during the extended period that he holds the security.

4. THE RISK–REWARD TRADEOFF

We now offer an interpretation of the efficient frontier of optimal strategies in terms of the utility function of the trader. We do this in two ways: by direct analogy with modern portfolio theory employing a utility function; and by a novel approach, value-at-risk. We conclude this section with some general observations concerning the importance of utility in forming execution strategies.

4.1 Utility Function

The utility function approach amounts to establishing that each point along the efficient frontier represents the unique optimal execution strategy for a trader with a certain degree of risk aversion.

Suppose we measure utility by a smooth concave function $u(w)$, where w is our total wealth. This function may be characterized by its risk-aversion

coefficient $\lambda_u = -u''(w)/u'(w)$. If our initial portfolio is fully owned, then, as we transfer our assets from the risky stock into the alternative investment, w remains roughly constant, and we may take λ_u to be constant throughout our trading period. If the initial portfolio is highly leveraged, then the assumption of constant λ is an approximation.

For short time horizons and small changes in w , higher derivatives of $u(w)$ may be neglected. Thus choosing an optimal execution strategy is equivalent to minimizing the scalar function

$$U_{\text{util}}(x) = \lambda_u V(x) + E(x). \tag{21}$$

The units of λ_u are $\text{\$}^{-1}$: we are willing to accept an extra square dollar of variance if it reduces our expected cost by λ_u dollars.

The combination $E + \lambda V$ is precisely the one we used to construct the efficient frontier in Section 3; the parameter λ , introduced as a Lagrange multiplier, has a precise definition as a measure of our aversion to risk. Thus the methodology above used above to construct the efficient frontier likewise produces a family of optimal paths, one for each level of risk aversion.

We now return to an important point raised earlier. We have computed optimal strategies by minimizing $E + \lambda V$ as measured at the initial time; this is equivalent to maximizing utility at the outset of trading. As one trades, information arrives that could alter the optimal trading path. The following theorem eliminates this possibility.

THEOREM 1 *For a fixed quadratic utility function, the static strategies computed above are “time-homogeneous”. More precisely, given a strategy beginning at time $t = 0$ and ending at time $t = T$, the optimal strategy computed at time $t = t_k$ is simply the continuation from time $t = t_k$ to $t = T$ of the optimal strategy computed at time $t = 0$.*

Proof. For the specific solutions above, this may be seen by direct algebraic computations. Suppose that at time k , with $k = 0, \dots, N - 1$, we were to compute a new optimal strategy. Our new strategy would be precisely (17) with X replaced by x_k , T replaced by $T - t_k$, and t_j replaced by $t_j - t_k$. Using superscript (k) to denote the strategy computed at time k , we would have

$$x_j^{(k)} = \frac{\sinh[\kappa(T - t_j)]}{\sinh[\kappa(T - t_k)]} x_k \quad \text{for } j = k, \dots, N,$$

and the trade lists

$$n_j^{(k)} = \frac{2 \sinh(\frac{1}{2}\kappa\tau)}{\sinh[\kappa(T - t_k)]} \cosh[\kappa(T - t_{j-\frac{1}{2}})] X \quad \text{for } j = k + 1, \dots, N,$$

It is then apparent that if x_k is the optimal solution from (17) (with $j \mapsto k$) then $x_j^{(k)} = x_j^0$ and $n_j^{(k)} = n_j^{(0)}$, where $x_j^0 = x_j$ and $n_j^{(0)} = n_j$ comprise the strategy from (17, 18). \square

For general nonlinear impact functions $g(v)$ and $h(v)$, the optimality condition (16) is replaced by a *nonlinear* second-order difference relation. The solution $x_j^{(k)}$ beginning at a given time is determined by the two boundary values x_k and $x_N = 0$. It is then apparent that the solution does not change if we reevaluate it at later times.

More fundamentally, solutions are stable in time because, in the absence of serial correlation in the asset price movements, we have no more information about price changes at later times than we do at the initial time. Thus, the solution which was initially determined to be optimal over the entire time interval is optimal as a solution over each subinterval. This general phenomenon is well known in the theory of optimal control (Bertsekas 1976).

4.2 Value-at-Risk

The concept of value-at-risk is traditionally used to measure the greatest amount of money (maximum profit and loss) a portfolio will sustain over a given period of time under “normal circumstances”, where “normal” is defined by a confidence level.

Given a trading strategy $x = (x_1, \dots, x_N)$, we define the value-at-risk $\text{VaR}_p(x)$ of x to be the level of transaction costs incurred by trading strategy x that will not be exceeded p percent of the time. Put another way, it is the p th percentile level of transaction costs for the total cost of trading x .

Under the arithmetic Brownian motion assumption, total costs (market value minus capture) are normally distributed with known mean and variance. Thus the confidence interval is determined by the number of standard deviations λ_v from the mean by the inverse cumulative normal distribution function, and the value-at-risk for the strategy x is given by the formula

$$\text{VaR}_p(x) = \lambda_v \sqrt{V(x)} + E(x). \quad (22)$$

That is, with probability p , the trading strategy will not lose more than $\text{VaR}_p(x)$ of its market value in trading. Borrowing from the language of Perold (1988), the implementation shortfall of the execution will not exceed $\text{VaR}_p(x)$ more than a fraction p of the time. A strategy x is efficient if it has the minimum possible value-at-risk for the confidence level p .

Note that $\text{VaR}_p(x)$ is a complicated nonlinear function of the x_j composing x : we can easily evaluate it for any given trajectory, but finding the minimizing trajectory directly is difficult. However, once we have the one-parameter family of solutions which form the efficient frontier, we need only solve a one-dimensional problem to find the optimal solutions for the value-at-risk model, i.e., to find the value of λ_u corresponding to a given value of λ_v . Alternatively, we may characterize the solutions by a simple graphical procedure, or we may read off the confidence level corresponding to any particular point on the curve.

Figure 3 shows the same curve as Figure 1, except that the x -axis is the square root of variance rather than variance. In this coordinate system, lines of optimal

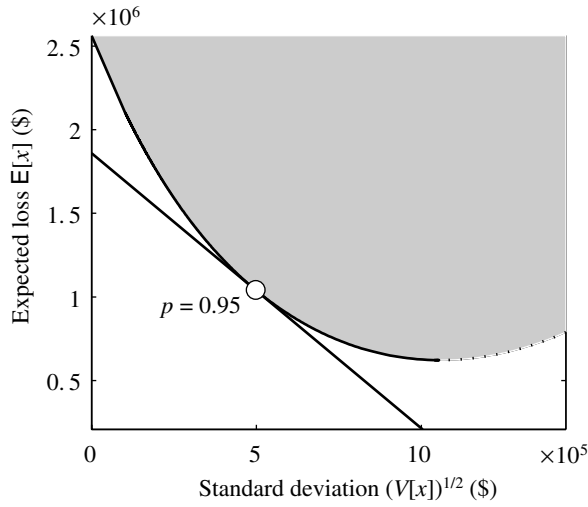


FIGURE 3. Efficient frontier for value-at-risk. The efficient frontier for parameters given in Table 1, in the plane of $V^{1/2}$ and E . The point of tangency is the optimal value-at-risk solution for a 95% confidence level, or $\lambda_v = 1.645$.

