

NBER WORKING PAPER SERIES

EXECUTION RISK

Robert Engle
Robert Ferstenberg

Working Paper 12165
<http://www.nber.org/papers/w12165>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2006

The authors are indebted to Lasse Pedersen and participants in the Morgan Stanley Market Microstructure conference, Goldman Sachs Asset Management, Rotman School at University of Toronto and NYU QFE Seminar for helpful comments. This paper is the private opinion of the authors and does not necessarily reflect policy or research of Morgan Stanley. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2006 by Robert Engle and Robert Ferstenberg. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Execution Risk
Robert Engle and Robert Ferstenberg
NBER Working Paper No. 12165
April 2006
JEL No. G2

ABSTRACT

Transaction costs in trading involve both risk and return. The return is associated with the cost of immediate execution and the risk is a result of price movements during a more gradual trading. The paper shows that the trade-off between risk and return in optimal execution should reflect the same risk preferences as in ordinary investment. The paper develops models of the joint optimization of positions and trades, and shows conditions under which optimal execution does not depend upon the other holdings in the portfolio. Optimal execution however may involve trades in assets other than those listed in the order; these can hedge the trading risks. The implications of the model for trading with reversals and continuations are developed. The model implies a natural measure of liquidity risk.

Robert Engle
Department of Finance, Stern School of Business
New York University, Salomon Center
44 West 4th Street, Suite 9-160
New York, NY 10012-1126
and NBER
rengle@stern.nyu.edu

Robert Ferstenberg
Morgan Stanley
1585 Broadway
New York, NY 10036
Robert.Ferstenberg@morganstanley.com

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 measuring 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 now incorporated in all textbooks.

Although much attention has been paid to the cost of trading, little has been devoted to the risks of trading. Analysis has typically focused on the costs of executing a single trade or, in some cases, a sequence of trades. In a series of papers, Almgren and Chriss(1999)(2000) and Almgren(2003) and Grinold and Kahn (1999), and most recently Obizhaeva and Wang(2005) developed 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 these immediately by sending them to a block desk or other intermediary who will take on the risk, or executing a sequence of smaller trades. These might be planned and executed by a floor broker, by an in-house trader, an 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. The delay in trading introduces price risk due to price movements beyond that which can be anticipated as a natural response to the trade itself. Different trading strategies will have different probability distributions of the costs and thus customers will need to choose the trading strategy that is optimal for them.

This paper addresses 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?

In this paper we will 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 many others.

The paper initially introduces the theoretical optimization problems in section II and synthesizes them into one problem in section III. Section IV discusses the implications of trading strategy on the Sharpe Ratio. A specific assumption is made on price dynamics in Section V leading to specific solutions for the optimal trades. This section also shows the role of non-traded assets. Section VI introduces more sophisticated dynamics allowing reversals. Section VII uses this apparatus to discuss measures of liquidity risk and section VIII concludes.

II. TWO PROBLEMS

A. Portfolio 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:

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

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 $t=0, 1, \dots, T$. The dollar return on the full period is the sum of the dollar returns on the individual periods. Furthermore, 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

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

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

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

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

Treating each of the sub periods separately could then solve this problem; however, this would not in general be optimal. Better solutions involve forecasts and dynamic programming or hedge portfolios. See inter alia Merton(1973), Constantinides(1986), Colacito and Engle(2004).

B. Trade Optimization

The classic trading problem is conveniently formulated with 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), Grinold and Kahn(1999), Almgren and Chriss(1999)(2000) 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 with 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 in dollars is given by

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

Since in the liquidation example, the change in position is negative, a transaction price below the arrival price corresponds to positive transaction costs. If on the other hand the trade is a purchase, then 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. This express gives

$$TC\% = TC / (x_T - x_0)' p_0 \quad (5)$$

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

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

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 time period over which the trade is executed, the more uncertainty there is in the eventual transaction cost. We can consider both the mean and variance of the transaction cost as being important to the investment decision.

