Measuring and Modeling Execution Cost and Risk[1]

Robert Engle
NYU and Morgan Stanley

Robert Ferstenberg
Morgan Stanley

Jeffrey Russell
University of Chicago

April 2006

Preliminary
Please do not quote without authors' permission

Abstract:

We introduce a new analysis of transaction costs that explicitly recognizes the importance of the timing of execution in assessing transaction costs. Time induces a risk/cost tradeoff. The price of immediacy results in higher costs for quickly executed orders while more gradual trading results in higher risk since the value of the asset can vary more over longer periods of time. We use a novel data set that allows a sequence of transactions to be associated with individual orders and measure and model the expected cost and risk associated with different order execution approaches. The model yields a risk/cost tradeoff that depends upon the state of the market and characteristics of the order. We show how to assess liquidation risk using the notion of liquidation value at risk (LVAR).

[1] This paper is the private opinion of the authors and does not necessarily reflect policy or research of Morgan Stanley. We thank Peter Bolland for very useful discussions as well as seminar participants at the University of Pennsylvania and the NYSE Economics Research Group. Jeffrey Russell gratefully acknowledges Morgan Stanley and NYU for funding a visiting position at NYU Stern School of Business where this research was conducted.

1. Introduction

Understanding execution costs has important implications for both practitioners and regulators and has attracted substantial attention from the academic literature. Traditional analysis of transaction costs focus on the average distance between observed transaction prices and an "efficient" or fair market price. These types of analysis, however, are disconnected from transaction costs faced in practice since they neglect any notion of risk. Specifically, a buy order could be filled by submitting a market order and paying a price near the ask. Alternatively, the order could be submitted as a limit order and either execute at a better price, or not execute at all. Similarly, a single order is often broken up into a sequence of smaller ones spread out over time. This temporal dimension to the problem yields a natural cost/risk tradeoff. Orders executed over a short period of time will have a high expected cost associated with immediate execution but the risk will be low since the price is (nearly) known immediately. Orders executed over a long period of time may have a smaller price impact and therefore smaller expected cost but may be more risky since the asset price can vary more over longer periods of time than shorter periods of time. Using a novel data set that allows transactions to be associated with individual orders we measure and model the expected cost and risk associated with different order execution strategies.

Our empirical work builds directly on the recent research of Almgren and Chriss (1999, 2000), Almgren (2003), Grinold and Kahn (1999), Obizhaeva and Wang (2005), and Engle and Ferstenberg (2006). These papers examine execution quality involving not just the expected cost but also the risk dimension. Order execution strategies that are guaranteed to execute quickly offer a different risk/reward tradeoff than transaction strategies that can take a longer time to be filled. The result is a frontier of risk/reward tradeoffs that is familiar in finance and analogous to classic mean variance analysis of portfolios. In fact, the work of Engle and Ferstenberg (2006) show that this analogy is deeper than might appear at first glace. Namely, they show how to integrate the portfolio decision and execution decision into a single problem and how to optimize these choices jointly.

Our work differs in important ways from most traditional approaches to the analysis of transaction costs. The classic measures of transaction costs such as Roll's measure, (realized) effective spreads, or the half spread measure average (positive) deviations of transaction prices from a notional efficient price[2]. The midquote is often taken as the efficient price. As such, these measures focus purely on expected cost and are not well suited to analyze the cost of limit order strategies or the splitting up of orders into smaller components. Part of the limitations of the traditional analysis of transaction costs is driven by data availability. Standard available data does not generally include information about how long it took before a limit order executed. Even more rare is information providing a link between individual trades and the larger orders.

Using a unique data set consisting of 233,913 orders executed by Morgan Stanley in 2004, we are able to construct measures of both the execution risk and cost[3]. Our data includes information about when the order was submitted and the times, prices, and quantities traded in filling the order. This data allows to take a novel view of the costs and risks associated with order execution.

The expected cost and variance tradeoffs that the trader faces will depend upon the liquidity conditions in the market and the characteristics regarding the order. We model both the expected cost and the risk as a function of a series of conditioning variables. In this way, we are able to generate a time varying menu of expected cost and risk tradeoffs given the state of the market and order characteristics. The result is a conditional frontier

[2] For a survey of the literature see the special issue on transaction costs in the Journal of Financial Markets.
[3] We do not know the identities of the traders and the data never left the confines of Morgan Stanley.

of different cost/risk tradeoffs. This frontier represents a menu of expected cost and variance tradeoffs faced by the trader.

The paper is organized as follows. Section 2 discusses measuring the order execution cost and risk. Section 3 presents the data used in our analysis and some preliminary analysis. Section 4 presents a model for conditional cost and risk with estimates. Section 5 presents an application of the model to liquidity risk and finally, section 6 concludes.

2. Measuring order execution cost and risk.

Our measure of trading costs captures both the expected cost and risk of execution. A key element of the measure takes the price available at the time of order submission as the benchmark price. The order may be executed using a larger number of small trades. Each transaction price and quantity traded might be different. The cost of the trade is always measured relative to a benchmark price which is taken to be the price available at the time of order submission. The transaction cost measure is then a weighted sum of the difference between the transaction price and the benchmark arrival price where the weights are simply the quantities traded. See Chan and Lakonishok (1995), Grinold and Kahn (1999), Almgren and Chriss (1999, 2000), Bertismas and Lo (1998) among others. In this paper the term order refers to the total volume that the agent desires to transact. We will use the term transaction to refer to a single trade. An order may be filled using multiple transactions.

More formally, let the position measured in shares at the end of time period t be x_t so that the number of shares transacted in period t is simply the change in x_t. Let p_0 denote the fair market value of the asset at the time of the order arrival. This can be taken to be the midquote at the time of the order arrival for this price in practice. Let $\tilde{p}_t$ denote the transaction price of the asset in period t. The transaction cost for a given order is then given by:

(1) $$TC = \sum_{t=1}^{T} \Delta x_t \left(\tilde{p}_t - p_0 \right)$$

If the order is purchasing shares then the change in the number of shares will be non-negative. Transactions that occur above the reference price will therefore contribute positively toward transaction costs. Alternatively, when liquidating shares, the change in shares will be non-positive. Transaction prices that occur below the reference price will therefore contribute positively to transaction costs. For a given order, the transaction costs can be either negative or positive depending upon whether the price moved with or against the direction of the order. However, because each trade has a price impact that tends to move the price up for buys and down for sells we would expect the transaction cost to be positive on average. Given transaction cost, both a mean and variance of the transaction cost can be constructed.

Of course, a measure of the transaction cost per dollar traded is obtained by dividing the transaction cost by the arrival value:

(2)
$$TC\% = \frac{TC}{(x_T - x_0)P_0}$$

This measure allows for more meaningful comparison of costs across different orders and it used in our analysis.

The transaction cost can be decomposed into two components that provide some insight. Specifically, the transaction cost can be written as

(3)
$$TC = \sum_{t=1}^{T} \Delta x_t (\tilde{p}_t - p_t) + \sum_{t=1}^{T} (x_T - x_{t-1}) \Delta p_t$$

The first term represents the deviation of the transaction price from the local arrival price p_t. The former is closely related to traditional measures of transaction costs capturing local effects. The second term captures an additional cost due to the price impact. Each trade has the potential to move the value of the asset. This change in the asset price has an effect on all subsequent trades executed. Since the price impact typically moves the price to a less desirable price for the trader, this term will generally increase the cost of executing an order that would be missed by traditional measures that lack this temporal component.

3. The data

In order to analyze our transaction cost measure we need detailed order execution data that includes the arrival price, trade sizes and transaction prices that associated with all the transactions that were used to fill a given order. We obtained such data from Morgan Stanley. We do not know the identity of the traders that placed these orders and more importantly we do not know their motives. The orders could have been initiated by Morgan Stanley traders on behalf of their clients or by a buy side trader on behalf of a portfolio manager. Regardless, we do not know the identity of the Morgan Stanley trader or the client. We use the word “trader” to refer to either one. The order data never left the confines of Morgan Stanley and will not be made available outside of the confines of Morgan Stanley.

The orders were executed by Morgan Stanley’s Benchmark Execution Strategies™ (BXS) strategies during 2004. BXS is a order execution strategy that minimizes the expected cost of the trade for a given level of risk relative to a benchmark. The trades are “optimally” chosen relying on an automated trading procedure that specifies when and how much to trade. The algorithm changes the trading trajectory as the current trading conditions in the market vary[4].

We consider two types of orders. The arrival price (AP) strategy and the volume weighted average price (VWAP). The AP strategy attempts to minimize the cost for a given level of risk around the arrival price p_0. The trader can specify a level of urgency given by high, medium, and low urgency. The level of urgency is inversely related to the level of risk that the trader is willing to tolerate. High urgency orders have relatively low risk, but execute at a higher average cost. The medium and low urgency trades execute with progressively higher risk but at a lower average cost. The trader chooses the urgency and the algorithm derives the time to complete the trade given the state of the market and the trader’s constraints. For a given order size and market conditions, lower

[4] The trading algorithm is a variant of Almgren and Chriss (2000) and the interested reader is referred to this paper for more details.

urgency orders tend to take longer to complete than higher urgency trades. However, since the duration to completion depends upon the market conditions and other factors there is not perfect correspondence between the urgency level and the time to complete the order. All orders in our sample, regardless of urgency, are filled within a single day.