VaR have constant slope, and, for a given value of λ_v , we simply find the tangent to the curve where the slope is λ_v .

Now the question of reevaluation is more complicated and subtle. If we reevaluate our strategy halfway through the execution process, we will choose a new optimal strategy which is not the same as the original optimal one. The reason is that we now hold λ_v constant, and so λ_u necessarily changes. The value-at-risk approach has many flaws from a mathematical point of view, as recognized by Artzner *et al.* (1997). The particular problem we have uncovered here would arise in any problem in which the time of measurement is a fixed date, rather than one maintained a fixed distance in the future. We regard it as an open problem to formulate suitable measures of risk for general time-dependent problems.

Despite this shortcoming, we suggest the smallest possible value of VaR_p as an informative measure of the possible loss associated with the initial position, in the presence of liquidity effects. This value, which we shall call L-VaR, for “liquidity-adjusted value-at-risk,” depends on the time to liquidation and on the confidence level chosen, in addition to market parameters such as the impact coefficient (see Almgren and Chriss 1999).

The optimal trajectories determined by minimizing value-at-risk do *not* have the counterintuitive scaling behavior described at the end of Section 3.2: even for linear impact functions, large portfolios will be traded closer to the straight-line trajectory. This is because here the cost assigned to uncertainty scales *linearly* with the portfolio size, while temporary impact cost scales *quadratically* as before. Thus the latter is relatively more important for large portfolios.

4.3 The Role of Utility in Execution

In this section, we use the structure of the efficient frontier and the framework we have established to make some general observations about optimal execution.

The naïve strategy and execution strategies Let's restrict ourselves to the situation in which a trader has no directional view concerning the security being traded. Recall that in this case the naïve strategy is the simple straight-line strategy in which a trader breaks the block being executed into equal-sized blocks to be sold over equal time intervals. We will use this strategy as a benchmark for comparison to other strategies throughout this section.

A crucial insight is that the curve defining the efficient frontier is a smooth convex function $E(V)$ mapping levels of variance V to the corresponding minimum mean transaction cost levels.

Write (E_0, V_0) for the mean and variance of the naïve strategy. Regarding (E_0, V_0) as a point on the smooth curve $E(V)$ defined by the frontier, we have dE/dV evaluated at (E_0, V_0) is equal to zero. Thus, for (E, V) near (E_0, V_0) , we have

$$E - E_0 \approx \frac{1}{2}(V - V_0)^2 \left. \frac{d^2 E}{dV^2} \right|_{V=V_0},$$

where $d^2 E/dV^2|_{V_0}$ is positive by the convexity of the frontier at the naïve strategy.

By definition, the naïve strategy has the property that any strategy with lower cost variance has greater expected cost. However, a special feature of the naïve strategy is that a first-order decrease in variance can be obtained (in the sense of finding a strategy with lower variance) while only incurring a second-order increase in cost. That is, for small increases in variance, one can obtain much larger reductions in cost. Thus, unless a trader is risk neutral, it is always advantageous to trade a strategy that is at least to some degree "to the left" of the naïve strategy. We conclude that from a theoretical standpoint it never makes sense to trade a strictly risk-neutral strategy.

The role of liquidity An intuitive proposition is that all things being equal, a trader will execute a more liquid basket more rapidly than a less liquid basket. In the extreme, this is particularly clear. A broker given a small order to work over the course of a day will almost always execute the entire order immediately. How do we explain this? The answer is that the market impact cost attributable to rapid trading is negligible compared with the opportunity cost incurred in breaking the order up over an entire day. Thus, even if the expected return on the security over the day is zero, the perception is that the risk of waiting is outweighed by any small cost of immediacy. Now, if a trader were truly risk neutral, in the absence of any view he would always use the naïve strategy and use the allotted time fully. This would make sense because any price to pay for trading immediately is worthless if you place no premium on risk reduction.

It follows that any model that proposes optimal trading behavior should predict that more liquid baskets are traded more rapidly than less liquid ones. Now, a model that considers only the minimization of transaction costs, such as

that of Bertsimas and Lo (1998), is essentially a model that excludes utility. In such a model and under our basic assumptions, traders will trade all baskets at the same rate irrespective of liquidity, unless, that is, they have an explicit directional view on the security or the security possesses extreme serial correlation in its price movements.¹³ Another way of seeing this is that the half-life of all block executions, under the assumption of risk-neutral preferences, is infinite.

4.4 Choice of Parameters

In this section, we compute some numerical examples for the purpose of exploring the qualitative properties of the efficient frontier. Throughout the examples we will assume we have a single stock with current market price $S_0 = 50$, and that we initially hold 5 million shares, for an initial portfolio size of \$50 million. The stock will have 30% annual volatility, a 10% expected annual return of return, a bid–ask spread of $\frac{1}{8}$ and a median daily trading volume of 5 million shares.

With a trading year of 250 days, this gives daily volatility of $0.3/\sqrt{250} = 0.019$ and expected fractional return of $0.1/250 = 4 \times 10^{-4}$. To obtain our absolute parameters σ and α , we must scale by the price, so $\sigma = 0.019 \times 50 = 0.95$ and $\alpha = (4 \times 10^{-4}) \times 50 = 0.02$. Table 1 summarizes this information.

Suppose that we want to liquidate this position in 1 week, so that $T = 5$ days. We divide this into daily trades, so τ is 1 day and $N = 5$.

Over this period, if we held our initial position with no trading, the fluctuations in value would be Gaussian with a standard deviation of $\sigma\sqrt{T} = 2.12$ \$/share, and the fluctuations in value would have standard deviation $\sqrt{V} = \$2.12$ million. As expected, this is precisely twice the value of \sqrt{V} for the lowest point in Figure 3, since that point corresponds to selling along a linear trajectory rather than holding a constant amount.

TABLE 1. Parameter values for test case.

Initial stock price:	$S_0 = 50$ \$/share
Initial holdings:	$X = 10^6$ share
Liquidation time:	$T = 5$ days
Number of time periods:	$N = 5$
30% annual volatility:	$\sigma = 0.95$ (\$/share)/day ^{1/2}
10% annual growth:	$\alpha = 0.02$ (\$/share)/day
Bid–ask spread = $\frac{1}{8}$:	$\epsilon = 0.0625$ \$/share
Daily volume 5 million shares:	$\gamma = 2.5 \times 10^{-7}$ \$/share ²
Impact at 1% of market:	$\eta = 2.5 \times 10^{-6}$ (\$/share)/(share/day)
Static holdings 11 000 shares:	$\lambda_u = 10^{-6}$ /\$
VaR confidence $p = 95\%$:	$\lambda_v = 1.645$

¹³ We remind the reader that in Section 4 we note that our model in the case of linear transaction costs does not predict more rapid trading for smaller versus larger baskets of the same security. However, this is a result of choosing linear temporary impact functions and the problem goes away when one considers more realistic superlinear functions.

