Execution Risk

It's the same as investment risk.

Robert Engle and Robert Ferstenberg

The trade-off between risk and return is the central feature of both academic and practitioner finance. Financial managers must decide which risks to take and how much to take. This involves measurement of the risks and modeling the relation between risk and return. This setting is the classic framework for optimal portfolio construction pioneered by Markowitz [1952] and reflected today in all textbooks.

Although much attention has been paid to the cost of trading, little has been devoted to the risks of trading. Analysis typically focuses on the costs of executing a single trade or in some cases a sequence of trades. Almgren and Chriss [1999, 2000], Grinold and Kahn [1999], Almgren [2003], and Obizhaeva and Wang [2005] notably develop models to focus on the risk of as well the mean cost of execution.

What is this risk? There are many ways to execute a trade, and these have different outcomes. For example, a small buy order submitted as a market order will most likely execute at the asking price. If it is submitted as a limit order at a lower price, the execution will be uncertain. If it does not execute and is converted to a market order at a later time or to another limit order, the ultimate price at which the order is executed will be a random variable.

This random variable can be thought of as having both a mean and a confidence interval. In a mean-variance framework, often we consider the mean to be the expected cost, while the variance is the measure of the risk of this transaction.

More generally for large trades, the customer can either execute trades immediately by sending them to a block desk or other intermediary that will take on the risk, or execute a sequence of smaller trades. These might

Robert Engle
is the Michael Armelin professor of Finance in the Stern School of Business at New York University in New York City.
rengle@stern.nyu.edu

Robert Ferstenberg
is a managing director at Morgan Stanley in New York City.
Robert.ferstenberg@morganstanley.com

be planned and executed by a floor broker, or an in-house or institutional trader, or by an algorithmic trading system. The ultimate execution will be a random variable primarily because some portions of the trade will be executed after prices have moved.

Delay in trading introduces price risk when prices move more than could be anticipated as a natural response to the trade itself. Different trading strategies will have different probability distributions of the costs, so customers will need to choose the trading strategy that is optimal for them.

We address the relation between the risk-return trade-off that is well understood for investment and the risk-return trade-off that arises in execution. For example, would it be sensible to trade in a risk-neutral fashion when a portfolio is managed very conservatively? Will execution risk on different names and at different times average out to zero? Should the transaction strategy depend on what else is in the portfolio? Should execution risks be hedged?

We integrate the portfolio decision and the execution decision into a single problem to show how to optimize these choices jointly. In this way we will answer the four questions posed above and several others.

We first introduce the theoretical optimization problems and synthesize them into one single problem. The implications of trading strategy for the Sharpe ratio are developed. A specific assumption on price dynamics produces specific solutions for optimal trades. This also shows the role of non-traded assets. The model may include sophisticated dynamics allowing reversals. Finally, we discuss measures of liquidity risk.

PORTFOLIO OPTIMIZATION AND TRADE OPTIMIZATION

The classic portfolio problem in its simplest form seeks portfolios with minimum variances that attain at least a specific expected return. If y is the portfolio value, the problem is simply stated as:

$$\min_{s.t. E(y) \geq \mu_0} V(y) \tag{1}$$

In this expression, the mechanism for creating the portfolio is not explicitly indicated, nor is the time period specified. Let us suppose that the portfolio is evaluated over the period $(0, T)$ and that we define the dollar returns on each period $t = 0, 1, \ldots, T$. The dollar return on the full period is the sum of the dollar returns on the individual periods. The variance of the sum is the sum of the variances of the individual periods, at least if there is no autocorrelation in returns.

The problem can then be formulated as:

$$\min_{\sum_{t=1}^{T} E(y_t) \geq \mu_0} \sum_{t=1}^{T} V(y_t) \tag{2}$$

By varying the required return, we can map out the entire efficient frontier. The optimal point on this frontier depends upon the investor's tolerance for risk. If we define the coefficient of risk aversion to be λ, the solution obtained in one step is:

$$\max \sum_{t=1}^{T} [E(y_t) - \lambda V(y_t)] \tag{3}$$

Generally this problem is defined in returns, but this dollar-based formulation is equivalent. If there is a collection of assets available with known means and covariance matrix, the solution to this problem yields an optimal portfolio. When the problem is reformulated relative to a benchmark, the value of the portfolio at each time as well as the price of each asset at each time is measured relative to the benchmark portfolio. This will not affect anything in the subsequent analysis.

Treating each of the subperiods separately could solve this problem, but this would not in general be optimal. Better solutions involve forecasts and dynamic programming or hedge portfolios; see Merton [1973], Constantinides [1986], and Colacito and Engle [2004].

The classic trading problem is conveniently formulated using the "implementation shortfall" of Perold [1988]. This is now often described as measuring trading costs relative to an arrival price benchmark. See, for example, Chan and Lakonishok [1995], Almgren and Chriss [1999, 2000], Grinold and Kahn [1999], and Bertismas and Lo [1998], among others.

If a large position is sold in a sequence of small trades, each part will trade at potentially a different price. The average price can be compared to the arrival price to determine the shortfall.

Let the position measured in shares at the end of time period t be x_t so that the trade is the change in x. Let the transaction price at the end of time period t be $\tilde{p}_t$ and the fair market value measured perhaps by the midquote be p_t. The price at the time the order was submitted is p_0, so the transaction cost (TC) in dollars is given by:

$$TC = \sum_{t=1}^{T} \Delta x_t'(\tilde{p}_t - p_0) \tag{4}$$

In the liquidation example, the change in position is negative, so a transaction price below the arrival price corresponds to positive transaction costs. If the trade is a purchase, the trades will be positive, and if the executed price rises the transaction cost will again be positive. When multiple assets are traded, the position and price can be interpreted as vectors giving the same expression.