The problem then can be formulated as finding a sequence of trades to solve

$$\min_{s.t. V(TC) \leq K} E(TC) \quad (7)$$

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. Equivalently, by postulating a

mean variance utility function for trading with risk aversion parameter λ^* , we could solve for the trading strategy by

$$\min E (TC) + \lambda^* V (TC). \quad (8)$$

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

III. ONE PROBLEM

We now formulate these 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 will be denoted by x_t and the market value per share at the end of the period will be p_t , which may be interpreted as the midquote. The portfolio value at time t is therefore given by

$$y_t = x_t' p_t + c_t \quad (9)$$

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

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

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

$$\Delta c_t = -\Delta x_t' \tilde{p}_t \quad (11)$$

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 (10) and (11) to be

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

The gain is simply the capital gains on the previous period holdings less the transactions 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:

$$y_T - y_0 = x_T' (p_T - p_0) - TC \quad (13)$$

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

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

or equivalently

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

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 notice 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 at 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 $\{t=1, \dots, T-1\}$ 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 $T_2 \gg T$. During the period from T to T_2 the holdings will be constant at x_T . The problem becomes

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

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

$$\begin{aligned} \max_{\{x_1, \dots, x_{T-1}\} \{x_T\}} E \left[x_T' (p_{T_2} - p_T) \right] - \lambda \mathcal{W} \left[x_T' (p_{T_2} - p_T) \right] \\ + E \left[x_T' (p_T - p_0) - TC \right] - \lambda \mathcal{W} \left[x_T' (p_T - p_0) - TC \right] \end{aligned} \quad (17)$$

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

$$\max_{x_T} E \left(x_T' (p_{T_2} - p_T) \right) - \lambda \mathcal{W} \left(x_T' (p_{T_2} - p_T) \right) \quad (18)$$

$$\max_{\{x_1, \dots, x_{T-1}\}} E \left(x_T' (p_T - p_0) - TC \right) - \lambda \mathcal{W} \left(x_T' (p_T - p_0) - TC \right) \quad (19)$$

This corresponds to the institutional structure as well. Orders are decided based 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 for it to choose the trades jointly with the target portfolio. In this case, the optimal holdings will depend upon transaction costs and price impacts. If sufficient investors trade in this way, then 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 among others O'Hara(2003), Easley Hvidkjaer and O'Hara(2002) and Acharya and Pederson(2005). Some of these authors consider liquidity to be time varying and add risks of liquidating the position as well.

IV 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 greater than the initial position so that the order is a buy, then 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 inferior to the anticipated ratio. This could occur either from ignoring the expected transaction costs, 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 the following two inappropriate objective functions. We might call the first, pure Markowitz suggesting that this is the classical portfolio problem with no adjustment for transaction costs.

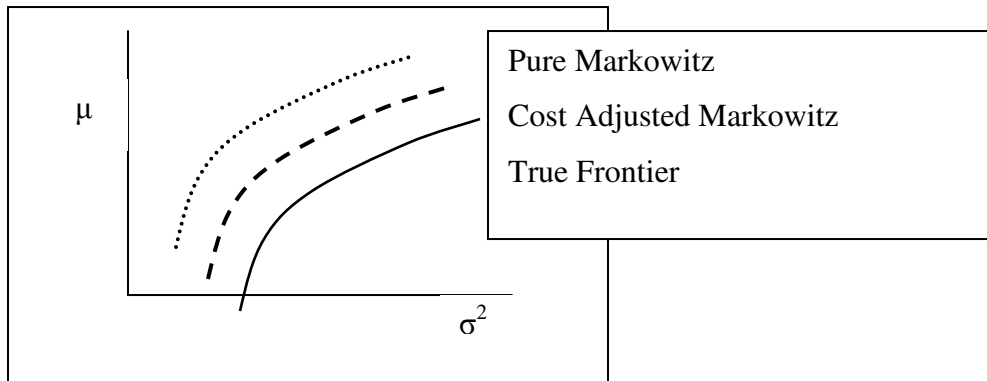
$$\max_{\{x_1, \dots, x_{T-1}\}, \{x_T\}} E[x_T'(p_T - p_0)] - \lambda V[x_T'(p_T - p_0)] \quad (23)$$

We call the second, Cost Adjusted Markowitz, which takes expected transaction costs into account but does not take transaction risks into account.