We also consider VWAP orders. For these orders, the trader selects a time horizon and the algorithm attempts to execute the entire order by trading proportional to the market volume over this time interval. We only consider VWAP orders where the trader directed that the order be filled over the course of the entire trading day or that the overall volume traded was a very small fraction of the market volume over that period. This can be interpreted as a strategy to minimize cost regardless of risk. As such, we consider this a risk neutral trading VWAP strategy. Generally, these orders take longer to fill than the low urgency orders and should provide the highest risk and the lowest cost.

We consider orders for both NYSE and NASDAQ stocks. In order to ensure that orders of a given urgency reflect the cost/risk tradeoff optimized by the algorithm we apply several filters to the orders. Only completed orders are considered. Hence orders that begin to execute and are then cancelled midstream are not included in order to ensure homogeneity of orders of a given type. We excluded short sales because the uptick rule prevents the economic model from being used "freely". We do not consider orders executed prior to 9:36 since the market conditions surrounding the open are quite different than non-opening conditions. Only stocks that have an arrival price greater than $5 are included. Orders that execute in less than 5 minutes tend to be very small orders that may be traded in a single trade. As such, they are not representative of the cost/risk tradeoff optimized by the algorithm. For similar reasons, orders smaller than 1000 shares are also not included. Finally, orders that are constrained to execute more quickly than the algorithm would dictate due to the approaching end of the trading day are also excluded. In the end, we are left with 233,913 orders.

For each order we construct the following statistics. The percent transaction cost are constructed using equation (2). The 5 day lagged bid ask spread weighted by time as a percent of the midquote. The annualized 21 day lagged close to close volatility. The order shares divided by the lagged 21 day median daily volume. Table 1 presents summary statistics of our data. The statistics weight each order by its fraction of dollar volume. The transaction cost standard deviation is large relative to the average cost. Hence the risk component appears to be substantial.

The rows labeled B and S break down the orders into buyer and seller orders respectively. 62% of the dollars traded were buys and 38% sells. Buy orders tend to be slightly more expensive on average in this sample. The risk is similar. We see that 75% of the dollar volume was for NYSE stocks and 25% for Nasdaq. We see that NYSE orders tend to cost less than NASDAQ by an average of about 5 basis points. It is important to note that these statistics are unconditional and do not control for differences in characteristics of the stocks traded on the two exchanges which might be driving some of the variation in the observed costs. For example, we see that the average volatility of NASDAQ stocks is substantially higher than that of NYSE.

The last four rows separate the orders by urgency. H, M, and L, correspond to high, medium and low urgencies and V is the VWAP strategy. Hence as we move down the rows we move from high cost, low risk strategies to low cost, high risk strategies. Almost half of the orders are the risk neutral VWAP strategy (46%). Only 10% of the order volume is high urgency, 24% is medium urgency and 20% is high urgency. This is reflected in the sample statistics. The average cost decreases from 11.69 basis points to 8.99 basis points as we move from high to low urgency orders. At the same time, the risk moves from 12.19 basis points up to 40.89 basis points for the same change in urgency. Contrary to the intent, the VWAP strategy does not exhibit the lowest cost at 9.69 basis points. It is the most risky however. Of course, the order submission may depend on the state of the market and characteristics of the order. These unconditional statistics will not reflect the market state and may blur the tradeoffs faced by the traders.

Table 2 presents the same summary statistics conditional on the size of the order relative to the 21 day median daily volume. The first bin is for orders less than a quarter of a percent of the 21 day median daily volume and the largest bin considered is for orders that exceed 1%. For each bin the statistics are presented for each type of order. The top of the table is for NYSE and the bottom half of the table is for NASDAQ stocks. Not surprisingly, for each order type, larger order sizes tend to be associated with higher average cost and higher risk indicating that larger orders are more difficult to execute along both the cost and risk dimensions.

This tradeoff can be seen clearly by plotting the cost/risk tradeoffs for each of the percent order size bins. Figure 1 presents the average cost/risk tradeoff for the NYSE stocks. Each contour indicates the expected cost/risk tradeoff faced for a given order size. Each contour is constructed using 4 points, the three urgencies and the VWAP. For a given contour, as we move from left to right we move from the high urgency orders to the VWAP. Generally speaking, the expected cost falls as the risk increases. This is not true for every contour, however. Increasing the percent order size shifts the entire frontier toward the north east indicating a less favorable average cost / risk tradeoff. Figure 2 presents the same plot but for NASDAQ stocks.

Contrary to what might be expected, some order size bins exhibit a cost increase as we move to less urgent strategies. These plots, however, do not consider the state of the market at the time the order is executed. It is entirely possible that the traders consider the state of the market when considering what type of urgency to associate with their order. If this is the case, a more accurate picture of the tradeoff faced by the trader can be obtained by considering the conditional frontier. This requires building a model for the expected cost and the standard deviation of the cost conditional on the state of the market. This is precisely the task considered in the next section of the paper.

4. Modeling the expected cost and risk of order execution.

Both the transaction cost and risk associated with trading a given order will vary depending on the state of the market. In this section we propose a modeling strategy for both the expected cost and risk of trading an order. The model is estimated using the Morgan Stanley execution data described in the previous section. In estimating this model for the expected cost and risk we are also estimating a conditional expected cost/risk frontier. This frontier depicts the expected cost/risk tradeoff faced by the agent given the current state of the market. This frontier will be a function of both the state of the market as well as the size of the order.

Both the mean and the variance of transaction costs are assumed to be an exponential function of the market variables and the order size. Specifically the transaction costs for the i^{th} order are given by:

$$TC\%_i = \exp(X_i\beta) + \exp\left(\frac{1}{2}X_i\gamma\right)\varepsilon_i \tag{4}$$

where $\varepsilon_i \sim iid\ N(0,1)$. The conditional mean is an exponential function of a linear combination of the X_i with parameter vector β. The conditional standard deviation is also an exponential function of a linear combination of the X_i with parameter vector γ. X_i is a vector of conditioning information. In our empirical work we find that the same vector X_i explains both the mean and the variance but this restriction is obviously not required.

The exponential specification for both the mean and the variance restricts both to be positive numbers. This is a natural restriction for both the mean and the variance. While the realized transaction cost for any given trade can be either positive or negative (and empirically we do find both signs), the *expected* transaction cost is positive.

We consider several factors that market microstructure theory predicts should contribute to the ease of executing a given order. The lagged 5 day time weighted average spread as a percent of the midquote. The log volatility constructed from the average close to close returns over the last 21 days. The log of the average historical 21 day median daily dollar volumes. In addition to these market variables we also condition on the log of the dollar value of the order and the urgency associated with the order. The urgency is captured by 3 dummy variables for high, medium, and low urgencies. The constant term in the mean and variance models therefore corresponds to the VWAP strategy.

The exponential specification for the mean is not commonly used in econometrics analysis. It is particularly useful here since it is natural to restrict the mean to be positive. The often used method of modeling the logarithm of the left hand side variable won't work here because the transaction cost often take negative values. Also, notice that $\ln[E(TC\%_i)] = X_i\beta$. Hence the coefficients can be interpreted as the percent change in TC% for a one unit change in X. Right hand side variables that are expressed as the logarithm of a variable (such as ln(value)) can be interpreted as an elasticity with respect to the non-logged variable (such as value).

The exponential model also allows for interesting nonlinear interactions that we might suspect should be present. Consider the expected transaction cost and the logged value and volatility variables. We have $E(TC\%) = \exp^{(\text{other variables})} value^{\beta_1} volatility^{\beta_2}$. If β_1 is larger than 1 then the cost increases more than proportionally to the value. If β_1 is smaller than 1 then the expected cost increases less than proportionally to the value. If β_1 and β_2 are both positive, then increases in the volatility result in larger increases in the expected cost for larger value trades. Alternatively, as the value of the trade goes to zero, so does the expected cost. It is entirely possible that the marginal impact of volatility might be different for different order sizes. The exponential model allows for this possibility in a very parsimonious fashion. Hence, what appears as a very simple nonlinear transformation allows for fairly rich nonlinear interactions. Obviously, using the exponential function for the variance has the same interpretation.

We estimate the model by maximum likelihood under the normality assumption for ε. It is well known that the normality assumption is a quasi maximum likelihood estimator. As long as the conditional mean and variance are correctly specified, we still obtain consistent estimates of the parameters even if the normality assumption is not correct. The standard errors, however, will not be correct in the event that the errors are not normal. Robust standard errors that are consistent in the event of non-normal errors can be constructed following White (1982) and are constructed for our parameter estimates.

The estimation is performed separately for NYSE and NASDAQ stocks. The two markets operate in a very different fashion and it is unlikely a single model would be appropriate for both trading venues. We have 166,508 NYSE orders and 67,405 Nasdaq orders. The parameter estimates for the variance equation for the NYSE stocks is given in table 3.