It may also be convenient to express this as a return relative to the arrival price valuation of the full order:

$$TC\% = TC / | x_T - x_0 |' p_0 \tag{5}$$

Transaction costs can also be written as the divergence of transaction prices from each local arrival price plus a price impact term:

$$TC = \sum_{t=1}^{T} \Delta x_t (\tilde{p}_t - p_t) + \sum_{t=1}^{T} (x_T - x_{t-1})' \Delta p_t \tag{6}$$

In some cases this is a more convenient representation.

On average we expect this measure to be positive. For a single small order executed instantly, there would still be a difference between the arrival price and the transaction price given by half the bid-ask spread. For larger orders and orders that are broken into smaller trades, there will be additional costs due to the price impact of the first trades and additional uncertainty due to unanticipated price moves. The longer the period over which the trade is executed, the more uncertainty there is in the eventual transaction cost.

Both the mean and variance of the transaction cost are important to the investment decision. The problem can be formulated as finding a sequence of trades to solve:

$$\min_{s.t. V(TC) \le K} E(TC) \tag{7}$$

where K is a measure of the risk that is considered tolerable. By varying K, the efficient frontier can be traced out and optimal points selected. Or, by postulating a mean-variance utility function for trading with risk aversion parameter λ^*, we could solve for the trading strategy by

$$\min E(TC) + \lambda^* V(TC) \tag{8}$$

This leaves unclear the question of integrating these two problems. Is this the same lambda, and can these various costs be combined for joint optimization?

ONE PROBLEM

We now formulate the two problems as a single optimization in order to see the relation between them. The vector of holdings in shares at the end of the period is denoted by x_t and the market value per share at the end of the period by p_t, which may be interpreted as the midquote. The portfolio value at time t is therefore given by

$$y_t = x_t' p_t + c_t \tag{9}$$

where c_t is the cash position. The change in value from $t = 0$ to $t = T$ is therefore given by:

$$y_T - y_0 = \sum_{t=1}^{T} \Delta y_t = \sum_{t=1}^{T} \left(x_{t-1}' \Delta p_t + \Delta x_t' p_t + \Delta c_t \right) \tag{10}$$

Assuming that the return on cash is zero and there are no dividends, and the change in cash position is just a result of purchases and sales, each at transaction prices. The equation is completed with:

$$\Delta c_t = -\Delta x_t' \tilde{p}_t \tag{11}$$

Here all trades take place at the end of the period so the change in portfolio value from $t - 1$ to t is immediately obtained from (9) and (11) to be

$$\Delta y_t = x_{t-1}' \Delta p_t - \Delta x_t' (\tilde{p}_t - p_t) \tag{12}$$

The gain is simply the capital gains on the previous-period holdings less the transaction costs of trades using end-of-period prices. It is a self-financing portfolio position.

Substituting (11) into (10) and then identifying transaction costs from (4) gives the key result:

$$y_T - y_0 = x_T' (p_T - p_0) - TC \tag{13}$$

The portfolio gain is simply the total capital gain if the transaction had occurred at time 0, less the transaction costs. The simplicity of the formula masks the complexity of the relation. The transactions will of course affect the evolution of prices, and therefore the decision of how to trade will influence the capital gain as well.

Proposition 1. The optimal mean-variance trade trajectory is the solution to:

$$\max_{\{x_t\}} E(x_T'' (p_T - p_0) - TC) - \lambda V(x_T'' (p_T - p_0) - TC \tag{14}$$

or equivalently

$$\max_{\{x_t\}} E\left(\sum_{t=1}^{T} (x_{t-1}' \Delta p_t - \Delta x_t' [\tilde{p}_t - p_t]) \right) - \lambda V\left(\sum_{t=1}^{T} (x_{t-1}' \Delta p_t - \Delta x_t' [\tilde{p}_t - p_t]) \right) \tag{15}$$

The two problems have become a single problem. The risk aversion parameter is the same in the two problems. The mean return is the difference of the two means, and the variance of the difference is the risk. It is important to recognize that this is not the sum of the variances, as there will likely be covariances. When x_T is zero as in a liquidation, the problems are identical for either a long or a short position. For purchases or sales with terminal positions that are not purely cash, more analysis is needed.

In this single problem, the decision variables are now the portfolio positions in all time periods including period T. In the static problem described in Equation (1), only a single optimized portfolio position is found, and we might think of this as x_T. In (7), portfolio positions at times ($t = 1, \ldots, T - 1$) are found, but the position at the end is fixed and in this case is zero. In Equation (14), the intermediate holdings as well as the terminal holding are determined jointly. To solve this problem jointly, we must know expected returns, the covariance of returns, and the dynamics of price impact and trading cost.

A conceptual simplification is therefore to suppose that the optimization is formulated from period 0 to T_2 where $T_2 >> T$. During the period from T to T_2, the holdings will be constant at x_T. The problem becomes:

$$\max_{\{x_1,\ldots,x_{T-1}\}\{x_T\}} E[(y_{T_2} - y_T) + (y_T - y_0)] - \lambda V[(y_{T_2} - y_T) + (y_T - y_0)] \quad (16)$$

or, more explicitly, assuming no covariance between returns during $(0, T)$ and (T, T_2):

$$\max_{\{x_1,\ldots,x_{T-1}\}\{x_T\}} E[x_T'(p_{T_2} - p_T)] - \lambda V[x_T'(p_{T_2} - p_T)] + E[x_T'(p_T - p_0) - TC] - \lambda V[x_T'(p_T - p_0) - TC] \quad (17)$$

Although this can be optimized as a single problem, it is clear that if the holdings before T do not enter into the optimization after T, and if the latter period is relatively long, there is little lost in doing this in two steps. It is natural to optimize x_T over the investment period, and then take this vector of holdings as given when solving for the optimal trades.