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

For a theory of transaction costs, the risk/return frontier can be calculated for each of these objective functions. In general, the Pure Markowitz frontier will be highest, followed by the Cost Adjusted Markowitz 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 the frontier.

In the next section, specific assumptions on trading costs will be added 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 to (19).



V. ASSUMPTIONS ON DISTRIBUTION OF RETURNS AND TRANSACTIONS COSTS

We consider the following two additional assumptions.

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

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

These assumptions should be explained. 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 the information set 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 . Clearly, in practice there could also be uncertainty in the instantaneous execution price and this effect would add additional terms in the expressions below.

Similarly, A.2 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 period. 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.

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

$$\begin{aligned} V(x_T(p_T - p_0)) &= Tx_T' \Omega x_T \\ V(TC) &= \sum_{t=1}^T (x_t - x_{t-1})' \Omega (x_t - x_{t-1}) \\ Cov &= \sum_{t=1}^T x_t' \Omega (x_t - x_{t-1}) \end{aligned} \quad (27)$$

For ease of presentation the conditioning information is suppressed. For one asset portfolios, the covariance term will be positive for buying orders and negative for selling orders leading the risk to be 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.

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

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

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

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

Furthermore, an expression for expected transaction costs can be obtained as

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

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

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

This solution depends upon desired or target holdings, 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 further simplified. If the target holdings are solved by optimization, as for example in the case when a post trade position is to be held for a substantial period of time, then there is a relation between these parameters that can be employed.

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

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

$$\max_{\{x_t\}} \sum_{t=1}^T \left[x_{t-1} \pi_t - \Delta x_t \tau_t + \lambda x_T' \Omega x_T - \lambda (x_T - x_{t-1})' \Omega (x_T - x_{t-1}) \right] \quad (33)$$

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 have the same risk as liquidations regardless of the other holdings in the portfolio. Essentially, risk is due to the distance away from the optimum at each point of time.

V 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

$$A.1.a \quad \tilde{p}_t - p_t = T \Delta x_t \quad (34)$$

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

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

$$\begin{aligned} E(TC) &= \sum_{t=1}^T \left\{ \Delta x_t T \Delta x_t + (x_T - x_{t-1}) (\mu + \Pi \Delta x_t) \right\} \\ V(TC) &= \sum_{t=1}^T (x_T - x_{t-1})' \Omega (x_T - x_{t-1}) \end{aligned} \quad (36)$$

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

$$\max_{\{x_t\}} \sum_{t=1}^T \left[x_{t-1}' (\mu + \Pi \Delta x_t) - \Delta x_t' \Gamma \Delta x_t \right] - \lambda \sum_{t=1}^T x_{t-1}' \Omega x_{t-1}. \quad (37)$$

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:

$$\max_{\{x_1, \dots, x_{T-1}\}} \sum_{t=1}^T \left(x_{t-1}' \Pi \Delta x_t - \Delta x_t' \Gamma \Delta x_t + \lambda x_T' \Omega x_T - \lambda (x_T - x_{t-1})' \Omega (x_T - x_{t-1}) \right) \quad (38)$$

The target holdings contain 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 $x_t = (x_{1,t}', x_{2,t}')' = (x_{1,t}', x_{2,T}')'$. Since 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, 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 the light of equation (14) this 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. However 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:

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

which has a solution:

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

In the simple case where 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 particular, 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 liquidation but also for trades with non-zero terminal positions.