The coefficient on the spread is positive indicating that wider spreads are associated with more risk for any given order type and order size. The coefficient on the log volatility is 1.2. A simple model where a given order type is always executed over the same time interval with roughly constant quantities traded implies that the variance of the transaction cost should be proportional to the variance of the traded asset. To see this, consider the variance of the transaction cost when the local effects are fixed so that the $\mathrm{var}(\tilde{p}_t - p_t) = 0$. If equal quantities are traded in each time interval so that $\frac{\Delta x_t}{(x_T - x_0)} = \frac{1}{T}$, and the variance of the asset is constant and given by σ^2 then the variance of the transaction cost:

$$(7) \quad Var(TC\%) = Var\left(\frac{\sum_{t=1}^{T} \Delta x_t (\tilde{p}_t - p_0)}{(x_T - x_0)P_0} \right) = \left(\frac{1}{T}\right)^2 \sum_{t=1}^{T} Var(\tilde{p}_t - p_0) = \left(\frac{1}{T}\right)^2 \sum_{t=1}^{T} t\sigma^2 = \sigma^2 \left(\frac{T^2 + T}{2T^2} \right)$$

For large T this is approximately $\frac{\sigma^2}{2}$ but the variance of the transaction costs should be proportional to the variance of the asset even for small T. Recall that $Var(TC\%) = \exp^{(\text{other variables})} value^{\beta_1} volatiltiy^{\beta_2}$ so that it is therefore interesting to compare the estimated coefficient to the value 1. Squaring the volatility to convert the standard deviations to the variance $(volatility^2)^{\frac{\beta_2}{2}}$ yields a coefficient on the variance that is half the coefficient on the standard deviation which is .6 for the NYSE data. The variance of the transaction cost therefore increases less than proportionally to the variance of the asset. Thus, the Morgan Stanley BXS algorithm reduces the risk of the order relative to the simple constant volume, constant time interval strategy. This could happen for a number of reasons including front loading the trades, or more rapid execution in higher volatility markets.

The coefficient on the log of the average 21 day median volume is -.51. Every 1% increase in the volume translates into a half a percent decrease in the trading cost. The order size has a coefficient of .53 indicating larger orders have a higher risk. A 1% increase in the order size translates into about a half of a percent increase in the variance. It is interesting to notice that the coefficient on the order size is roughly the negative of the coefficient on the volume. This indicates that logarithm of the order size as a fraction of the daily volume that predicts the variance. Not surprisingly, the variance of the transaction cost is decreasing as the urgency increases. This is consistent with the high urgency orders executing more quickly than the low urgency orders.

Next we turn to the mean cost parameter estimates. The spread is positively related to the transaction cost. A 1% increase in the spread translates into about a 1% increase in the transaction cost. Recall that the transaction costs are already expressed as a percent so this is a percent increase in the percent transaction cost. Wider spreads are consistent with markets that are less liquid. The volatility has a coefficient of .50. Every 1% increase in the 21 day volatility translates into a half of a percent increase in the expected trading cost. High volatility is often thought to be associated more uncertainty and less liquid markets as we find here. The coefficient on the average 21 day median volume is -.47. Every 1% increase in the daily volume translates into about a half of a percent decease in the expected trading costs. The greater the volume the more liquid is the market.

The value has a coefficient of .43 indicating that a 1% increase in the value of the order translates into a little less than a half of a percent increase in the trading cost. It is again interesting to note that the coefficient on the value is roughly the same magnitude, but opposite sign as the coefficient on the volume. It appears that the size of the trade relative to the daily volume that predicts the cost. Finally, the cost is strictly increasing as the urgency increases.

The variance and mean model estimates for NASDAQ are presented in tables 5 and 6 respectively. While the magnitude of some of the estimates differs across the two exchanges the results are qualitatively very similar. We test the null hypothesis that the mean and variance models for the NYSE and NASDAQ are not different. This null hypothesis can be tested by an likelihood ratio test based on the difference between the sum of the likelihoods for the two unrestricted NYSE and NASDAQ models and the restricted model using the pooled data. Twice the difference in these two likelihoods will have a chi-squared distribution with degrees of freedom given by the number of restricted parameters, or 16. Twice the difference in the two likelihoods is 1954.76. The critical value is 26.29 so we overwhelmingly reject the null with a p-value near 0. Hence, while the models are qualitatively similar, there are statistically meaningful quantitative differences.

The parameter estimates provide intuitive interpretations regarding the transaction costs. It is nevertheless interesting to evaluate the statistical fit of the assumed exponential form. Toward this end we consider a variety of lagrange multiplier tests. The test can be performed for both omitted terms in the mean and the variance equations. Our null is that the exponential specification is sufficient while under the alternative we consider omitted linear and squared terms $E(TC\%_i) = \exp(X_i\beta) + Z_i\theta$ where Z will be taken to be X and X^2 or a combination of linear and squared terms. The test for the mean is performed by regressing the standardized error term on potential omitted terms. The standardized error term is given by $\hat{\varepsilon}_i = \frac{TC\%_i - E(TC\%_i)}{sd(TC\%_i)}$. We regress $\hat{\varepsilon}_i = \exp\left(X_i\hat{\beta}\right)X_i\theta_0 + Z_i\theta_1$ where X_i^2 is taken to mean the element by element square of each variable (ie no cross products are included). The θ_1 and θ_2 are conforming parameter vectors. Similarly, the test for the variance is performed by regressing $\hat{\varepsilon}_i^2 = \exp\left(X_i\hat{\beta}\right)X_i\phi_0 + Z_i\phi_1$ where the ϕ_0 and ϕ_1 are again conforming parameter vectors. The results of these test and the special cases of omitted linear terms only and omitted squared terms only are presented in table 7. (TABLE 7 IS NOT READY YET).

Generally speaking, larger orders are cost more to execute than smaller orders. We next look more closely at how the expected cost and risk vary as the order size increases. Figures 3 and 4 plot the expected cost and the standard deviation as a function of the order size relative to the 21 day average median volume. The plots consider orders ranging from near 0 percent up to 2% of the daily volume. The plots are done for an average stock on an average day. Figures 5 and 6 present the same plots, but for the NASDAQ stocks. In the expected cost plots, the higher curves correspond to the more urgent orders. The opposite is true for the standard deviation plots.

We can also look at the conditional risk/cost trade off by plotting the mean and volatility conditional upon the state of the market for each order type. We again consider the risk/cost tradeoff for an average stock under average conditions. These contours are plotted in figures 7 and 8. The ellipses represent 95% confidence intervals for the true mean and true variance for each order submission strategy. As we move from left to right we move from high urgency to medium, to low and finally VWAP or the risk neutral strategy. Perhaps the most interesting conclusion is that this analysis suggests that there is not much benefit to moving from low urgency to VWAP for either NYSE or NASDAQ stocks. The change in the expected cost is nearly zero while the increase in risk is substantial. If the agent cares at all about risk, the VWAP strategy does not appear viable.

Given the model, we can evaluate the cost/risk tradeoff under any stock. To get an idea of how this tradeoff varies as we examine how the frontier changes as we vary the order size for a typical stock on a typical day. Again, it is natural to express the order size relative to the average median 21 day volume. These plots are presented in figures 9 and 10 for typical NYSE and NASDAQ stocks respectively. The larger orders shift the cost/risk tradeoff to less desirable north east region. We again see that the order size effects on the cost/risk tradeoff are substantial.

5. Liquidation Value at Risk (LVAR)

Liquidation risk is the uncertainty about how much it costs to liquidate a position in a timely manner if the need should arise. Liquidation risk is important from both an asset management/risk perspective, as well as a more recent literature on asset pricing and liquidity (see for example Easley and O'Hara (2003), Pastor and Stambaugh (2003), Pedersen and Acharya (2005)). The conditional distribution of transaction costs is fundamentally related to liquidation risk. We show how the losses associated with liquidating an asset can be bounded with some probability. We call this measure liquidation value at risk or LVAR. Like the traditional value at risk (VaR), LVAR tells us the minimum number of dollars that will be lost with some probability α, when liquidating an asset.

For a given liquidation order the conditional mean and variance can be constructed. Under a normality assumption one can construct an α% LVAR given by:

(8)
$$LVAR(\alpha) = \exp\left(X\hat{\beta}\right) + \exp\left(\frac{1}{2}X\hat{\gamma}\right)z_{1-\alpha}$$

where $z_{1-\alpha}$ is the 1-α % quantile.

More generally, we might not wish to impose the normality assumption and instead use a more non-parametric approach. In the first stage, consistent estimates or the parameters can be estimated by QMLE. In the second stage, the standardized residuals can be used to construct a non-parametric estimate of the density function of the errors ε. The standardized residuals are given by:

(9)
$$\hat{\varepsilon}_i = \frac{TC\%_i - \exp\left(X_i\hat{\beta}\right)}{\exp\left(\frac{1}{2}X_i\hat{\gamma}\right)}$$

A non-parametric estimate of the density or perhaps just the quantiles themselves can then be used to construct a semi-parametric LVAR. Specifically, let $\hat{\varepsilon}_{1-\alpha}$ denote a non-parametric estimate of the α% quantile of the density function of the error term ε. Then the semi-parametric $\alpha\%$ LVAR is obtained by replacing $z_{1-\alpha}$ with the non-parametric quantile $\hat{\varepsilon}_{1-\alpha}$:

$$LVAR(\alpha) = \exp\left(X\hat{\beta}\right) + \exp\left(\frac{1}{2}X\hat{\gamma}\right)\hat{\varepsilon}_{\alpha} \tag{10}$$

Figures 11 and 12 present the standardized residuals for the NYSE and NASDAQ models. The residuals are clearly non-normal. We use the empirical quantiles of the data to construct the LVAR. Figures 13 and 14 present the 1% LVAR associated with the high, medium and low urgency orders as well as the VWAP. The LVAR estimates are constructed for typical stocks on a typical day. The vertical axis is the transaction cost in basis points. As we move from left to right we move from LVAR to low urgency to the high urgency orders. The LVAR is given by the upper bar for each order type. The expected cost for each order type is given by the smaller bar in near the origin. The differences in the mean are small relative to the changes in the risk across the different order types. Since the risk dominates, the minimum LVAR order type here is given by the most aggressive strategy, the high urgency order. The 1% LVAR for this order type is just under around half a percent for NYSE and 1% for NASDAQ. For each order type the lower dashed line completes a 98% prediction interval.