Formally the approximate problem can be expressed as:

$$\max_{x_T} E(x_T'[p_{T_2} - p_T]) - \lambda V(x_T'[p_{T_2} - p_T]) \quad (18)$$

$$\max_{\{x_1,\ldots,x_{T-1}\}} E(x_T'(p_T - p_0) - TC) - \lambda V(x_T'(p_T - p_0) - TC) \quad (19)$$

This corresponds to the institutional structure as well. Orders are dispatched depending on models of expected returns and risks, and these orders are transmitted to brokers for trading. The traders thus take the orders as given and seek to exercise them optimally. Any failure to fully execute the order is viewed as a failure of the trading system.

Clearly, an institution that trades frequently enough will not have this easy separation, and it will be important to choose the trades jointly with the target portfolio. In this case, the optimal holdings will depend upon transaction costs and price impacts. If enough investors trade in this way, asset prices will be determined in part by liquidity costs.

There is a large literature exploring this hypothesis starting with Amihud and Mendelsohn [1986] and including Easley, Hvidkjaer, and O'Hara [2002], O'Hara [2003], and Acharya and Pederson [2005]. Some of these authors consider liquidity to be time-varying and add risks of liquidating the position as well.

SHARPE RATIO

The Sharpe ratio from trading can be established from Equations (18) and (19). The earnings from initial cash and portfolio holdings accumulated at the risk-free rate r^f, for the period $(0, T)$ would yield:

$$RF = r^f y_0 T \quad (20)$$

Hence the annualized Sharpe ratio is given by:

$$\text{Sharpe Ratio} = \frac{E[x_T'(p_T - p_0)] - E(TC) - RF}{\sqrt{T}\sqrt{V(x_T'(p_T - p_0) - TC)}} \quad (21)$$

Clearly transaction costs reduce the expected return and potentially increase the risk. These will both reduce the Sharpe ratio over levels that would be expected in the absence of transaction costs. This can be expressed in terms of the variance and covariance as:

$$\text{Sharpe Ratio} = \frac{E[x_T'(p_T - p_0)] - E(TC) - RF}{\sqrt{T}\sqrt{V[x_T'(p_T - p_0)] + V(TC) - 2Cov(x_T'(p_T - p_0), TC)}} \quad (22)$$

so that the covariance between transaction costs and portfolio gains enters the risk calculation.

The covariance term will have the opposite sign for buys and sells. If the final position is larger than the initial position so that the order is a buy, transaction costs will be especially high if prices happen to be rising, but in this circumstance so will the portfolio value. Sells are the opposite. Hence, for buys, the covariance will reduce the impact of the execution risk, while for sells it will exaggerate it.

In practice, portfolio managers sometimes ignore these aspects of transaction costs. On average, this means that the realized Sharpe ratio will be lower than the anticipated ratio. This could occur from ignoring either the expected transaction costs, or the risk of transaction costs, or both. This leads not only to disappointment, but also to inferior planning. Optimal allocations selected with an incorrect objective function are of course not really optimal.

Consider the outcome using the optimal objective function in (17) as compared with two inappropriate objective functions. We might call the first pure Markowitz, suggesting that this is the classic portfolio problem with no adjustment for transaction costs:

$$\max_{\{x_1,\ldots,x_{T-1}\}\{x_T\}} E[x_T{}'(p_T - p_0)] - \lambda V[x_T{}'(p_T - p_0)] \quad (23)$$

We call the second cost-adjusted Markowitz, which takes expected transaction costs but not transaction risks into account:

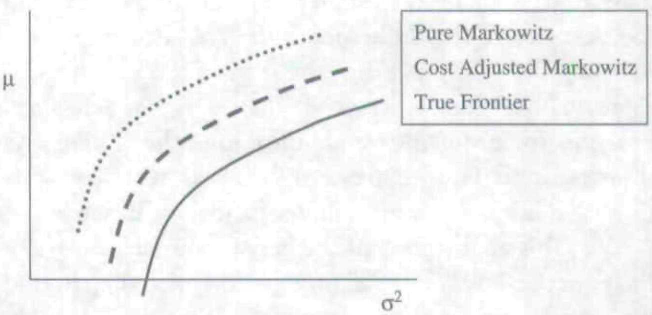
$$\max_{\{x_1,\ldots,x_{T-1}\}\{x_T\}} E[x_T{}'(p_T - p_0) - TC] - \lambda V[x_T{}'(p_T - p_0)] \quad (24)$$

For a theory of transaction costs, the risk-return frontier can be calculated for each of these objective functions. In general, as in Exhibit 1, the pure Markowitz frontier will be highest, followed by the cost-adjusted Markowitz frontier followed by the true frontier. A portfolio that is optimal with respect to the pure Markowitz or cost-adjusted Markowitz will not generally be optimal with respect to the true frontier, and will typically lie inside it.

ASSUMPTIONS ON DISTRIBUTION OF RETURNS AND TRANSACTION COSTS

We add specific assumptions on trading costs to the problem to solve for the optimal trajectory of trades. The risk will not be zero but will be reduced until the corresponding increase in expected transaction costs leads to an optimum solution to problem (19).