If some of the positions are unchanged from the initial to final period, expression (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 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. In the case where the permanent and transitory impact matrices and the covariance matrix are block diagonal, then the second assets would not be traded and the optimal trade would not depend upon the position in the second assets. However, in general 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.

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

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 a 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 large, 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 will incur minimal transaction costs and can be used to hedge a wide range of trades.

To see better the appearance of these trades, a small excel spreadsheet version of this model was constructed with 20 periods of trading. Rather typical parameters were used. In the following figure a buying trade is illustrated under various conditions.

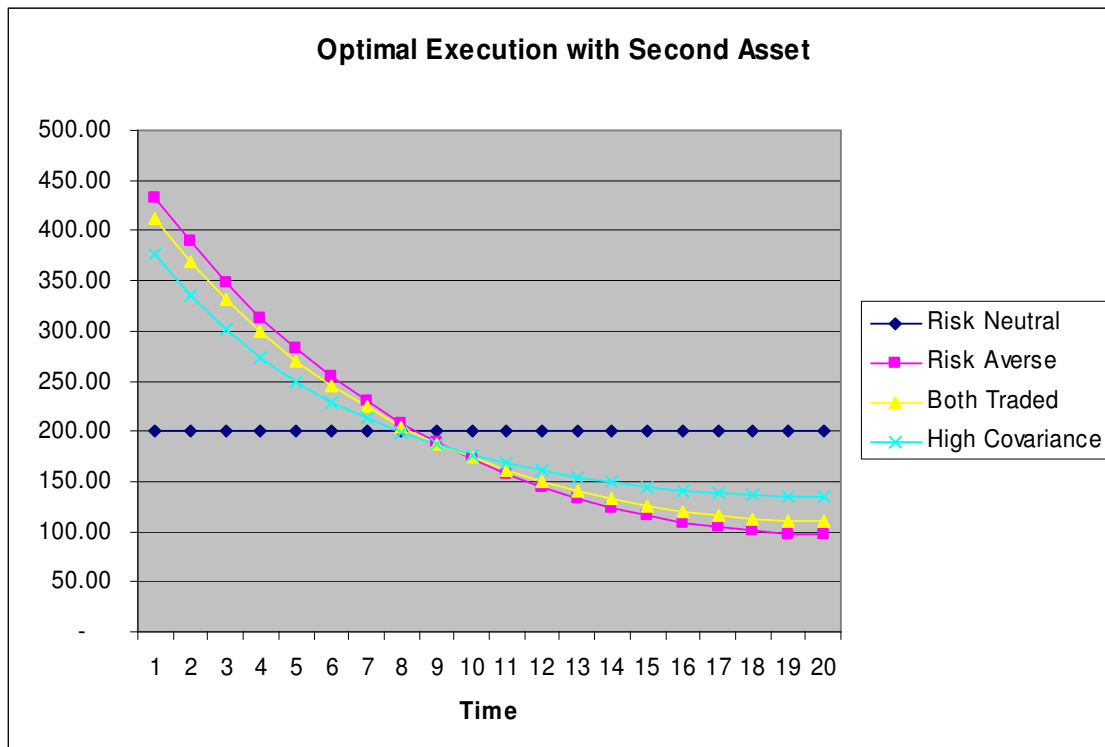


Figure 1

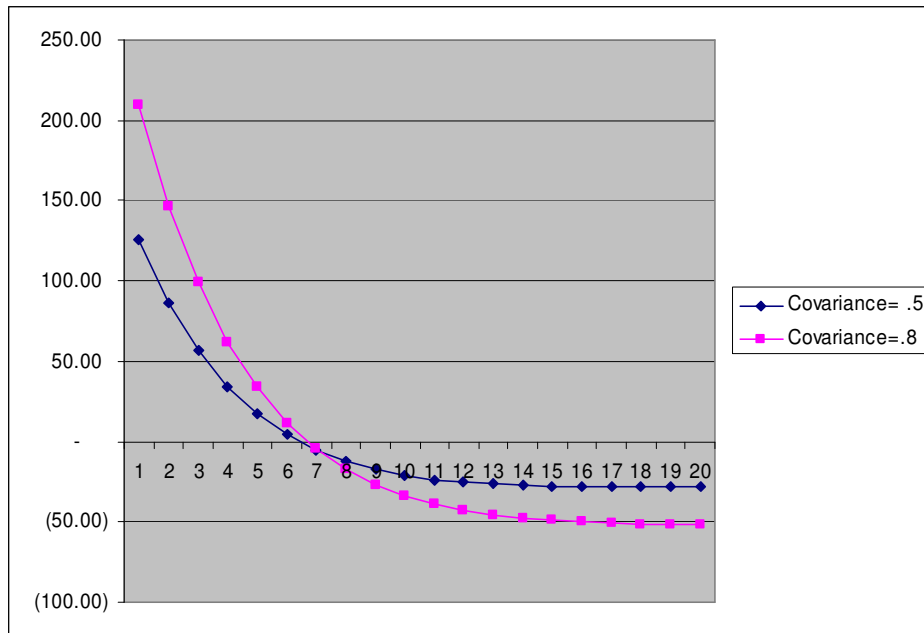


Figure 2

The horizontal line indicates the risk neutral solution which is to evenly space the trades. The squares give the risk averse solution 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 Figure 2. Clearly, the asset is purchased initially and then gradually resold to arrive back at the initial holdings. The higher the covariance, the bigger are the positions.

VI. 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 is however still a special case of A.1 and A.2.

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

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

The lagged effects allow both continuations and reversals. For example, if there were a strong set of buying orders the last period, then it may be that the transaction price will be elevated this period as well as last period. This would be a transitory continuation. It has the implication that a continued buy program will increase transaction prices and spreads substantially while it continues, but afterward, they will revert back to normal levels. In the permanent equation, it might be that a buy order in the previous period raises the price in this period an additional amount above the increase last period. This would be a permanent continuation. However 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 since 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 into 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 due to the lags introduced. The solution is to pick a time $T_1 > T$ and then choose the trading strategy to maximize 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.

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

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

and adding assumption A.3, it is

$$\max_{\{x_1, \dots, x_{T-1}\}} \sum_{t=1}^{T_1} \left(x_{t-1}' \Pi(L)\Delta x_t - \Delta x_t' T(L)\Delta x_t + \lambda x_T' \Omega x_T - \lambda (x_T - x_{t-1})' \Omega (x_T - x_{t-1}) \right) \quad (45)$$

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. Figure 3 shows the trade trajectories for a buy order just as used above where the constant path is risk neutral and the pink 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. On the other hand, a one period transitory reversal encourages more early trades even when there is also a permanent reversal. Of course these offsetting effects would be sensitive to the size of the coefficients.