6. Conclusion

This paper demonstrates that expected cost and risk components of transaction costs can be estimated from detailed transaction data. We show that we can construct a cost/risk tradeoff in the spirit of classical portfolio analysis. We find that the expected cost and risk components can be successfully modeled using an exponential specification for the mean and variance. Characteristics of the order and state of the market play a major role in determining the cost/risk tradeoff faced by the trader.

We provide an example of how this approach can be used to asses liquidation risk using the notion of liquidation value at risk (LVAR). This is, of course, only one approach that could be taken in assessing liquidation risk. More generally, we have the entire conditional distribution of transaction costs so there are potentially many approaches that one could take in assessing liquidation risk.

Finally, our data here consists of the transaction costs. Another direction to go would be to directly consider the raw transaction data set. In this way, we could better asses the dynamics of the price impact functions. For example, how large are the local vs. price impact effects?

Exchange	Side	Benchmark	Urgency	Weight	Count	Price	Spread	Volatility	Volume	Capitalization (000)	Order Value	Order Shares	Cost (BP)	StDev (BP)
				100%	233,913	$ 45.07	0.09%	26%	1.59%	$ 59,609,060	$ 310,472	9,154	10.09	47.24
	B			62%	147,649	$ 45.06	0.09%	26%	1.57%	$ 58,137,900	$ 302,812	8,946	10.77	47.17
	S			38%	86,264	$ 45.09	0.08%	26%	1.62%	$ 61,965,453	$ 323,583	9,512	8.99	47.31
NYSE				75%	166,508	$ 48.01	0.09%	23%	1.68%	$ 66,717,110	$ 326,031	8,701	8.82	43.28
NASDAQ				25%	67,405	$ 36.38	0.08%	36%	1.33%	$ 38,565,201	$ 272,037	10,273	13.84	57.19
		A	H	10%	15,616	$ 47.81	0.08%	26%	1.18%	$ 60,838,460	$ 475,462	12,845	11.69	23.19
		A	M	24%	54,095	$ 44.73	0.09%	27%	1.47%	$ 48,482,781	$ 320,909	9,688	11.09	32.20
		A	L	20%	51,588	$ 46.44	0.08%	26%	1.13%	$ 68,894,693	$ 285,018	8,106	8.99	40.89
		V		46%	112,614	$ 44.04	0.09%	26%	1.95%	$ 61,042,206	$ 294,240	8,867	9.69	59.01

Table 1. Summary statistics for Morgan Stanley trades. B and S are buy and sell orders respectively. A denotes arrival price strategy and V denotes VWAP strategy. H, M, and L denote high medium and low urgency trades.

Exchange	Benchmark	Urgency	Volume Range	Weight	Count	Price	Spread	Volatility	Volume	Capitalization (000)	Order Value	Cost (BP)	StDev (BP)
NYSE	A	H	≤ 0.25%	0.69%	2,630	$ 47.77	0.08%	22%	0.19%	$ 94,061,545	$ 190,107	4.22	10.97
NYSE	A	M		3.10%	13,664	$ 48.06	0.08%	22%	0.16%	$ 95,714,257	$ 164,767	3.69	11.64
NYSE	A	L		4.17%	19,379	$ 50.15	0.08%	22%	0.14%	$ 98,487,506	$ 156,371	2.71	12.74
NYSE	V			6.54%	38,116	$ 47.46	0.08%	22%	0.13%	$ 90,733,082	$ 124,606	1.97	34.56
NYSE	A	H	≤ 0.5%	1.69%	3,559	$ 51.00	0.08%	22%	0.37%	$ 81,476,811	$ 345,605	6.16	11.95
NYSE	A	M		3.30%	9,557	$ 49.09	0.08%	23%	0.36%	$ 68,991,562	$ 250,562	5.68	15.53
NYSE	A	L		2.99%	8,027	$ 50.18	0.08%	21%	0.36%	$ 92,105,328	$ 270,537	4.15	19.65
NYSE	V			5.00%	14,890	$ 48.23	0.08%	23%	0.37%	$ 65,205,449	$ 244,088	3.06	42.27
NYSE	A	H	≤ 1.0%	2.39%	2,979	$ 52.69	0.08%	22%	0.73%	$ 72,154,182	$ 582,738	8.93	15.98
NYSE	A	M		3.64%	6,907	$ 48.85	0.08%	24%	0.72%	$ 54,822,342	$ 383,195	7.54	20.67
NYSE	A	L		3.17%	6,035	$ 50.22	0.08%	22%	0.72%	$ 82,177,737	$ 381,487	6.76	28.64
NYSE	V			6.40%	11,822	$ 46.88	0.09%	23%	0.73%	$ 71,181,217	$ 393,015	5.17	47.42
NYSE	A	H	> 1.0%	2.67%	2,549	$ 51.53	0.09%	23%	2.31%	$ 48,375,509	$ 760,186	14.36	25.12
NYSE	A	M		7.34%	7,052	$ 47.41	0.10%	24%	3.02%	$ 36,941,147	$ 755,825	15.55	38.97
NYSE	A	L		4.77%	5,632	$ 47.78	0.10%	23%	2.55%	$ 52,146,046	$ 615,387	12.64	52.73
NYSE	V			16.88%	13,710	$ 45.67	0.09%	23%	3.95%	$ 54,313,103	$ 894,247	14.66	65.61
NASDAQ	A	H	≤ 0.25%	0.28%	715	$ 36.70	0.06%	33%	0.18%	$ 76,072,230	$ 288,664	6.58	12.91
NASDAQ	A	M		1.59%	5,771	$ 37.81	0.06%	34%	0.15%	$ 52,884,846	$ 200,367	6.05	17.13
NASDAQ	A	L		1.36%	5,065	$ 37.49	0.06%	33%	0.14%	$ 77,185,322	$ 195,338	5.29	17.93
NASDAQ	V			3.73%	16,227	$ 39.02	0.06%	34%	0.11%	$ 71,786,694	$ 167,134	3.96	42.05
NASDAQ	A	H	≤ 0.5%	0.82%	1,131	$ 39.07	0.06%	32%	0.37%	$ 63,042,552	$ 527,242	10.76	18.02
NASDAQ	A	M		1.35%	4,455	$ 36.72	0.08%	36%	0.36%	$ 30,091,409	$ 220,816	9.15	21.53
NASDAQ	A	L		0.89%	2,578	$ 38.02	0.07%	36%	0.37%	$ 37,310,972	$ 251,330	8.33	31.04
NASDAQ	V			1.64%	5,597	$ 35.82	0.07%	36%	0.36%	$ 45,881,902	$ 213,288	6.07	59.83
NASDAQ	A	H	≤ 1.0%	0.85%	992	$ 38.30	0.07%	35%	0.70%	$ 42,108,883	$ 623,028	15.81	21.17
NASDAQ	A	M		1.31%	3,453	$ 36.58	0.09%	38%	0.71%	$ 15,709,279	$ 275,812	14.78	30.01
NASDAQ	A	L		0.97%	2,003	$ 39.11	0.07%	37%	0.72%	$ 32,386,856	$ 352,916	12.36	45.50
NASDAQ	V			1.80%	5,310	$ 34.08	0.08%	36%	0.73%	$ 36,671,191	$ 246,208	10.62	66.66
NASDAQ	A	H	> 1.0%	0.83%	1,061	$ 37.41	0.10%	36%	2.91%	$ 10,249,188	$ 565,871	26.99	47.16
NASDAQ	A	M		2.26%	3,236	$ 32.85	0.11%	39%	3.10%	$ 8,045,030	$ 508,137	22.96	58.51
NASDAQ	A	L		1.91%	2,869	$ 36.91	0.10%	38%	2.85%	$ 15,236,379	$ 484,223	26.02	79.01
NASDAQ	V			3.62%	6,942	$ 33.30	0.11%	37%	3.46%	$ 23,073,584	$ 379,150	24.65	94.67

Table 2. Summary statistics for Morgan Stanley trades. Volume is the order size as a percent of the average daily volume. A denotes arrival price strategy and V denotes VWAP strategy. H, M, and L denote high medium and low urgency trades.

VARIABLE	COEFFICIENT	ROBUST T-STAT
Const	11.80559	90.86393
Spread	1.815802	14.39896
Log volatility	1.207152	64.98954
Log volume	-0.51614	-55.6044
Log value	0.536306	46.08766
Low urg	-1.45436	-83.2275
Med urg	-1.92541	-61.058
High urg	-2.33731	-88.2665

Table 3. Variance parameter estimates for NYSE stocks.

VARIABLE	COEFFICIENT	ROBUST T-STAT
Const	5.173827	30.0342
Spread	0.969804	8.395586
Log volatility	0.503987	21.14475
Log volume	-0.47084	-43.4163
Log value	0.43783	40.27979
Low urg	0.094929	2.41284
Med urg	0.305623	8.796438
High urg	0.41034	11.28093

Table 4. Mean parameter estimates for NYSE stocks.