We now choose parameters for the temporary cost function (7). We choose $\epsilon = \frac{1}{16}$, i.e., the fixed part of the temporary costs will be one-half the bid–ask spread. For η , we will suppose that for each 1% of the daily volume we trade, we incur a price impact equal to the bid–ask spread. For example, trading at a rate of 5% of the daily volume incurs a one-time cost on each trade of $\frac{5}{8}$. Under this assumption, we have $\eta = \frac{1}{8}/(0.01 \times 5 \times 10^6) = 2.5 \times 10^{-6}$.

For the permanent costs, a common rule of thumb is that price effects become significant when we sell 10% of the daily volume. If we suppose that “significant” means that the price depression is one bid–ask spread, and that the effect is linear for smaller and larger trading rates, then we have $\gamma = \frac{1}{8}/(0.1 \times 5 \times 10^6) = 2.5 \times 10^{-7}$. Recall that this parameter gives a fixed cost independent of path.

We have chosen $\lambda = \lambda_u = 10^{-6}$. For these parameters, we have from (19) that, for the optimal strategy, $\kappa \approx 0.6/\text{day}$, so $\kappa T \approx 3$. Since this value is near 1 in magnitude, the behavior is an interesting intermediate in between the naïve extremes.

For the value-at-risk representation, we assume a 95% desired confidence level, giving $\lambda_v = 1.645$.

5. THE VALUE OF INFORMATION

Up to this point, we have discussed optimal execution under the assumption that price dynamics follow an arithmetic random walk with zero drift. Since past price evolution provides no information as to future price movements, we concluded that optimal execution trajectories can be statically determined. There are three relevant ways that random walk with zero drift may fail to correctly represent the price process.

First, the price process may have drift. For example, if a trader has a strong directional view, then he will want to incorporate this view into the liquidation strategy. Second, the price process may exhibit serial correlation. The presence of first-order serial correlation, for example, implies price moves in a given period provide nontrivial information concerning the next-period movement of the asset.¹⁴ Lastly, at the start of trading, it may be known that at some specific point in time an event will take place whose outcome will cause a material shift in the parameters governing the price process.¹⁵ For example, Brown, Harlow, and Tinic (1988) show that events cause temporary shifts in both risk and return

¹⁴ Bertsimas and Lo (1998) study a general form of this assumption, wherein an investor possesses (possibly) private information in the form of a serially correlated “information vector” that acts as a linear factor in asset returns.

¹⁵ Such event-induced parameter shifts include quarterly and annual earnings announcements, dividend announcements, and share repurchases. Event studies documenting these parameter shifts and supplying theoretical grounding for their existence include Beaver (1968), Campbell, Lo, and MacKinlay (1997), Dann (1981), Easterwood and Nutt (1999), Fama *et al.* (1969), Kalay and Loewentstein (1985), Kim and Verrecchia (1991), Patell and Wolfson (1984), and Ramaswami (1999).

of individual securities, and that the extent of these shifts depends upon the outcome of the event. In particular, securities react more strongly to bad news than to good news.

We study a stylized version of events in which a known event at a known time (e.g., an earnings announcement) has several possible outcomes. The probability of each outcome is known and the impact that a given outcome will have on the parameters of the price process is also known. Clearly, optimal strategies must explicitly use this information, and we develop methods to incorporate event-specific information into our risk–reward framework. The upshot is a piecewise strategy that trades statically up to the event, and then reacts explicitly to the outcome of the event. Thus, the burden is on the trader to determine which of the possible outcomes occurred, and then trade accordingly.

5.1 Drift

It is convenient to regard a drift parameter in a price process as a directional view of price movements. For example, a trader charged with liquidating a block of a single security may believe that this security is likely to rise. Intuitively, it would make sense to trade this issue more slowly to take advantage of this view.

To incorporate drift into price dynamics, we modify (1) to

$$S_k = S_{k-1} + \sigma\tau^{1/2}\xi_k + \alpha\tau - \tau g\left(\frac{n_k}{\tau}\right), \tag{23}$$

where α is an expected drift term. If the trading proceeds are invested in an interest-bearing account, then α should be taken as an *excess* rate of return of the risky asset.

We readily write the modified version of (8) as

$$E(x) = \frac{1}{2}\gamma X^2 - \alpha \sum_{k=1}^N \tau x_k + \epsilon \sum_{k=1}^N |n_k| + \frac{\tilde{\eta}}{\tau} \sum_{k=1}^N n_k^2. \tag{24}$$

The variance is still given by (5). The optimality condition (16) becomes

$$\frac{1}{\tau^2}(x_{k-1} - 2x_k + x_{k+1}) = \tilde{\kappa}^2(x_k - \bar{x}),$$

in which the new parameter

$$\bar{x} = \frac{\alpha}{2\lambda\sigma^2} \tag{25}$$

is the optimal level of security holding for a time-independent portfolio optimization problem. For example, the parameters of Section 4.4 give approximately $\bar{x} = 1100$ shares, or 0.11% of our initial portfolio. We expect this fraction to be very small, since, by hypothesis, our eventual aim is complete liquidation.

The optimal solution (17) becomes

$$x_j = \frac{\sinh[\kappa(T - t_j)]}{\sinh(\kappa T)} X + \left(1 - \frac{\sinh[\kappa(T - t_j)] + \sinh(\kappa t_j)}{\sinh(\kappa T)}\right) \bar{x}$$

for $j = 0, \dots, N$, (26)

with associated trade sizes

$$n_j = \frac{2 \sinh(\frac{1}{2}\kappa\tau)}{\sinh(\kappa T)} \cosh[\kappa(T - t_{j-\frac{1}{2}})] X + \frac{2 \sinh(\frac{1}{2}\kappa\tau)}{\sinh(\kappa T)} \{\cosh(\kappa t_{j-\frac{1}{2}}) - \cosh[\kappa(T - t_{j-\frac{1}{2}})]\} \bar{x}. \quad (27)$$

This trading trajectory is the sum of two distinct trajectories: the zero-drift solution as computed before, plus a “correction” which profits by capturing a piece of the predictable drift component. The size of the correction term is proportional to \bar{x} , and thus to α ; it is *independent* of the initial portfolio size X .¹⁶

The difference between this solution and the no-drift one of (17) may be understood by considering the case when $\kappa T \gg 1$, corresponding to highly liquid markets. Whereas the previous solution relaxed from X to zero on a time scale $\theta = 1/\kappa$, this one relaxes instead to the optimal static portfolio size \bar{x} . Near the end of the trading period, it sells the remaining holdings to achieve $x_N = 0$ at $t = T$.

In this case, we require $0 \leq \bar{x} \leq X$ in order for all the trades to be in the same direction. This breaks the symmetry between a buy program and a sell program; if we wanted to consider buy programs, it would be more logical to set $\alpha = 0$.