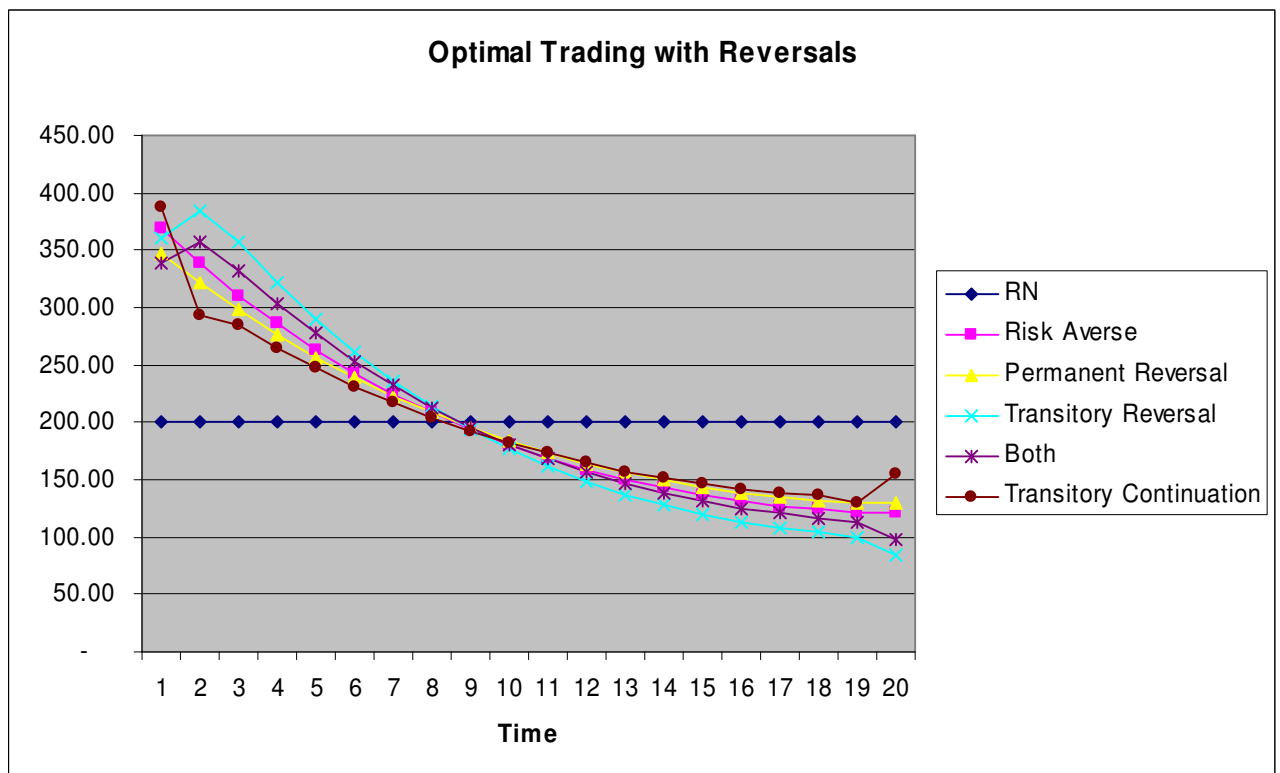


Figure 3

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.

VII 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 has become abundantly clear above, 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 apparatus developed above allows such a calculation. If the portfolio at time $t=0$ has positions x_0 , then the mark-to-market value is

$$y_0 = p_0' x_0 + c_0 \quad (46)$$

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 $y_T = c_T$. From (13), reproduced below,

$$y_T - y_0 = x_T' (p_T - p_0) - TC,$$

this simply depends on the transaction costs during the liquidation. Since 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 the investor can select his preferred distribution. If the liquidation is aggressive, then the costs will be large but rather certain. However if it 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 as in Malz(2003), Bangia et al.(1999), Harris(2003), 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. Taking the quantile approach from VaR, a bad outcome might be the 99% quantile of the transactions cost associated with the liquidation. This gives a close parallel with market risk. If market risk is the 1% quantile from holding the portfolio fixed for 10 trading days, then the liquidation risk could be the 1% quantile of the cash position after optimally liquidating the portfolio over perhaps the same 10 days. The liquidity risk could be more or less 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 less than market risk, however 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 since 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. Instead, expected shortfall is often now used as for example 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 magnitudes of the price impact terms are 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; but in addition, price impacts are often found to be greater when markets are more volatile so that the parameters τ and π will themselves depend upon 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 approximately modeled by increasing these parameters and re-computing portfolio liquidity risk. The concept of “liquidity black holes” as in Persaud(2003) could perhaps be modeled in this way.