VARIABLE	COEFFICIENT	ROBUST T-STAT
Const	11.40519	71.44173
Spread	2.016026	17.00666
Log volatility	1.078963	40.27656
Log volume	-0.44182	-46.7497
Log value	0.453704	42.64865
Low urg	-1.04398	-42.1013
Med urg	-1.70511	-65.0131
High urg	-2.10623	-48.7398

Table 5. Variance parameter estimates for NASDAQ stocks.

VARIABLE	COEFFICIENT	ROBUST T-STAT
Const	5.354067	26.04098
Spread	1.014023	8.734035
Log volatility	0.513628	16.96502
Log volume	-0.41447	-29.9304
Log value	0.376208	24.99588
Low urg	0.025356	0.243943
Med urg	0.230764	5.716912
High urg	0.282479	6.156517

Table 6. Mean parameter estimates for NASDAQ stocks.

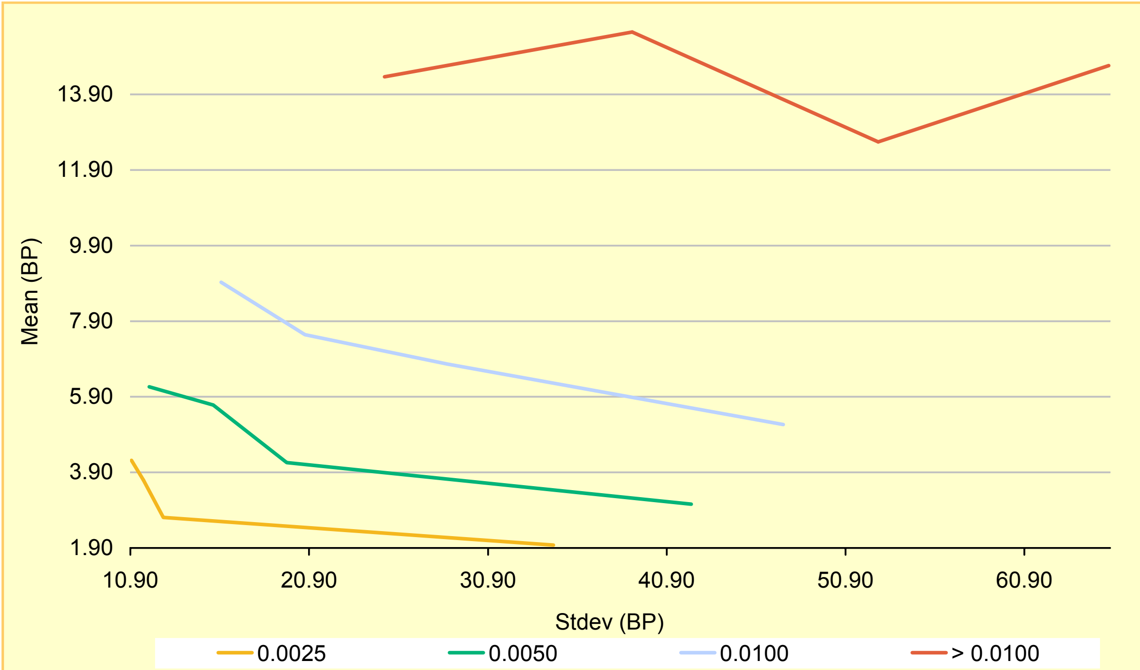


Figure 1: NYSE average cost/risk tradeoff given the order size. The order size is expressed as a fraction of the median 21 day daily volume.

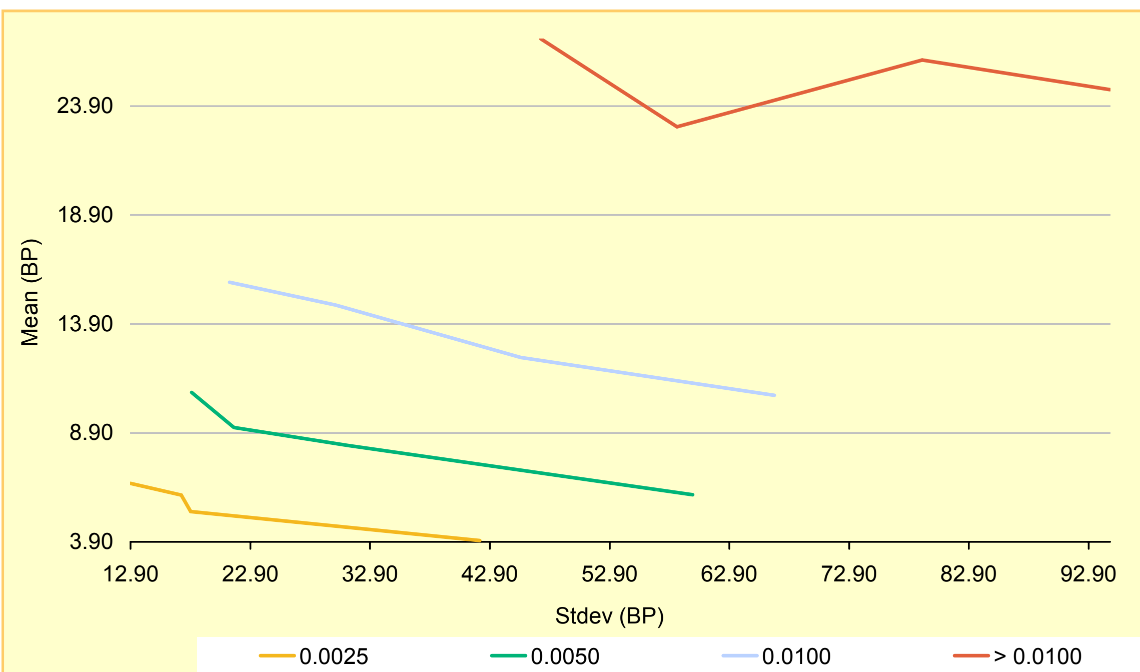


Figure 2: NASDAQ average cost/risk tradeoff given the order size. The order size is expressed as a fraction of the median 21 day daily volume.

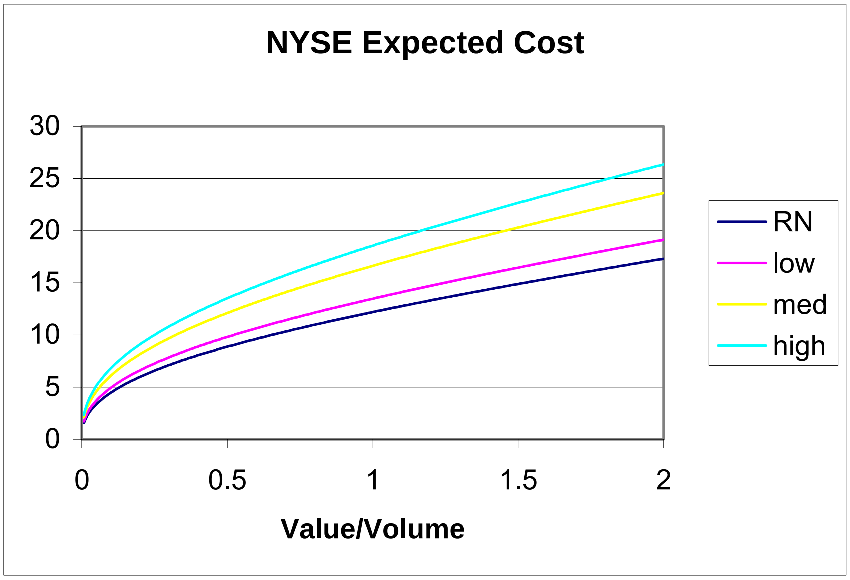


Figure 3: Expected Cost as a function of the order size expressed as a fraction of average daily volume for NYSE stocks.

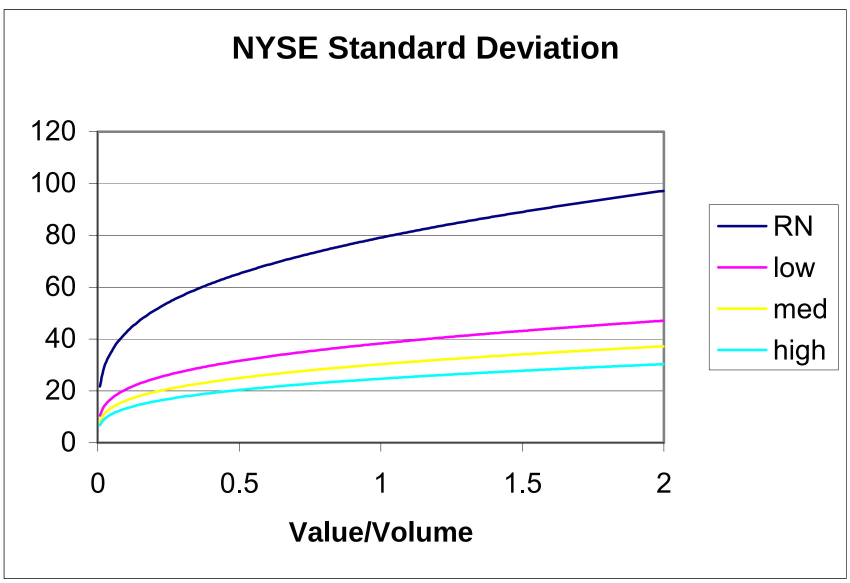


Figure 4: Standard deviation of transaction cost as a function of the order size expressed as a fraction of average daily volume for NYSE stocks.