Gain due to drift Now suppose that the price dynamics is given by (23), with $\alpha > 0$, but we choose to determine a solution as though $\alpha = 0$. This situation might arise, for example, in the case where a trader is trading a security with a nondrift, but *unknowingly* assumes the security has no drift. We now explicitly calculate the loss associated with ignoring the drift term.

Write x_j^* for the optimal solution (26) with $\alpha > 0$, and x_j^0 for the suboptimal solution (17), or (26) with $\alpha = 0$. Also write $E^*(X)$ and $V^*(X)$ for the optimal expected cost and its variance, measured by (24) and (5) with $x_j = x_j^*$; let us write $E^0(X)$ and $V^0(X)$ for the suboptimal values of (24) and (5) evaluated with $x_j = x_j^0$. The corresponding objective functions are $U^*(X) = E^*(X) + \lambda V^*(X)$ and $U^0(X) = E^0(X) + \lambda V^0(X)$. Then we define the *gain due to drift* to be the

¹⁶ To place this in an institutional framework, consider a program trading desk that sits in front of customer flow. If this desk were to explicitly generate alphas on all securities that flow through the desk in an attempt to, say, hold securities with high alpha and sell securities more rapidly with low alpha, the profit would not scale in proportion to the average size of the programs. Rather, it would only scale with the number of securities that flow through the desk. An even stronger conclusion is that, since the optimal strategy disconnects into a static strategy unrelated to the drift term and a second strategy related to the drift term, there is no particular advantage to restricting trading to securities which the desk currently holds positions in.

difference $U^0(X) - U^*(X)$; this is the amount by which taking account of the drift term reduces our cost and variance. Clearly $U^0 - U^* \geq 0$, since x^* is the unique optimal strategy for the model with $\alpha > 0$.

Now, the value of the terms in U^0 that come from (8) and (5) is only *increased* by going from x^0 to x^* , since x^0 and not x^* was the optimum strategy with $\alpha = 0$. Therefore, an *upper bound* for the gain is

$$U^0 - U^* \leq \alpha\tau \sum_{k=1}^N (x_k^* - x_k^0).$$

That is, in response to the positive drift, we should increase our holdings throughout trading. This reduces our net cost by the amount of asset price increase we capture, at the expense of slightly increasing our transaction costs and volatility exposure. An upper bound for the possible benefit is the amount of increase we capture.

But $x_k^* - x_k^0$ is just the term in square brackets in (26) times \bar{x} , which is clearly independent of X . Indeed, we can explicitly evaluate this difference to find

$$\alpha\tau \sum_{k=1}^N (x_k^* - x_k^0) = \alpha\bar{x}T \left(1 - \frac{\tau \tanh(\frac{1}{2}\kappa T)}{T \tanh(\frac{1}{2}\kappa\tau)} \right).$$

Since $\tanh(x)/x$ is a positive decreasing function, this quantity is positive and bounded above by $\alpha\bar{x}T$, the amount you would gain by holding portfolio \bar{x} for time T . Any reasonable estimates for the parameters show that this quantity is negligible compared with the impact costs incurred in liquidating an institutional-size portfolio over a short period.

5.2 Serial Correlation

Now let us suppose that the asset prices exhibit serial correlation, so that at each period we discover a component of predictability of the asset price in the next period. In the model (1), with mean drift $\alpha = 0$, we now suppose that the ξ_k are serially correlated with period-to-period correlation ρ ($|\rho| < 1$). We can determine ξ_k at time k based on the observed change $S_k - S_{k-1}$ and our own sale n_k .

With serial correlation, the optimal strategy is no longer a static trajectory determined in advance of trading; since each price movement gives us some information about the immediate future price movements, the optimal trade list can be determined only one period at a time. Thus a fully optimal solution requires the use of dynamic programming methods. However, since information is still roughly local in time, we can estimate the gain attainable by an optimal strategy. We state the conclusion in advance of our estimate. The value of information contained in price movements due to serial correlation is independent of the size of the portfolio being traded. The calculation below lends intuition to this counterintuitive statement.

Consider two consecutive periods, during which our base strategy has us selling the same number, n , of shares in each period. With the linear price

impact model, in each period we depress the price by $\epsilon + \eta(n/\tau)$ \$/share. We pay this cost on each sale of n shares, so the total cost due to market impact per period is

$$\text{per-period impact cost of "smooth" strategy} = \left(\epsilon + \eta \frac{n}{\tau} \right) n.$$

Suppose we have some price information due to correlations. If we know ξ_k at the previous period, then the predictable component of the price change is roughly $\rho\sigma\tau^{1/2}$. If we shift the sale of δn shares from one period to the next, then the amount of extra money we can earn per period is roughly

$$\text{per-period gain by adapting to correlations} \approx \rho\sigma\tau^{1/2} \delta n.$$

But this adaptation increases our impact costs. After the shift, in the first period the price depression is $\epsilon + \eta[(n - \delta n)/\tau]$, while in the second period it is $\epsilon + \eta[(n + \delta n)/\tau]$. We pay these costs on $n - \delta n$ and $n + \delta n$ shares, respectively, so the market impact cost per period is now given by

per-period cost of adapted strategy

$$\begin{aligned} &= \frac{1}{2} \left[\left(\epsilon + \eta \frac{n - \delta n}{\tau} \right) (n - \delta n) + \left(\epsilon + \eta \frac{n + \delta n}{\tau} \right) (n + \delta n) \right] \\ &= \left(\epsilon + \eta \frac{n}{\tau} \right) n + \frac{\eta}{\tau} \delta n^2. \end{aligned}$$

To determine how many shares we should shift, we solve the quadratic optimization problem

$$\max_{\delta n} \left(\rho\sigma\tau^{1/2} \delta n - \frac{\eta}{\tau} \delta n^2 \right).$$

We readily find the optimal δn

$$\delta n^* = \frac{\rho\sigma\tau^{3/2}}{2\eta}$$

and the maximum possible gain

$$\text{maximum per-period gain} = \frac{\rho^2\sigma^2\tau^2}{4\eta}. \tag{28}$$

This heuristic analysis can be confirmed by a detailed dynamic programming computation, which accounts for optimal shifts across multiple periods.¹⁷

¹⁷ We briefly explain the limitation of this approximation. When ρ is close to zero, this approximation is extremely close to correct, because the persistence of serial correlation effects dies down very quickly after the first period. When $|\rho|$ is too large to ignore, the approximation is too small for $\rho > 0$. That is, equation (28) understates the possible gains over ignoring serial correlation. Conversely, when $\rho < 0$, (28) overstates the possible gains due to serial correlation. As the former is empirically more frequently the case, we assert that (28) is useful for bounding the possible gains in most situations available from serial correlation.