VIII. CONCLUSION

In conclusion, execution risk and investment risk are the same thing. They can be traded off against each other and the same risk tolerance should be used to evaluate trading strategies as investment strategies. The optimization of investment and trading strategies can be separated if the time allowed for trading is small relative to the time for

holding the investment portfolio. When there is 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 follow directly from the analysis as well as measures of liquidity risk.

REFERENCES

- Acharya, Viral and Lasse Pederson(2005), “Asset Pricing with Liquidity Risk” *Journal of Financial Economics*,vol 11, pp.375-410
- Almgren, Robert and Neil Chriss, (1999) “Value under Liquidation”, *Risk*, 12
- Almgren, Robert and Neil Chriss,(2000) “Optimal Execution of Portfolio Transactions,” *Journal of Risk*, 3, pp 5-39
- Almgren, Robert,(2003) “Optimal Execution with Nonlinear Impact Functions and Trading-enhanced Risk”, *Applied Mathematical Finance*, 10,pp1-18
- Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17, 223-249.
- Amihud, Y., Mendelson, H., 1989. The Effects of Beta, Bid-Ask Spread, Residual Risk, and Size on Stock Returns. *Journal of Finance* 44, 479-486.
- Bangia, A., Diebold, F. X., Schuermann, T. and Stroughair, J. D. (1999). Modeling liquidity risk with implications for traditional market risk measurement and management, *Working Paper 99-06*, Wharton Financial Institutions Center.
- Bertsimas, Dimitris, and Andrew W. Lo, (1998) “Optimal Control of Execution Costs,” *Journal of Financial Markets*, 1, pp1-50
- Chan, Louis K. and Josef Lakonishok, (1995) “The Behavior of Stock Prices around Institutional Trades”, *Journal of Finance*, Vol 50, No. 4, pp1147-1174
- Colacito, Riccardo and Robert Engle(2004), “Multiperiod Asset Allocation with Dynamic Volatilities”, manuscript
- Constantinides, George,(1986), Capital Market Equilibrium with Transaction Costs, *Journal of Political Economy*, Vol 94, No 4, pp842-862
- Duffie, D. and Ziegler, A. (2003). Liquidation risk, *Financial Analysts Journal* **59**(3): 42–51
- Easley, David, Soeren Hvidkjaer, and Maureen O’Hara, 2002, Is information risk a determinant of asset returns? *Journal of Finance* 57, 2185-2222.
- Grinold R. and R. Kahn(1999) *Active Portfolio Management* (2nd Edition) Chapter 16 pp473-475, McGraw-Hill
- Harris, Larry(2003) *Trading and Exchanges*, Oxford University Press

- Huberman, Gur, and Werner Stanzl, (2004) “Price Manipulation and Quasi-Arbitrage”,
Econometrica, 72, No 4, pp1247-1275
- Kyle, A. S. (1985). Continuous auctions and insider trading, *Econometrica* **53**(6): 1315–1335.
- Lo, A. W., Petrov, C. and Wierzbicki, M. (2003). It’s 11pm—do you know where your
liquidity is? The mean-variance-liquidity frontier., MIT Sloan School, manuscript
- Malz, Alan, “Liquidity Risk, Current Research and Practice”, *Riskmetrics Journal*, Fall
2003
- Markowitz, Harry(1952) “Portfolio Selection”, *Journal Finance*, Vol 7,No 1, pp 77-91
- Merton, Robert C. (1973)”An Intertemporal Capital Asset Pricing Model”,
Econometrica, 41, pp 867-887
- Obizhaeva, Anna and Jiang Wang (2005) “Optimal Trading Strategy and Supply/Demand
Dynamics” manuscript
- O’Hara, Maureen(2003), “Liquidity and Price Discovery”, *Journal of Finance* 58,(4),
2003 pp 1335-1354
- Perold, Andre,(1988) “The Implementation Shortfall: Paper Versus Reality”, *Journal of
Portfolio Management*, 14, pp4-9
- Persaud, Avinash(ed.),(2003), *Liquidity Black Holes*, , RiskBooks,