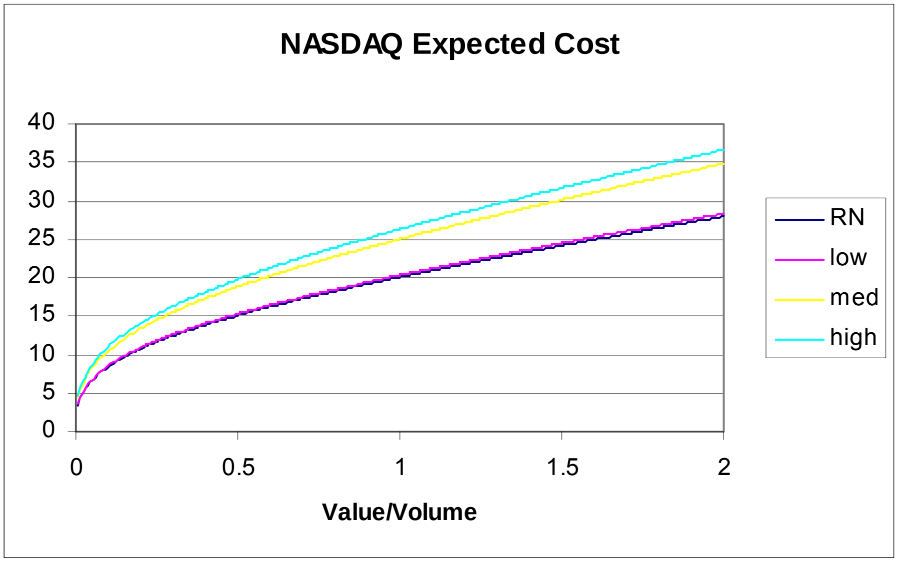


Figure 5. Expected Cost as a function of the order size expressed as a fraction of average daily volume for NASDAQ stocks.

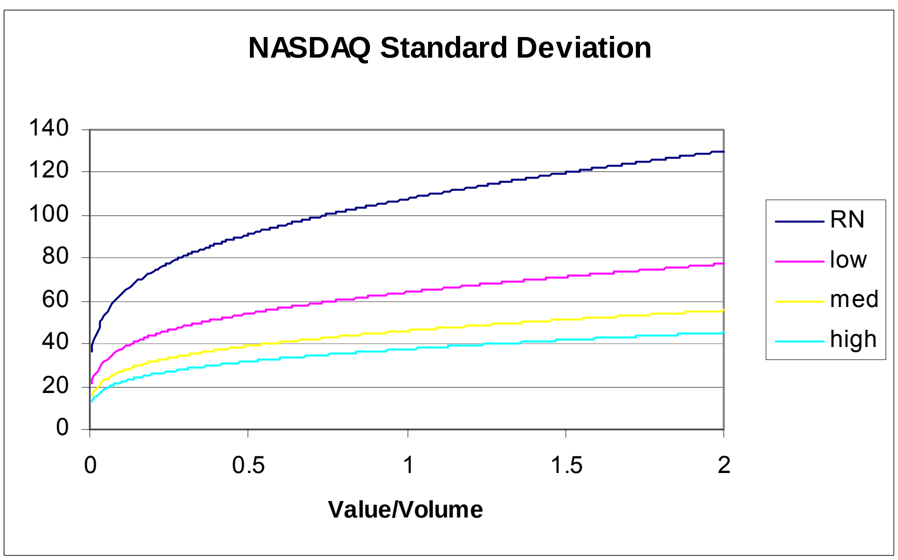


Figure 6. Standard deviation of transaction cost as a function of the order size expressed as a fraction of average daily volume for NASDAQ stocks.

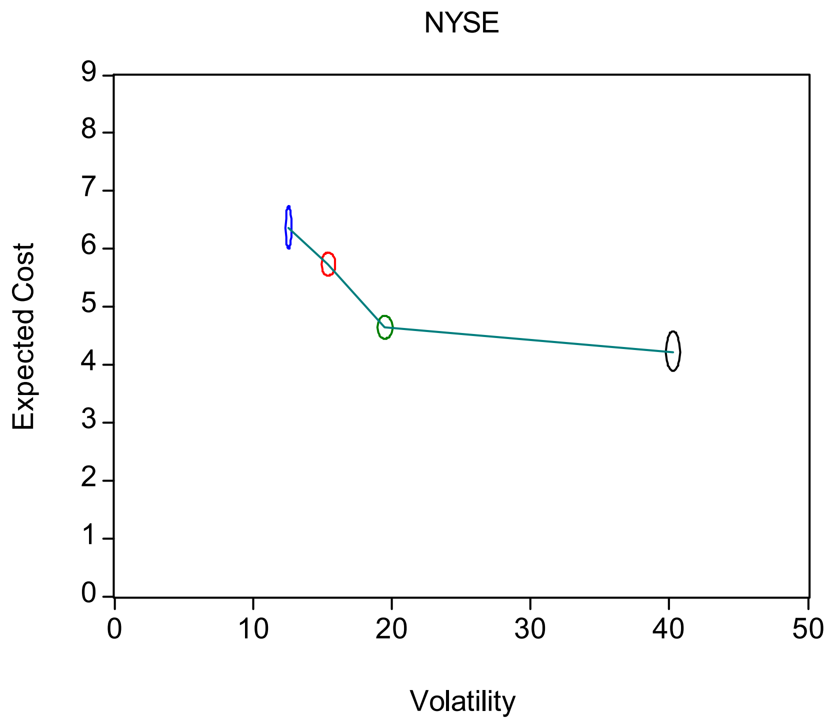


Figure 7. Expected cost and risk frontier for a typical NYSE stock on a typical day.

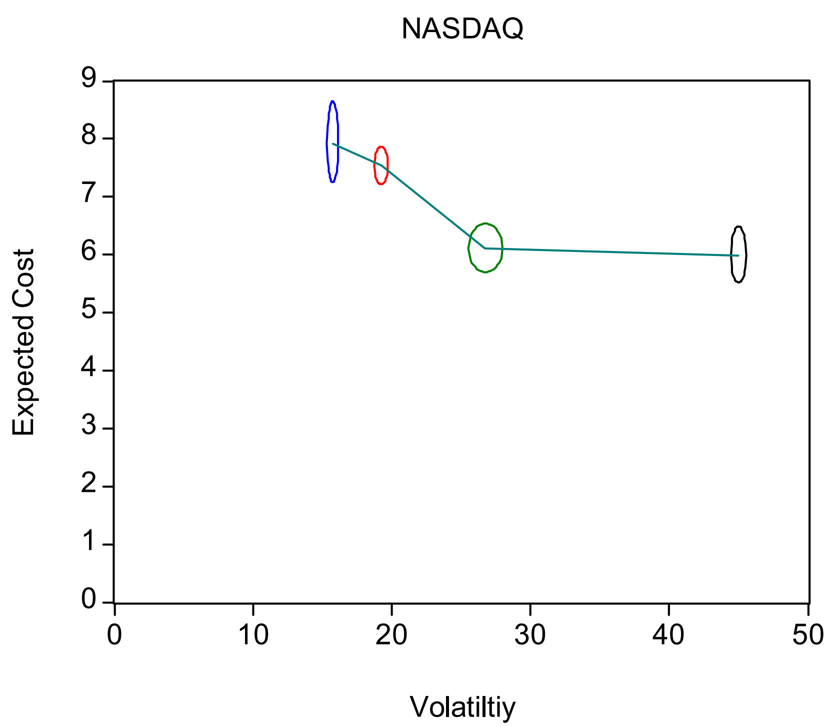


Figure 8. Expected cost and risk frontier for a typical NASDAQ stock on a typical day.

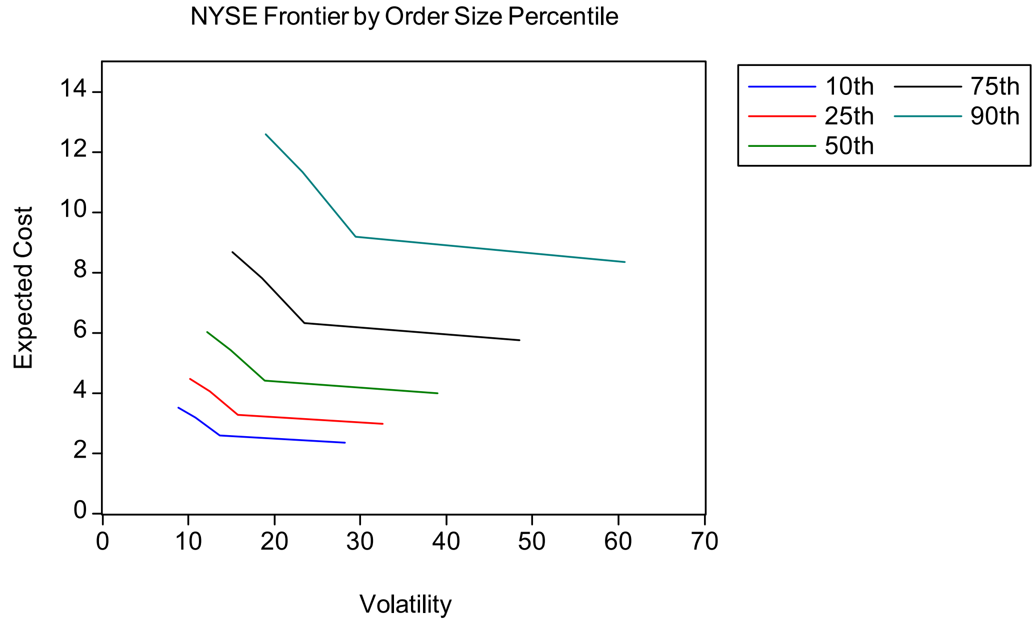


Figure 9. Expected cost/risk frontier for a typical NYSE stock on a typical day. Each contour represents the frontier for a different quantile of order size expressed as a fraction of average daily volume.