EXHIBIT 1



Two additional assumptions are:

$$\text{A.1:} \quad V_0(\tilde{p}_t - p_t \mid \{x_t\}) = 0, \quad E_0(\tilde{p}_t - p_t \mid \{x_t\}) \equiv \tau_t \quad (25)$$

$$\text{A.2:} \quad V_0(\Delta p_t \mid \{x_t\}) = \Omega_0, \quad E_0(\Delta p_t \mid \{x_t\}) \equiv \mu + \pi_t \quad (26)$$

The first supposes that the difference between the price at which a trade can instantaneously be executed and the current fair market price is a function of things that are known. The variance is conditional on information at the beginning of the trade such as market conditions, and it is conditional on the selected trajectory of trades. The mean is a function of market conditions and trades and is denoted by τ_t. Of course there could also be uncertainty in the instantaneous execution price, which would add additional terms to the expressions below.

Similarly, the second assumption implies that the evolution of prices will have variances and covariances that are not related to the trades and can be based on the covariance matrix at the initial time. This does not mean that the trades have no effect on prices; it simply means that once the mean of these effects is subtracted, the covariance matrix is unchanged. Expected price change is μ, the alpha of the asset, plus Π_t, the permanent price impact due to trading.

With these two assumptions, the variances and covariances from Equations (14) or (22) can be evaluated:

$$V[x_T(p_T - p_0)] = Tx_T{}'\Omega x_T$$

$$V(TC) = \sum_{t=1}^{T} (x_T - x_{t-1})'\Omega(x_T - x_{t-1})$$

$$Cov = \sum_{t=1}^{T} x_T{}'\Omega(x_T - x_{t-1}) \quad (27)$$

In the interest of clarity, we suppress the conditioning information. For one-asset portfolios, the covariance term will be positive for buying orders and negative for selling orders, meaning risk is reduced for buys and increased for sells. When only a subset of the portfolio is traded, there will again be differences in the covariance between buy and sell trades depending on the correlations with the remaining assets.

Putting these three equations together gives the unsurprising result that the risk depends on the full trajectory of trades:

$$V(x_T'(p_T - p_0) - TC) = \sum_{t=1}^{T} x_{t-1}'\Omega x_{t-1} \tag{28}$$

The net risk when some positions are being increased and others are being reduced depends on the timing of the trades. Carefully designed trading programs can mitigate this risk.

To solve for the optimal timing of trades, the assumptions A.1 and A.2 are substituted into Equation (14):

$$\max_{\{x_t\}} x_T'\sum_{t=1}^{T}(\mu + \pi_t) - E(TC) - \lambda\sum_{t=1}^{T} x_{t-1}'\Omega x_{t-1} \tag{29}$$

An expression for expected transaction costs can be obtained as:

$$E(TC) = \sum_{t=1}^{T} \Delta x_t \tau_t + (x_T - x_{t-1})(\mu + \pi_t) \tag{30}$$

Proposition 2. Under assumptions A.1 and A.2, the optimal trajectory is given by the solution to:

$$\max_{\{x_t\}} \sum_{t=1}^{T}[x_{t-1}(\mu + \pi_t) - \Delta x_t \tau_t] - \lambda\sum_{t=1}^{T} x_{t-1}'\Omega x_{t-1} \tag{31}$$

This solution depends upon desired or target holdings and permanent and transitory transaction costs, as well as expected returns and the covariance matrix of returns. Under specific assumptions on these parameters and functions, the optimal trajectory of trades can be computed and the Sharpe ratio evaluated. Because the costs π_t, τ_t are potentially non-linear functions of the trade trajectory, this is a non-linear optimization.

Under a special assumption, this problem can be simplified further. If the target holdings are solved by optimization, as for example when a post-trade position is to be held for a substantial time, a relation between these parameters can be employed:

A.3 $$x_T = \frac{1}{2\lambda}\Omega^{-1}\mu \tag{32}$$

Proposition 3. Under assumptions A.1, A.2, and A.3, the optimal trajectory is the solution to:

$$\max_{\{x_t\}} \sum_{t=1}^{T}[x_{t-1}\pi_t - \Delta x_t \tau_t +$$

$$\lambda x_T'\Omega x_T - \lambda\,(x_T - x_{t-1})'\Omega(x_T - x_{t-1})] \tag{33}$$

This solution no longer depends upon the expected return but does depend upon the target holding. The appropriate measure of risk is simply the variance of TC, which is the price risk of unfinished trades. Thus buys and sells entail the same risk as liquidations, whatever the other holdings in the portfolio. Essentially, risk is due to the distance away from the optimum at each time.

ALMGREN-CHRISS DYNAMICS

To solve this problem, we must specify the functional form of the permanent and transitory price impacts. A useful version is formulated in Almgren and Chriss [2000]. Suppose that:

A.1.a $$\tilde{p}_t - p_t = \mathrm{T}\Delta x_t \tag{34}$$

A.2.a $$\Delta p_t = \Pi\Delta x_t + \mu + \varepsilon_t,\ \varepsilon_t \sim D(0,\Omega) \tag{35}$$

describe the evolution of transaction prices and market values, respectively. Now *T* is a matrix of transitory price impacts, and *Π* is a matrix of permanent price impacts. The parameters (μ, Ω) represent the conditional mean vector and the covariance matrix of returns.

From Huberman and Stanzel [2004] we learn that the permanent effect must be time-invariant and linear to avoid arbitrage opportunities, although the temporary impact has no such restrictions. Substituting into (6) and rearranging gives:

$$E(TC) = \sum_{t=1}^{T}\{\Delta x_t \mathrm{T}\Delta x_t + (x_T - x_{t-1})(\mu + \Pi\Delta x_t)\}$$

$$V(TC) = \sum_{t=1}^{T}(x_T - x_{t-1})'\Omega(x_T - x_{t-1}) \tag{36}$$

Substituting (34) and (35) into (31) gives:

$$\max_{\{x_t\}} \sum_{t=1}^{T}[x_{t-1}'(\mu + \Pi\Delta x_t) - \Delta x_t'\mathrm{T}\Delta x_t] - \lambda\sum_{t=1}^{T} x_{t-1}'\Omega x_{t-1} \tag{37}$$

The solution to this problem depends on the initial and final holdings as well as the mean and covariance matrix of dollar returns. As the problem is quadratic, it has a closed-form solution for the trade trajectory in terms of these parameters. In this setting, it is clear that the full vector of portfolio holdings and expected returns will be needed to optimize the trades.