TABLE 2. Gain from serial correlation. Approximate gain, in dollars per day, earned by trading to serial correlation of price movements, for different temporary price impact coefficients and different autocorrelation coefficients. The parameter ρ is the serial correlation across a 10-minute time period. The temporary impact coefficient is estimated by specifying the percentage of the market average volume we can trade before incurring one bid–ask spread in impact cost.

ρ	$\eta =$	1%	2%	5%	10%
		2.5	1.25	0.5	0.25×10^{-6}
-0.1		19	38	96	192
-0.2		64	129	321	643
-0.5		257	514	1286	2571

Note that both the size of the adaptation, and the resulting gain, are independent of the amount n of shares that we would sell in the unadapted strategy. That is, they are independent of the size of our initial portfolio. Instead, the binding constraint is the liquidity of the security being traded, and the magnitude of the correlation coefficient. The more information available due to correlation and the more liquid the security, the more overall gain that is available due to adapting the strategy to correlations.¹⁸

To indicate the size of gains that can be expected by adapting to correlations, we give a numerical example based on that of Section 4.4. We take parameters as in Section 4.4, except that, for the temporary impact parameter η , we suppose that steady trading at a rate of either 1%, 2%, 5%, or 10% of the market’s average volume requires a price concession of one bid–ask spread, or 25 basis points. We suppose that, over a 10-minute time period ($\tau = 0.0256$ day), the correlation in successive price motions is $\rho = 0.5, 0.2, \text{ or } 0.1$. Table 2 shows the

¹⁸ This result is especially simple because we are assuming linear impact functions. Let us briefly show what happens in the more general case of a nonlinear impact function $h(v) = h(n/\tau)$. The cost per period due to market impact is

$$\begin{aligned} \text{impact cost of adapted strategy} &= \frac{1}{2} \left[h\left(\frac{n - \delta n}{\tau}\right)(n - \delta n) + h\left(\frac{n + \delta n}{\tau}\right)(n + \delta n) \right] \\ &\approx h\left(\frac{n}{\tau}\right)n + \left[\frac{1}{2} h''\left(\frac{n}{\tau}\right) \frac{n}{\tau} + h'\left(\frac{n}{\tau}\right) \right] \frac{\delta n^2}{\tau} \end{aligned}$$

for small δn . Now the optimal shift and maximum gain are

$$\delta n^* = \frac{\rho \sigma \tau^{3/2}}{v h'' + 2h'}, \quad \text{maximum gain per period} = \frac{\rho^2 \sigma^2 \tau^2}{2(v h'' + 2h')},$$

where h' and h'' are evaluated at the base sale rate $v = n/\tau$. The linear case is recovered by setting $h(v) = \epsilon + \eta v$; this has the special property that h' is independent of v and $h'' = 0$.

In general, suppose $h(v) \sim O(v^\alpha)$ as $v \rightarrow \infty$. We require $\alpha > 0$, so that $h(v)$ is increasing: selling more shares always pushes the price down more. The marginal cost is $h'(v) \sim O(v^{\alpha-1})$; $\alpha > 1$ corresponds to an increasing marginal impact, $\alpha < 1$ to a decreasing marginal impact. Then the per-period cost we pay on our base strategy is $\sim O(v^{\alpha+1})$ for large initial portfolios and hence large rates of sale. The marginal gain from adapting to correlation is $\sim O(v^{\alpha-1})$ in the same limit.

Thus, while for nonlinear impact functions, the gain available by adapting to correlation may increase with v if $\alpha > 1$, it is always asymptotically smaller than the impact cost paid on the underlying program. Thus the correlation gain can always be neglected for sufficiently large portfolios.

resulting gain in dollars per day (39 periods) given by the exact formula from a dynamic programming computation.¹⁹ Only in the case of an extremely liquid stock with extremely high serial correlation are these gains significant for institutional trading.

5.3 Parameter Shifts

We now discuss the impact on optimal execution of scheduled news events such as earnings or dividend announcements. Such events have two key features which make them an important object of study. First, the outcome of the event determines a shift in the parameters governing price dynamics (see Brown, Harlow, and Tinic 1988; Ramaswami 1999; Easterwood and Nutt 1999). Second, the fact that they are scheduled increases the likelihood that we can detect what the true outcome of the event is. We formalize this situation below and give explicit formulas for price trajectories before and after the event takes place.

Suppose that at some time T_* , between now and the specified final time T , an event will occur, the outcome of which may or may not cause a shift in the parameters of price dynamics. We use the term *regime* or *parameter set* to refer to the collection $R = \{\sigma, \eta, \dots\}$ of parameters that govern price dynamics at any particular time, and events of interest to us are those that have the possibility of causing *parameter shifts*.

Let $R_0 = \{\sigma_0, \eta_0, \dots\}$ be the parameters of price dynamics at the time we begin to liquidate. Suppose the market can shift to one of p possible new sets of parameters R_1, \dots, R_p , so that R_j is characterized by parameters σ_j, η_j, \dots for $j = 1, \dots, p$. We suppose also that we can assign probabilities to the possible new states, so that P_j is the probability that regime R_j occurs. These probabilities are *independent* of the short-term market fluctuations represented by ξ_k . Of course, it is possible that some R_j have the same values as R_0 , in which case P_j is the probability that no change occurs.

We consider a dynamic trading strategy that yields globally optimal strategies in the presence of a parameter shift at time T_* . Suppose that $T_* = t_s = \sigma\tau$. We precompute an initial trajectory $x^0 = (x_0^0, \dots, x_s^0)$, with $x_0^0 = X$; we write $X_* = x_s^0$. We also compute a family of trajectories $x^j = (x_s^j, \dots, x_N^j)$ for $j = 1, \dots, p$, all of which have $x_s^j = X_*$ and $x_N^j = 0$. We follow trajectory x^0 until the time of the shift. Once the shift occurs, we assume that we can quickly identify the outcome of the event and the new set of parameters governing price dynamics. With this settled, we complete trading using the corresponding trajectory x^j . We shall show that we can determine each trajectory using static optimization, although we cannot choose which one to use until the event occurs. Furthermore, the starting trajectory x^0 will *not be the same* as the trajectory we would choose if we believed that regime R_0 would hold through the entire time T .

To determine the trajectories x^0, x^1, \dots, x^p , we reason as follows. Suppose we

¹⁹ For this comparison, we have neglected permanent impact, setting $\gamma = 0$, so $\bar{\eta} = \eta$.

fixed the common value $X_* = x_s^0 = x_s^j$. Then, by virtue of the independence of the regime shift itself from the security motions, the optimal trajectories conditional on the value of X_* are simply those that we have already computed with a small modification to include the given nonzero final value. We can immediately write

$$x_k^0 = X \frac{\sinh[\kappa_0(T_* - t_k)]}{\sinh(\kappa_0 T_*)} + X_* \frac{\sinh(\kappa_0 t_k)}{\sinh(\kappa_0 T_*)} \quad \text{for } k = 0, \dots, s,$$

where κ_0 is determined from σ_0, η_0, \dots . This trajectory is determined in the same way as in Section 3.2; it is the unique combination of exponentials $\exp(\pm\kappa_0 t)$ that has $x_0^0 = X$ and $x_s^0 = X_*$. Similarly,

$$x_k^j = X_* \frac{\sinh[\kappa_j(T - t_k)]}{\sinh[\kappa_j(T - T_*)]} \quad \text{for } k = s, \dots, N; j = 1, \dots, p.$$

Thus we need only determine X_* .