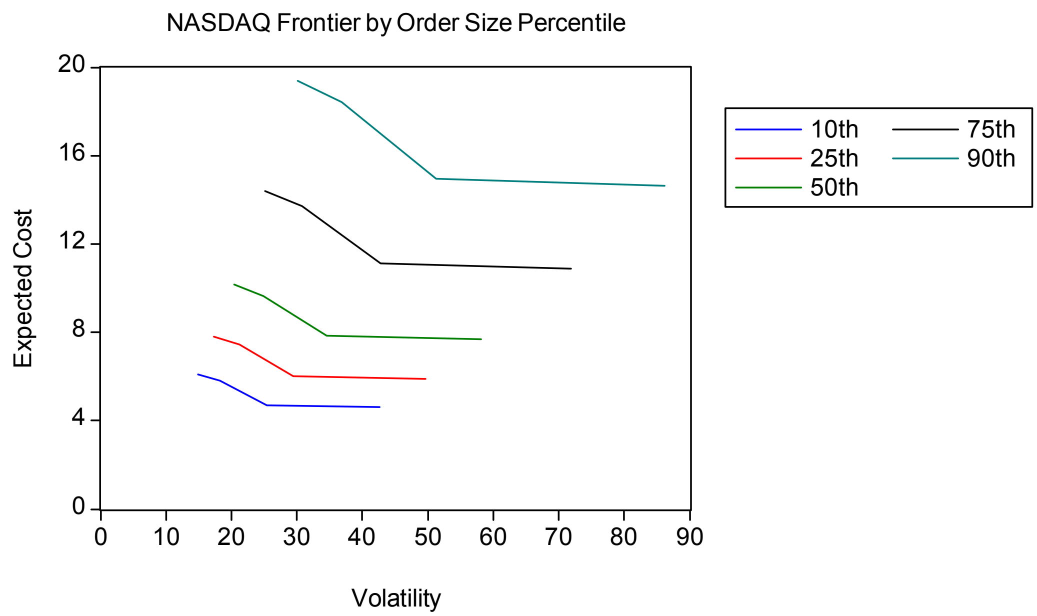


Figure 10. Expected cost/risk frontier for a typical NASDAQ stock on a typical day. Each contour represents the frontier for a different quantile of order size expressed as a fraction of average daily volume.

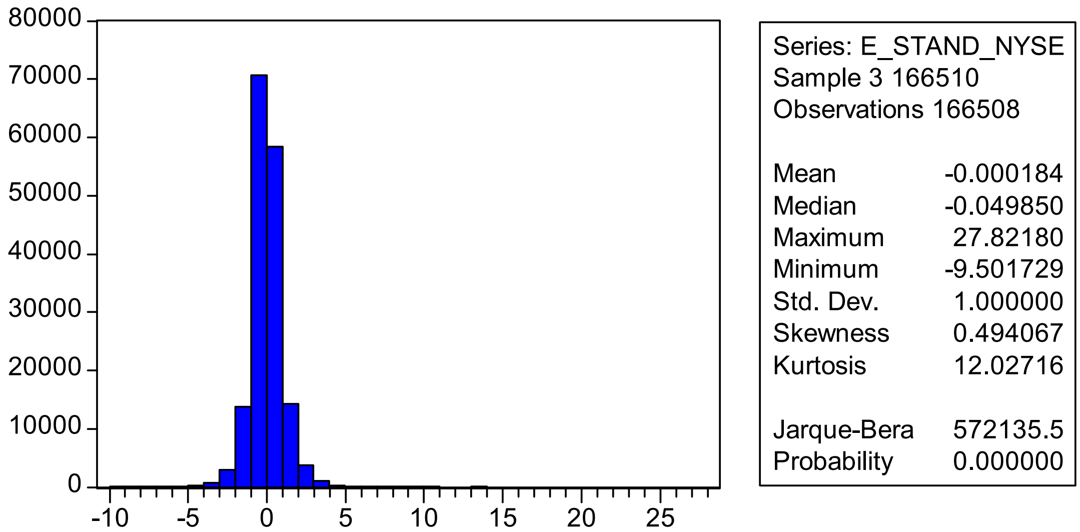


Figure 11. Standardized residuals for NYSE stocks.

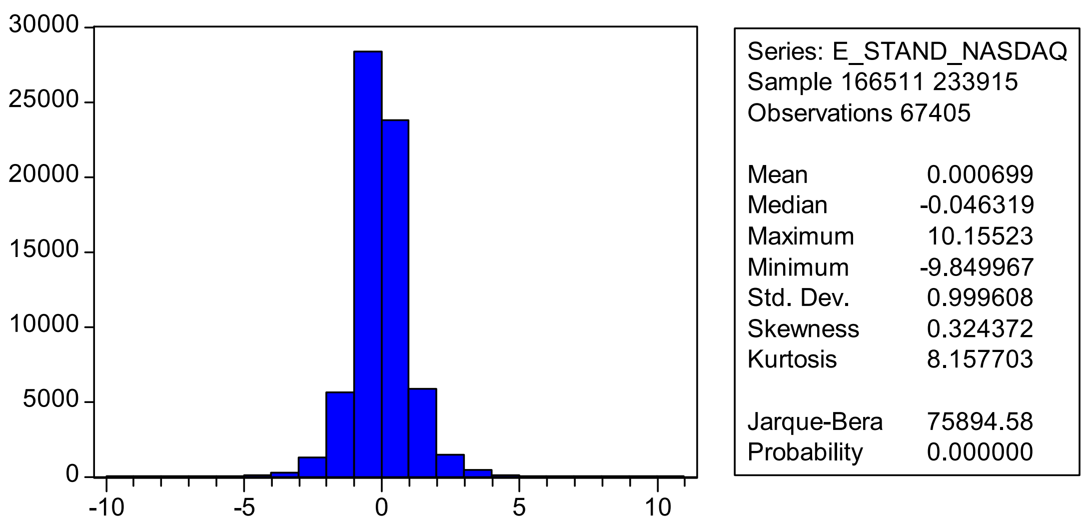


Figure 12. Standardized residuals for NASDAQ stocks.

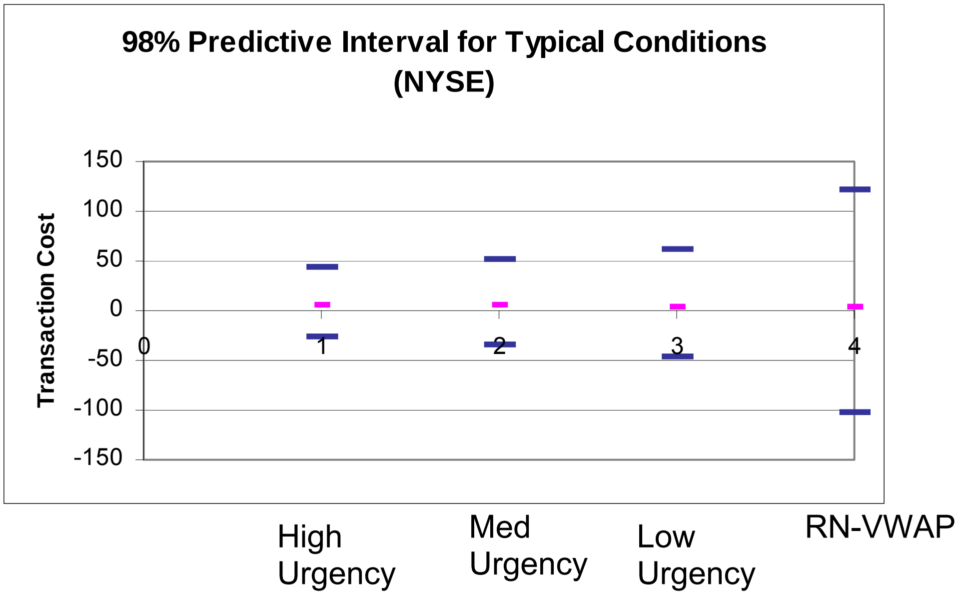


Figure 13. This plot shows the 98% predictive interval for the transaction cost for a typical NYSE stock on a typical day. 0 corresponds to VWAP, 1 to low urgency, 2 to medium urgency and 3 to high urgency. For each trade type, the upper bar denotes the 1% LVAR.