If in addition it is assumed that the target holdings are chosen optimally so that A.3 holds, the optimization problem is:

$$\max_{\{x_1,\ldots,x_{T-1}\}} \sum_{t=1}^{T} [x_{t-1}'\Pi\Delta x_t - \Delta x_t' \mathrm{T}\Delta x_t + \lambda x_T'\Omega x_T \\ -\lambda\,(x_T - x_{t-1})'\Omega(x_T - x_{t-1})] \tag{38}$$

The target holdings provide all the relevant information on the portfolio alpha and lead to this simpler expression.

An important implication of this framework is that it gives a specific instruction for portions of the portfolio that are not being traded. Suppose that the portfolio includes N assets but that only a subset of these is to be adjusted through the trading process. Call these assets x_1 and the remaining assets x_2 so that $x_t = (x_{1,t}', x_{2,t}')' = (x_{1,t}', x_{2,T}')'$.

Because the shares of non-traded assets are held constant, the levels of these holdings disappear in Equation (38) in all but one of the terms that include interaction with the trades of asset 1. This remaining term is $x_{t-1}'\Pi\Delta x_t$. If trades in asset 1 are assumed to have no permanent impact on the prices of asset 2, then the holdings of asset 2 will not affect the optimal trades of asset 1.

Typically the permanent impact matrix is assumed to be diagonal or block-diagonal, so even this effect is not present. Of course, it is easy to find cases where cross-asset permanent price impacts could be important.

Proposition 4. Assuming A.1.a, A.2.a, and A.3, and that the permanent impact of traded assets on non-traded assets is zero, portfolio holdings that are fixed during the trade have no effect on the optimal execution strategy.

In light of Equation (14), Proposition 4 might seem surprising, as the risk of trading will have a covariance with portfolio risk. In particular, for positively correlated assets, the covariance when buying will reduce risk and when selling will increase it. This effect is exactly offset by the expected return when assuming optimal target portfolios.

Several important implications of this framework are easily computed in a simple three-period problem. We suppose that these are $(0, t, T)$ and that the holdings at 0 and T are given, but that the holdings at t are to be found to solve (38). Rewriting this expression gives the equivalent problem as long as all the matrices are symmetric:

$$\max_{\{x_t\}} x_0'(\Pi + 2T)x_t + x_T'(\Pi + 2T + 2\lambda\Omega)x_t \\ - x_t'(\lambda\Omega + \Pi + 2T)x_t \tag{39}$$

which has a solution:

$$x_t = \frac{1}{2}(\Pi + 2T + \lambda\Omega)^{-1} \times \\ \{(\Pi + 2T)x_0 + (\Pi + 2T + 2\lambda\Omega)x_T\} \tag{40}$$

In the simple case, when there is no risk aversion, this model gives the widely known solution that half the trades should be completed by half the time. As risk aversion increases, the weights change. In the scalar case, the trades will be advanced so that the holding at the midpoint is closer to the final position than the initial. We see now that this is true not only for liquidations but also for trades with non-zero terminal positions.

If some of the positions are unchanged from the initial to the final period, Equation (40) still gives the solution for the intermediate holding. Notice that this implies that there could be trading in these assets. There could even be trading in assets that are not held at either the initial or the final period.

Partition $x_t = (x_{1,t}', x_{2,t}')'$ and similarly partition the initial and final positions. Then, assuming, $x_{2,0} = x_{2,T}$, we can ask whether the second group of assets would be traded at all in an optimal trade.

When the permanent and transitory impact matrices and the covariance matrix are block-diagonal, the second assets would not be traded, and the optimal trade would not depend upon the position in the second assets. In general, however, Equation (40) would imply some trading in other assets.

Assuming that the price impact matrices are block-diagonal, a simple expression for trade in the second asset can be found in the three-period problem:

$$x_{2,t} = x_{2,T} - (\Pi_{22} + 2T_{22} + \lambda\Omega_{22})^{-1}\lambda\Omega_{12}(x_{1,t} - x_{1,T}) \tag{41}$$

For positively correlated assets, all the parameters would be positive. Hence, asset 2 will be above its target whenever asset 1 is below its target. This means that a buying trade for asset 1 would require buying asset 2 as well and then selling it back. The unfinished part of the trade would then be long asset 2 and short asset 1, leading to risk reduction. Similarly a selling trade for asset 1 would involve also selling asset 2 and then buying it back.

Notice that if lambda is high, or there are negligible transaction costs for asset 2, the relation is simply the beta of asset 2 on asset 1. Regressing returns of asset 2 on asset 1 would give a regression coefficient that would indicate the optimal position in asset 2 when its transaction costs are minimal. This would suggest considering a futures contract as the second asset, as it involves minimal transaction costs and can be used to hedge a wide range of trades.

To see the appearance of these trades, we construct a small Excel spreadsheet version of this model with 20 periods of trading and rather typical parameters. Exhibit 2 illustrates a buying trade under various conditions.

The horizontal line indicates the risk-neutral solution, which is to space the trades evenly. The squares give the solution under risk aversion when only one asset is traded. When a second asset is traded, the curve shifts to the triangles and then to the crosses for high correlations. Notice that the trading in this asset is less aggressive when the second asset is also traded. The better the hedge of the second asset, the less aggressive the trading needs to be.