To determine X_* , we determine the expected loss and its variance of the combined strategy. Let E_0 and V_0 denote the expectation and variance of the loss incurred by trajectory x^0 on the first segment $k = 0, \dots, s$. For $j = 1, \dots, p$, let E_j and V_j denote the expectation and variance of loss incurred by trajectory x^j on the second segment $k = s, \dots, N$. These quantities can readily be determined using

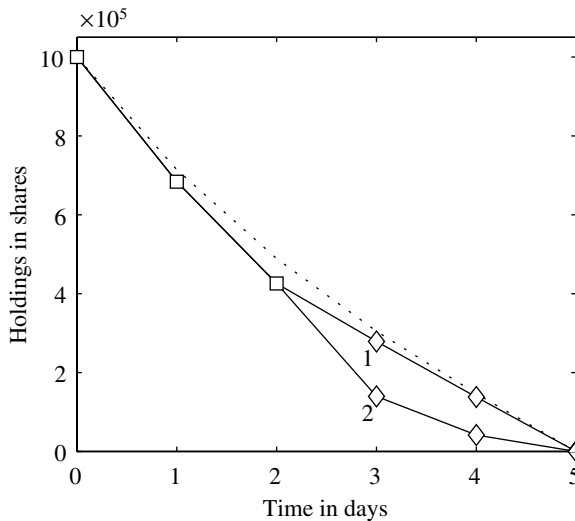


FIGURE 4. Optimal strategy with parameter shift. Parameters are as in Section 4.4, except that the initial volatility is 10% annually. After close of trading on the second day, an announcement is expected which will cause the volatility either to decrease to 5% (case 1) or to increase to 40% (case 2) for the remainder of the trading period. The solid lines show the risk-averse optimal strategy including the two possible branches to be taken following the announcement. The dashed line is the strategy that would be followed under the assumption that the initial parameters will last throughout the whole liquidation.

the formulas (8) and (5), respectively. Then, by virtue of the independence of the regime shift and the security motions, the expected loss of the compound strategy is

$$E = E_0 + P_1 E_1 + \cdots + P_p E_p,$$

and its variance is

$$V = V_0 + P_1 V_1 + \cdots + P_p V_p + \frac{1}{2} \sum_{i,j=1}^p P_i P_j (E_i - E_j)^2.$$

We can now do a one-variable optimization in X_* to minimize $E + \lambda V$. An example is shown in Figure 4.

6. CONCLUSIONS

The central feature of our analysis has been the construction of an *efficient frontier* in a two-dimensional plane whose axes are the expectation of total cost and its variance. Regardless of an individual's tolerance for risk, the only strategies which are candidates for being optimal are found in this one-parameter set. For linear impact functions, we give complete analytical expressions for the strategies in this set.

Then considering the details of risk aversion, we have shown how to select an optimal point on this frontier either by classic mean–variance optimization, or by the more modern concept of value-at-risk. These solutions are easily constructed numerically, and easily interpreted graphically by examination of the frontier.

Several conclusions of practical importance follow from this analysis:

1. Because the set of attainable strategies, and hence the efficient frontier, are generally a *smooth* and *convex*, a trader who is at all risk averse should never trade according to the “naïve” strategy of minimizing expected cost. This is because in the neighborhood of that strategy, a first-order reduction in variance can be obtained at the cost of only a second-order increase in expected cost.
2. We also observe that this careful analysis of the costs and risks of liquidation can be used to give a more precise characterization of the risk of holding the initial portfolio. For example, we can define “liquidity-adjusted value-at-risk” (L-VaR) to be, for a given time horizon, the minimum VaR of any static liquidation strategy.

Although it may seem counterintuitive that optimal strategies can be determined in advance of trading, in Section 5 we have argued that only very small gains can be obtained by adapting the strategy to information as it is revealed.

The model may be extended in several interesting ways:

- *Continuous time*: The limit $\tau \rightarrow 0$ is immediate in all our solutions. The trading strategy is characterized by a holdings function $x(t)$ and a *trading rate*

$v(t) = \lim_{\tau \rightarrow 0} n_k / \tau$. The minimum-variance strategy of Section 2 has infinite cost, but the optimal strategies for finite λ have finite cost and variance. However, this limit is at best a mathematical convenience, since our market model is implicitly a “coarse-grained” description of the real dynamics.

- *Nonlinear cost functions:* The conceptual framework we have outlined is not restricted to the linear temporary and permanent impact functions (6) and (7), though the exact exponential solutions of Section 3 are special to that case. For nonlinear functions $g(v)$ and $h(v)$ that satisfy suitable convexity conditions, optimal risk-averse trajectories are found by solving a nonquadratic optimization problem; the difficulty of this problem depends on the specific functional form chosen.
- *Time-varying coefficients:* Our framework also covers the case in which the volatility, market impact parameters, and perhaps expected drift are time-dependent, as long as their values are known at the start of liquidation; finding the optimal strategy entails solving a linear system of size equal to the number of time periods (times the number of stocks, for a portfolio problem). One example in which this is useful is if the price is expected to jump either up or down on a known future date (an earnings announcement, say), as long as we have a good estimate of the expected *size* of this jump.

We hope that these extensions will lead to further useful insights.

APPENDIX: MULTIPLE-SECURITY PORTFOLIOS

With m securities, our position at each moment is a column vector $x_k = [x_{1k}, \dots, x_{mk}]^T$, where T denotes transpose. The initial value $x_0 = X = [X_1, \dots, X_m]^T$, and our trade list is the column vector $n_k = x_{k-1} - x_k$. If $x_{jk} < 0$, then security j is held short at time t_k ; and if $n_{jk} < 0$, then we are buying security j between t_{k-1} and t_k .

1. Trading Model

We assume that the column vector of security prices S_k follows a multidimensional arithmetic Brownian random walk with zero drift. Its dynamics is again written as in (1), but now $\xi_k = [\xi_{1k}, \dots, \xi_{rk}]^T$ is a vector of r independent Brownian increments, with $r \leq m$, with σ an $m \times r$ additive volatility matrix. $C = \sigma\sigma^T$ is the $m \times m$ symmetric positive-definite variance-covariance matrix.

The permanent impact $g(v)$ and the temporary impact $h(v)$ are vector functions of a vector. We consider only the linear model

$$g(v) = \Gamma v, \quad h(v) = \epsilon \operatorname{sgn}(v) + H v,$$

where Γ and H are $m \times m$ matrices, and ϵ is an $m \times 1$ column vector multiplied componentwise by $\operatorname{sgn}(v)$. The (i, j) th element of Γ and of H represents the price depression on security i caused by selling security j at a unit rate. We require

that H be positive-definite, since if there were a nonzero v with $v^T H v \leq 0$ then by selling at rate v we would obtain a net benefit (or at least lose nothing) from instantaneous market impact. We do not assume that H and Γ are symmetric.