The holdings of the second asset are shown in Exhibit 3. Clearly, the asset is purchased initially and then gradually resold to return to the initial holdings. The higher the covariance, the larger the positions.

REVERSALS

A richer set of dynamic relations will lead to more interesting trading strategies. In particular, the assumption that the impact of trades is felt completely after one period may be too simple. A natural generalization of A.1.a and A.1.b allows both the temporary and permanent impact of a trade to have delayed impacts. This generalization, however, is still a special case of A.1 and A.2:

$$\text{A.2.b} \quad \begin{aligned} \tilde{p}_t &= p_t + T\Delta x_t + T_1\Delta x_{t-1} + \ldots + T_q\Delta x_{t-q} \\ &= p_t + T(L)\Delta x_t \end{aligned} \tag{42}$$

$$\text{A.3.b} \quad \begin{aligned} \Delta p_t &= \Pi\Delta x_t + \Pi_1\Delta x_{t-1} + \ldots + \Pi_q\Delta x_{t-q} + \mu + \varepsilon_t \\ &= \Pi(L)\Delta x_t + \mu + \varepsilon_t \end{aligned} \tag{43}$$

The lagged effects allow both continuations and reversals. If there were a strong set of buying orders in the last period, for example, the transaction price may be elevated this period as well as the last period. This would be a transitory continuation, implying that a continued buy program will raise transaction prices and spreads substantially while it continues, but afterward prices and spreads will revert to normal levels.

In the permanent equation, it might be that a buy order in the previous period raises the price in this period above the increase in the last period. This would be a permanent continuation. It could also be that the permanent effect would be a reversal, so that the lagged coefficient would be negative. In many ways, this is a very interesting effect, as it is highly undesirable to buy a stock at a rising price, only to find it drop back after the purchase is completed.

These two equations can be substituted in the optimization problems of proposition 2 in (31) or proposition 3 in (33). Now, however, the trade patterns before T will influence the prices for a short time after T because of the lags introduced. The solution is to pick a time $T_1 > T$ and then choose the trading strategy to maximize

EXHIBIT 2

Optimal Execution with Second Asset

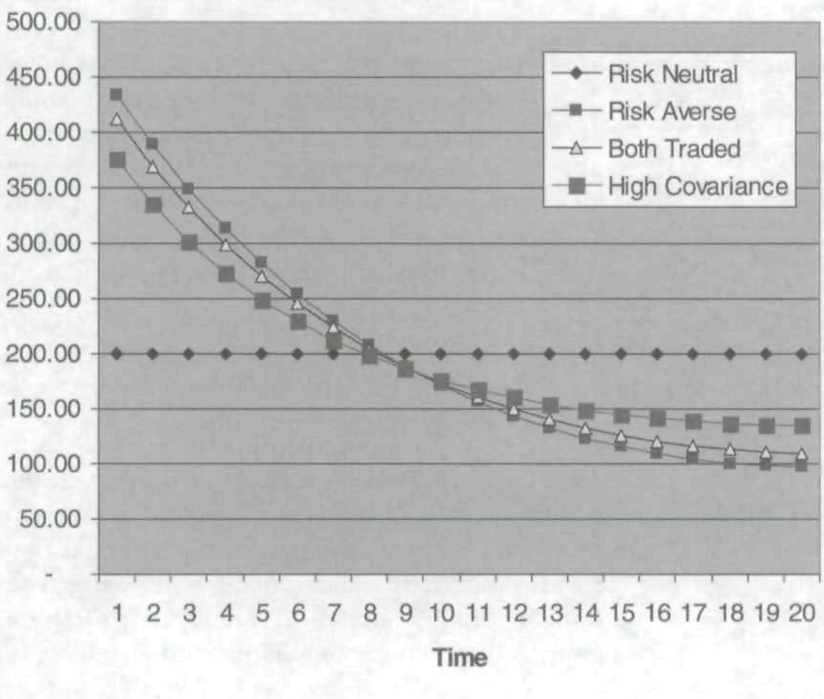


EXHIBIT 3

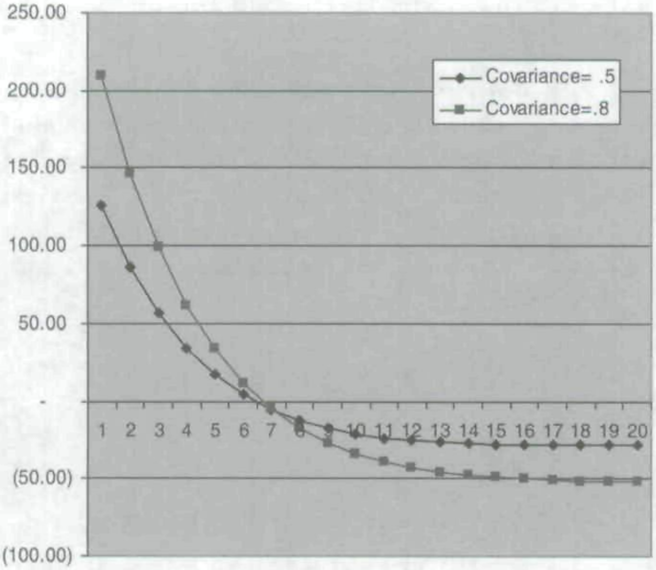
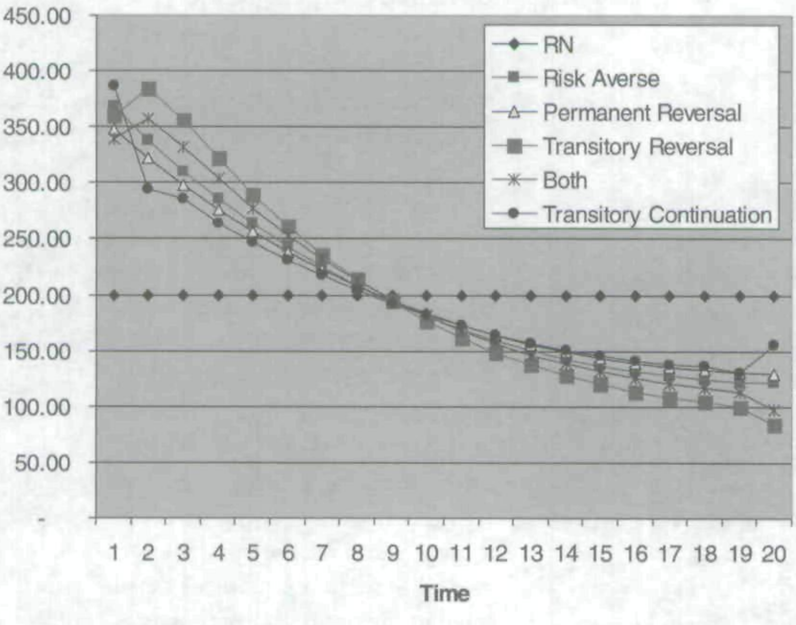