The market value of our initial position is $X^T S_0$. The loss in value incurred by a liquidation profile x_1, \dots, x_N is calculated just as in (3), and we find again, as in (4) and (5),

$$\begin{aligned} E(x) &= \epsilon^T |X| + \sum_{k=1}^N \tau x_k^T \Gamma v_k + \sum_{k=1}^N \tau v_k^T H v_k \\ &= \epsilon^T |X| + \frac{1}{2} X^T \Gamma^S X + \sum_{k=1}^N \tau v_k^T \tilde{H} v_k + \sum_{k=1}^N \tau x_k^T \Gamma^A v_k, \end{aligned} \tag{29}$$

$$V(x) = \sum_{k=1}^N \tau x_k^T C x_k, \tag{30}$$

with $\tilde{H} = H^S - \frac{1}{2} \tau \Gamma^S$. We use the superscripts S and A to denote the symmetric and antisymmetric parts, respectively, so $H = H^S + H^A$ and $\Gamma = \Gamma^S + \Gamma^A$, with

$$H^S = \frac{1}{2}(H + H^T), \quad \Gamma^S = \frac{1}{2}(\Gamma + \Gamma^T), \quad \Gamma^A = \frac{1}{2}(\Gamma - \Gamma^T).$$

Note that H^S is positive-definite as well as symmetric. We shall assume that τ is sufficiently small so that \tilde{H} is positive-definite and hence invertible. We have assumed that each component of v has a consistent sign throughout the liquidation.

Despite the multidimensional complexity of the problem, the set of all outcomes is completely described by these two scalar functionals. The utility function and value-at-risk objective functions are still given in terms of E and V by (21) and (22).

2. Optimal Trajectories

Determination of the optimal trajectory for the portfolio is again a linear problem. We readily find that stationarity of $E + \lambda V$ with respect to variation of x_{jk} gives the multidimensional extension of (16):

$$\frac{x_{k-1} - 2x_k + x_{k+1}}{\tau^2} = \lambda \tilde{H}^{-1} C x_k + \tilde{H}^{-1} \Gamma^A \frac{x_{k-1} - x_{k+1}}{2\tau} \quad \text{for } k = 1, \dots, N - 1. \tag{31}$$

Since $\tilde{H}^{-1} C$ is not necessarily symmetric and $\tilde{H}^{-1} \Gamma^A$ is not necessarily anti-symmetric, despite the symmetry of \tilde{H} , it is convenient to define a new solution variable y by

$$y_k = \tilde{H}^{1/2} x_k.$$

We then have

$$\frac{y_{k-1} - 2y_k + y_{k+1}}{\tau^2} = \lambda A y_k + B \frac{y_{k-1} - y_{k+1}}{2\tau},$$

in which

$$A = \tilde{H}^{-1/2} C \tilde{H}^{-1/2} \quad \text{and} \quad B = \tilde{H}^{-1/2} \Gamma^A \tilde{H}^{-1/2}$$

are symmetric positive-definite and antisymmetric, respectively. This is a linear system in $(N - 1)m$ variables which can easily be solved numerically.

3. Explicit Solution for Diagonal Model

To write explicit solutions, we make the *diagonal* assumption that trading in each security affects the price of that security only and no other prices. This corresponds to taking Γ and H to be diagonal matrices, with

$$\Gamma_{jj} = \gamma_j, \quad H_{jj} = \eta_j. \tag{32}$$

We require that each $\gamma_j > 0$ and $\eta_j > 0$. With this assumption, the number of coefficients we need in the model is proportional to the number of securities, and their values can plausibly be estimated from available data. For Γ and H diagonal, $E(x)$ decomposes into a collection of sums over each security separately, but the covariances still couple the whole system.

In particular, since Γ is now symmetric, we have $\Gamma^A = 0$ and hence $B = 0$; further, \tilde{H} is diagonal with

$$\tilde{H}_{jj} = \eta_j \left(1 - \frac{\gamma_j \tau}{2\eta_j} \right).$$

We require these diagonal elements to be positive, which will be the case if $\tau < \min_j(2\eta_j/\gamma_j)$. Then the inverse square root is trivially computed.

For $\lambda > 0$, the matrix λA has a complete set of positive eigenvalues, which we denote by $\tilde{\kappa}_1^2, \dots, \tilde{\kappa}_m^2$, and a complete set of orthonormal eigenvectors which form the columns of an orthogonal matrix U . The solution in the diagonal case is a combination of exponentials $\exp(\pm\kappa_j t)$, with

$$\frac{2}{\tau^2} [\cosh(\kappa_j \tau) - 1] = \tilde{\kappa}_j^2.$$

With $y_k = Uz_k$, we may write

$$z_{jk} = \frac{\sinh[\kappa_j(T - t_k)]}{\sinh(\kappa_j T)} z_{j0},$$

in which the column vector z_0 is given by

$$z_0 = U^T y_0 = U^T \tilde{H}^{1/2} X.$$

Undoing the above changes of variables, we have finally

$$x_k = \tilde{H}^{-1/2} Uz_k.$$

With multiple securities, it is possible for some components of the velocity to be nonmonotonic in time. For example, if our portfolio includes two securities

TABLE 3. Parameters for two-security example.

Share price	$\begin{bmatrix} \$50 \\ \$100 \end{bmatrix}$
Daily volume	$\begin{bmatrix} 5 \\ 20 \end{bmatrix}$ million
Annual variance	$\begin{bmatrix} 30\% & 10\% \\ 10\% & 15\% \end{bmatrix}$

whose fluctuations are highly correlated, with one much more liquid than the other, an optimal strategy directs us to rapidly go short in the liquid one to reduce risk, while we slowly reduce the whole position to zero. In this case, the above expressions are not exactly correct because of the changing sign of the cost associated with the bid–ask spread. Since this effect is probably very small, a reasonable approach in such a case is to set $\epsilon = 0$.

4. Example

We now briefly consider an example with only two securities. For the first security, we take the same parameters as for the example of Section 4.4. We choose the second security to be more liquid and less volatile, with a moderate amount of correlation. These parameters are summarized in Table 3. From this market data, we determine the model parameters just as in Section 4.4.

Our initial holdings are 10 million shares in each security; we take a time horizon $T = 5$ days and give ourselves $N = 5$ periods. Figure 5 shows the efficient frontier in the (V, E) -plane. The three trajectories corresponding to the points A, B, C are shown in Figure 6.