EXHIBIT 4

Optimal Trading with Reversals



portfolio value from $(0, T_1)$. If there are q lags, then T_1 must be at least $T + q$ in order to fully incorporate reversals and continuations into the trading optimization.

The resulting optimization problem can be expressed in:

Proposition 5. Under assumptions A.1.b and A.2.b, the optimal trajectory is the solution to:

$$\max_{\{x_t\}} \sum_{t=1}^{T_1} [x_{t-1}'(\mu + \Pi(L)\Delta x_t) - \Delta x_t' \mathrm{T}(L)\Delta x] - \lambda \sum_{t=1}^{T_1} x_{t-1}'\Omega x_{t-1} \qquad (44)$$

and adding assumption A.3, it is:

$$\max_{\{x_1,\ldots,x_{T-1}\}} \sum_{t=1}^{T_1} [x_{t-1}'\Pi(L)\Delta x_t - \Delta x_t'\mathrm{T}(L)\Delta x_t + \lambda x_T'\,\Omega x_T - \lambda\,(x_T - x_{t-1})'\Omega(x_T - x_{t-1})] \qquad (45)$$

The problem remains linear-quadratic and has a closed-form solution. It is of course still the case that the risk tolerance for trading and for investment should be the same.

Using the 20-period simulation model, it is simple to calculate trajectories for trades with various types of continuations and reversals. Exhibit 4 shows the trade trajectories for a buy order just as above, where the constant path is risk-neutral and the squares are the risk-averse trades when there are no reversals.

When there is a one-period permanent reversal, the trade is less aggressive, and when there is a one-period transitory continuation, the trade is also less aggressive. A one-period transitory reversal encourages more early trades even when there is also a permanent reversal, however. Of course, these offsetting effects would be sensitive to the size of the coefficients.

These models can also be blended with the portfolio of untraded or unhedged assets so the problems can be solved jointly. Qualitatively the results in the simple simulation are the same.

LIQUIDITY RISK

The value of a portfolio of assets is typically marked to market even though the assets could not be liquidated at these prices. An alternative approach is to value the portfolio at its liquidation value. This measure then incorporates both market risk and liquidity risk. As should be abundantly clear, both components of the valuation will be random variables, and it is natural to ask what the distribution of future liquidation values may be.

The process we have developed allows such a calculation. If the portfolio at time $t = 0$ has positions x_0, the mark-to-market value is:

$$y_0 = p_0{}'x_0 + c_0 \tag{46}$$

The cash equivalent of this requires time and execution costs. Thus setting positions $x_T = 0$ at time $t = T$, there is a distribution of values. From (13):

$$y_T - y_0 = x_T{}'(p_T - p_0) - TC \tag{47}$$

this depends simply on the transaction costs during the liquidation. As these are random, the liquidation value is not a number but a random variable. In fact, as there are many ways to trade out of a position, this is a family of random variables, and investors can select their preferred distribution. If the liquidation is aggressive, the costs will be high but rather certain. If the position is liquidated slowly, the average cost will be low but the range of possible outcomes will be wide.

Following the large literature on liquidity risk, or the much larger literature on market risk, it is natural to choose strategies based on expected utility maximization, but to measure risk as based on the likelihood of a particularly bad outcome (see Bangia et al. [1999], Harris, [2003], and Malz [2003]). According to the quantile approach of value at risk (VaR), a bad outcome might be the 99% quantile of the transaction cost associated with the liquidation. This gives a close parallel to market risk.

If market risk is the 1% quantile from holding the portfolio fixed for ten trading days, the liquidation risk could be the 1% quantile of the cash position after liquidating the portfolio optimally over perhaps the same ten days. The liquidity risk could be higher or lower than the market risk in this case.

The price risk faced by the portfolio owner will diminish as each piece is sold, making the risk lower than market risk, but there is a directional loss associated with the liquidation that could completely dominate the market risk. From this point of view, the liquidity risk might be relatively insensitive to the time allowed for liquidation, as the optimal liquidation would always be front-loaded so that longer times would be of little value. Only for large positions will the time to liquidation be an important constraint on minimizing the liquidation cost.

Just as VaR has theoretical disadvantages as a measure of market risk, the quantile has the same drawbacks for liquidity risk. Expected shortfall is an alternative, as in McNeil et al. [2005], and the same suggestion can be used for liquidity risk. The liquidity risk would then be defined as the expected liquidation cost given that it exceeds the 99% quantile.

From Equations (36), which incorporate the Almgren-Chriss dynamics, it is clear that the size of the price impact terms is important, and that the volatilities and correlations of the assets are also important. Thus there is a connection between market risk and liquidity risk, but the liquidity measure incorporates both effects.

Liquidity risk would be time-varying because of the risk of holding positions. Price impacts are also often found to be greater when markets are more volatile, so that the parameters τ and π will themselves depend on volatility as well as other factors. Thus liquidity risk may fluctuate more than proportionally to volatility.

Finally, in crisis scenarios, counterparties may also face the same liquidity risks. Hence the price impact coefficients could be even larger. Thus a liquidity failure can be modeled by increasing these parameters and recomputing portfolio liquidity risk. The concept of "liquidity black holes" as in Persaud [2003] could perhaps be modeled in this way.

CONCLUSION

Execution risk and investment risk are the same thing. They can be traded off against one another, and the same risk tolerance should be used to evaluate trading strategies and investment strategies. The optimization of investment and trading strategies can be separated if there is little time allowed for trading compared to the time for holding the investment portfolio.

When there are overshooting or reversals from trading, investment and trading returns affect each other, but these can be taken into account in the trade optimization by careful separation. Optimal approaches to hedging transaction risk as well as measures of liquidity risk follow directly from the analysis.

ENDNOTE

The authors are indebted to Lasse Pedersen and to participants in the Morgan Stanley Market Microstructure conference, Goldman Sachs Asset Management, the Rotman School at the University of Toronto, and the NYU QFE Seminar for helpful comments. This article represents the personal opinion of the authors and does not necessarily reflect policy or research of Morgan Stanley.

REFERENCES

Acharya, Viral, and Lasse Pederson. "Asset Pricing with Liquidity Risk." *Journal of Financial Economics*, 11(2005), pp. 375-410.

Almgren, Robert. "Optimal Execution with Nonlinear Impact Functions and Trading-Enhanced Risk." *Applied Mathematical Finance*, 10 (2003), pp. 1-18.

Almgren, Robert, and Neil Chriss. "Optimal Execution of Portfolio Transactions." *Journal of Risk*, 3 (2000), pp. 5-39.

——. "Value Under Liquidation." *Risk*, 12 (1999).

Amihud, Y., and H. Mendelson. "Asset Pricing and the Bid-Ask Spread." *Journal of Financial Economics*, 17 (1986), pp. 223-249.

——. "The Effects of Beta, Bid-Ask Spread, Residual Risk, and Size on Stock Returns." *Journal of Finance*, 44 (1989), pp. 479-486.

Bangia, A., F. X. Diebold, T. Schuermann, and J. D. Stroughair. "Modeling Liquidity Risk With Implications for Traditional Market Risk Measurement and Management." Working Paper 99-06, Wharton Financial Institutions Center, 1999.

Bertsimas, Dimitris, and Andrew W. Lo. "Optimal Control of Execution Costs." *Journal of Financial Markets*, 1 (1998), pp. 1-50.

Chan, Louis K., and Josef Lakonishok. "The Behavior of Stock Prices around Institutional Trades." *Journal of Finance*, 50, No. 4 (1995), pp. 1147-1174.

Colacito, Riccardo, and Robert Engle. "Multiperiod Asset Allocation with Dynamic Volatilities." Working paper, 2004.

Constantinides, George. "Capital Market Equilibrium with Transaction Costs." *Journal of Political Economy*, 94, No. 4 (1986), pp. 842-862.

Duffie, D., and A. Ziegler. "Liquidation Risk." *Financial Analysts Journal*, 59 (3) (2003), pp. 42-51.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara. "Is Information Risk a Determinant of Asset Returns?" *Journal of Finance*, 57 (2002), pp. 2185-2222.

Grinold R., and R. Kahn. *Active Portfolio Management*, 2nd ed. New York: McGraw-Hill, 1999, pp. 473-475.

Harris, Larry. *Trading and Exchanges*. Oxford: Oxford University Press, 2003.

Huberman, Gur, and Werner Stanzl. "Price Manipulation and Quasi-Arbitrage." *Econometrica*, 72, No. 4 (2004), pp. 1247-1275.

Kyle, A. S. "Continuous Auctions and Insider Trading." *Econometrica*, 53 (6) (1985), pp. 1315-1335.

Lo, A. W., C. Petrov, and M. Wierzbicki. "It's 11pm—Do You Know Where Your Liquidity Is? The Mean-Variance-Liquidity Frontier." *Journal of Investment Management*, 1 No.1 (2003).

Malz, Alan. "Liquidity Risk, Current Research and Practice." *Riskmetrics*, Fall 2003.

Markowitz, Harry. "Portfolio Selection." *Journal of Finance*, Vol 7, No. 1 (1952), pp. 77-91.

Merton, Robert C. "An Intertemporal Capital Asset Pricing Model." *Econometrica*, 41 (1973), pp. 867-887.

Obizhaeva, Anna, and Jiang Wang. "Optimal Trading Strategy and Supply/Demand Dynamics." NBER Working Paper No. 11444, 2005.

O'Hara, Maureen. "Liquidity and Price Discovery." *Journal of Finance*, 58 (4) (2003), pp. 1335-1354.

Perold, Andre. "The Implementation Shortfall: Paper Versus Reality." *The Journal of Portfolio Management*, 14 (1988), pp. 4-9.

Persaud, Avinash, ed. *Liquidity Black Holes*. New York: Risk-Books, 2003.