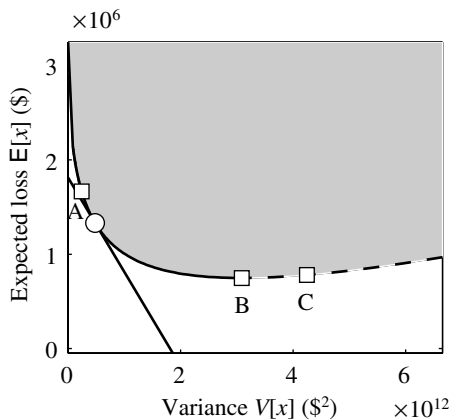


FIGURE 5. Efficient frontier for two securities. The straight line is the optimal point for $\lambda = 10^{-6}$; the three points A, B, C are optimal strategies for different values as illustrated in Figure 6.

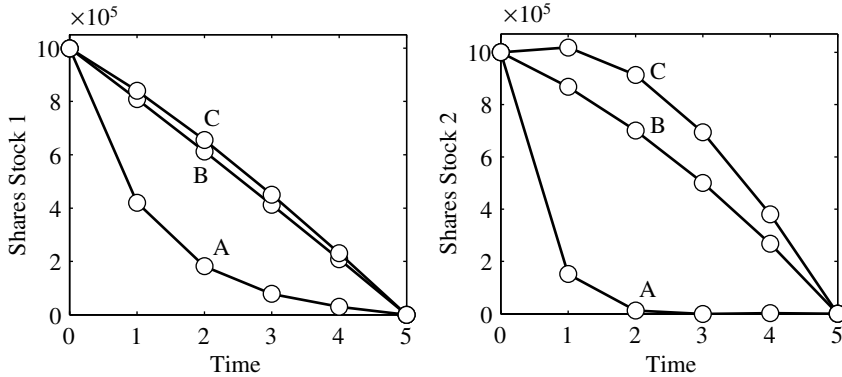


FIGURE 6. Optimal trajectories for two securities. As in Figure 5, A: $\lambda = 2 \times 10^{-6}$; B: the naïve strategy with $\lambda = 0$; C: $\lambda = -5 \times 10^{-8}$.

For these example parameters, the trajectory of security 1 is almost identical to its trajectory in the absence of security 2 (Section 4.4). Increasing the correlation of the two securities increases the interdependence of their trajectories; we expect that relaxing the assumption of diagonal transaction costs would have the same effect.

Acknowledgements

We thank Andrew Alford, Alix Baudin, Mark Carhart, Ray Iwanowski, and Giorgio De Santis (Goldman Sachs Asset Management), Robert Ferstenberg (ITG), Michael Weber (Merrill Lynch), Andrew Lo (Sloan School, MIT), and George Constantinides (Graduate School of Business, University of Chicago) for helpful conversations. This paper was begun while the first author was at the University of Chicago, and the second author was first at Morgan Stanley Dean Witter and then at Goldman Sachs Asset Management.

REFERENCES

- Admati, A., and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies*, **1**, 3–40.
- Almgren, R., and Chriss, N. (1999). Value under liquidation. *Risk*, **12**(12), 61–63.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1997). Thinking coherently. *Risk*, **10**(11), 68–71.
- Beaver, W. (1968). The information content of annual earnings announcements. In: *Empirical Research in Accounting: Selected Studies*. Supplement to *Journal of Accounting Research*, **6**, 67–92.
- Bertsekas, D. P. (1976). *Dynamic Programming and Stochastic Control*. Academic Press.

- Bertsimas, D., and Lo, A. W. (1998). Optimal control of liquidation costs. *Journal of Financial Markets*, **1**, 1–50.
- Brown, K., Harlow, W., and Tinic, S. (1988). Risk aversion, uncertain information and market efficiency. *Journal of Financial Economics*, **22**, 355–385.
- Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *Econometrics of Financial Markets*. Princeton University Press.
- Chan, L. K. C., and Lakonishok, J. (1993). Institutional trades and intraday stock price behavior. *Journal of Financial Economics*, **33**, 173–199.
- Chan, L. K. C., and Lakonishok, J. (1995). The behavior of stock prices around institutional trades. *Journal of Finance*, **50**, 1147–1174.
- Charest, G. (1978). Dividend information, stock returns and market efficiency—II. *Journal of Financial Economics*, **6**, 297–330.
- Dann, L. (1981). Common stock repurchases: An analysis of returns to bondholders and stockholders. *Journal of Financial Economics*, **9**, 113–138.
- Easterwood, J. C., and Nutt, S. R. (1999). Inefficiency in analysts' earning forecasts: Systematic misreaction or systematic optimism? *Journal of Finance*, **54**, 1777–1797.
- Fama, E., Fisher, L., Jensen, M., and Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, **10**, 1–21.
- Grinold, R. C., and Kahn, R. N. (1999). *Active Portfolio Management* (2nd edn.), Chap. 16, pp. 473–475. McGraw-Hill.
- Holthausen, R. W., Leftwich, R. W., and Mayers, D. (1987). The effect of large block transactions on security prices: A cross-sectional analysis. *Journal of Financial Economics*, **19**, 237–267.
- Holthausen, R. W., Leftwich, R. W., and Mayers, D. (1990). Large-block transactions, the speed of response, and temporary and permanent stock-price effects. *Journal of Financial Economics*, **26**, 71–95.
- Kalay, A., and Loewenstein, U. (1985). Predictable events and excess returns: The case of dividend announcements. *Journal of Financial Economics*, **14**, 423–449.
- Keim, D. B., and Madhavan, A. (1995). Anatomy of the trading process: Empirical evidence on the behavior of institutional traders. *Journal of Financial Economics*, **37**, 371–398.
- Keim, D. B., and Madhavan, A. (1997). Transactions costs and investment style: An inter-exchange analysis of institutional equity trades. *Journal of Financial Economics*, **46**, 265–292.
- Kim, O., and Verrecchia, R. (1991). Market reaction to anticipated announcements. *Journal of Financial Economics*, **30**, 273–310.
- Kraus, A., and Stoll, H. R. (1972). Price impacts of block trading on the New York Stock Exchange. *Journal of Finance*, **27**, 569–588.
- Krinsky, I., and Lee, J. (1996). Earnings announcements and the components of the bid–ask spread. *Journal of Finance*, **51**, 1523–1535.

- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, **53**, 1315–1336.
- Lee, C. M. C., Mucklow, B., and Ready, M. J. (1993). Spreads, depths, and the impact of earnings information: An intraday analysis. *Review of Financial Studies*, **6**, 345–374.
- Lo, A. W., and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, **1**, 41–66.
- Morse, D. (1981). Price and trading volume reaction surrounding earnings announcements: A closer examination. *Journal of Accounting Research*, **19**, 374–383.
- Patell, J. M., and Wolfson, M. A. (1984). The intraday speed of adjustment of stock prices to earnings and dividend announcements. *Journal of Financial Economics*, **13**, 223–252.
- Perold, A. F. (1988). The implementation shortfall: Paper versus reality. *Journal of Portfolio Management*, **14**, Spring, 4–9.
- Ramaswami, M. (1999). Stock volatility declines after earnings are announced, honest. Global Weekly Investment Strategy, Lehman Brothers.