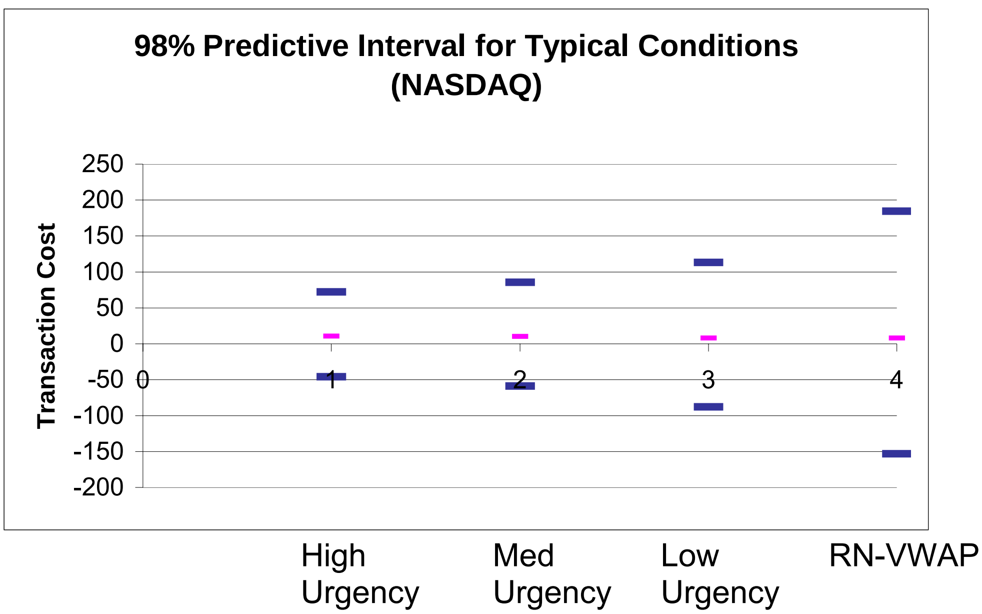


Figure 14. This plot shows the 98% predictive interval for the transaction cost for a typical NASDAQ stock on a typical day. 0 corresponds to VWAP, 1 to low urgency, 2 to medium urgency and 3 to high urgency. For each trade type, the upper bar denotes the 1% LVAR.

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

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 Proices around Institutional Trades", *Journal of Finance,* Vol 50, No. 4, pp1147-1174

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is information risk a determinant of asset returns? *Journal of Finance* 57, 2185-2222.

Engle, Robert, and Robert Ferstenberg, 2006, Execution Risk, NYU manuscript

Grinold R. and R. Kahn(1999) *Active Portfolio Management* (2nd Edition) Chapter 16 pp473-475, McGraw-Hill

Obizhaeva, Anna and Jiang Wang (2005) "Optimal Trading Strategy and Supply/Demand Dynamics" manuscript

Pastor, Lubos, Robert Stambaugh, (2003), "Liquidity Risk and Expected Stock Returns", *Journal of Political Economy* 111, 642–685.

White, Halbert, 1982, Maximum Likelihood Estimation of Misspecified Models, Econometrica, 50, 1-25.

EXECUTION RISK

Robert Engle and Robert Ferstenberg[1]
New York University and Morgan Stanley

CENTER FOR FINANCIAL ECONOMETRICS

March 27, 2006

PRELIMINARY

PLEASE DO NOT QUOTE WITHOUT AUTHORS PERMISSION

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 and price impacts during a more gradual trading trajectory. 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.

[1] 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 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:

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

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,…,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

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

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:

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

Generally this problem is defined in returns but this dollar-based formulation is equivalent. If there is a collection of assets available with known 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

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

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

$$TC\% = TC / \left(x_T - x_0\right)' p_0 \tag{5}$$

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

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

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 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

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

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. Equivalently, by postulating a mean variance utility function for trading with risk aversion parameter $\lambda *$, we could solve for the trading strategy by

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

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 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

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

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, 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 (10) and (11) to be

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

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:

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

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

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

or equivalently

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

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 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,...,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

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

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

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

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

$$\max_{x_T} E\left(x_T'\left(p_{T_2}-p_T\right)\right)-\lambda V\left(x_T'\left(p_{T_2}-p_T\right)\right) \tag{18}$$

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

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:

$$RF = r^f y_0 T \tag{20}$$

hence the annualized Sharpe ratio is given by

$$Sharpe\,Ratio = \frac{E\left(x_T\,'\left(p_T - p_0\right)\right) - E\left(TC\right) - RF}{\sqrt{T}\sqrt{V\left(x_T\,'\left(p_T - p_0\right) - TC\right)}}\ . \tag{21}$$

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

$$Sharpe\,Ratio = \frac{E\left(x_T\,'\left(p_T - p_0\right)\right) - E\left(TC\right) - RF}{\sqrt{T}\sqrt{V\left(x_T\,'\left(p_T - p_0\right)\right) + V\left(TC\right) - 2Cov\left(x_T\,'\left(p_T - p_0\right), TC\right)}} \tag{22}$$

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.

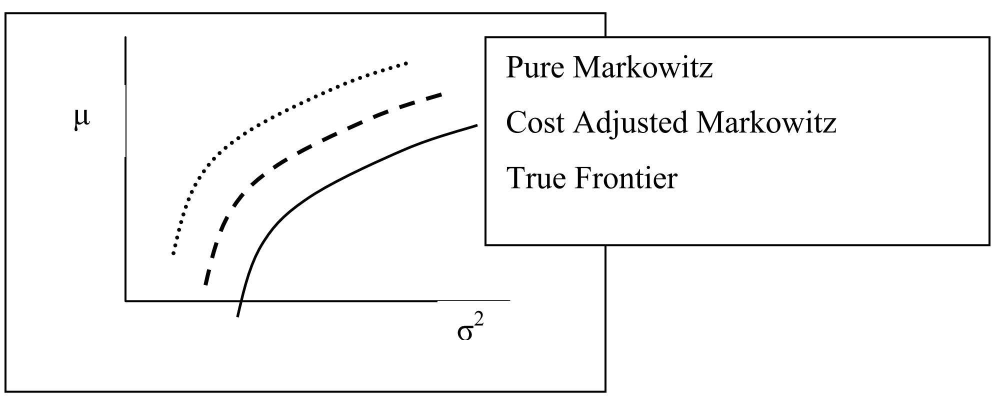
$$\max_{\{x_1,\ldots,x_{T-1}\}\{x_T\}} E\left[x_T\,'\left(p_T - p_0\right)\right] - \lambda V\left[x_T\,'\left(p_T - p_0\right)\right] \tag{23}$$

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

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

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.

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

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

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.

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

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.

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

The net risk when some positions are being increased and others are being 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)

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

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

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

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

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

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 $$x_T = \frac{1}{2\lambda}\Omega^{-1}\mu \qquad (32)$$

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

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

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 $$\tilde{p}_t - p_t = \mathrm{T}\Delta x_t \qquad (34)$$

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

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 $\left(\mu,\Omega\right)$ 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

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

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

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

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:

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

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 $x_t = \left(x_{1,t}', x_{2,t}' \right)' = \left(x_{1,t}', x_{2,T}' \right)'$. 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:

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

which has a solution:

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

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.

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

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 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

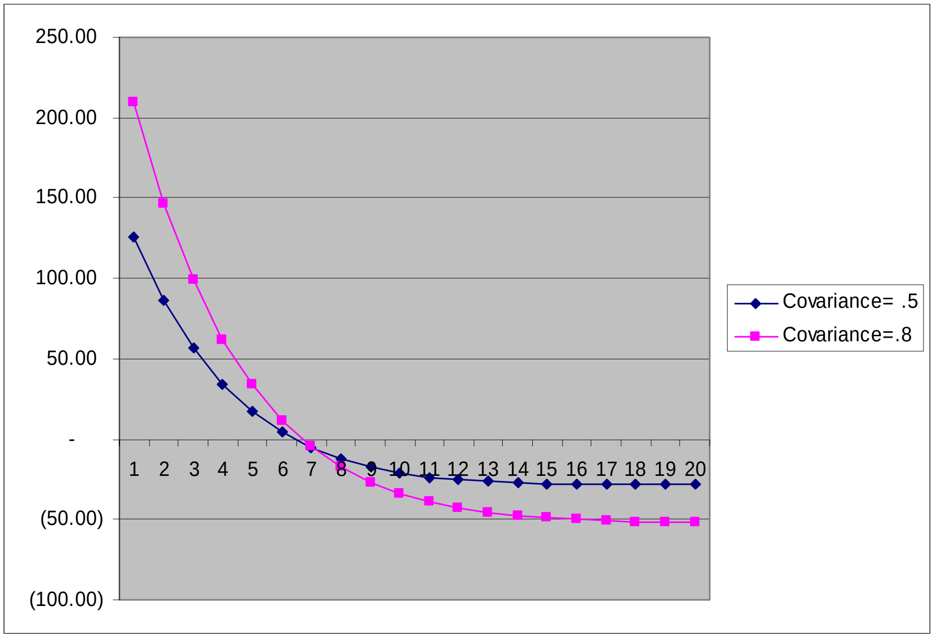


Figure 2

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.

A.2.b $$\tilde{p}_t = p_t + \mathrm{T}\Delta x_t + \mathrm{T}_1\Delta x_{t-1} + ... + \mathrm{T}_q\Delta x_{t-q} = p_t + \mathrm{T}(L)\Delta x_t \qquad (42)$$

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

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

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

and adding assumption A.3, it is

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

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. 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.

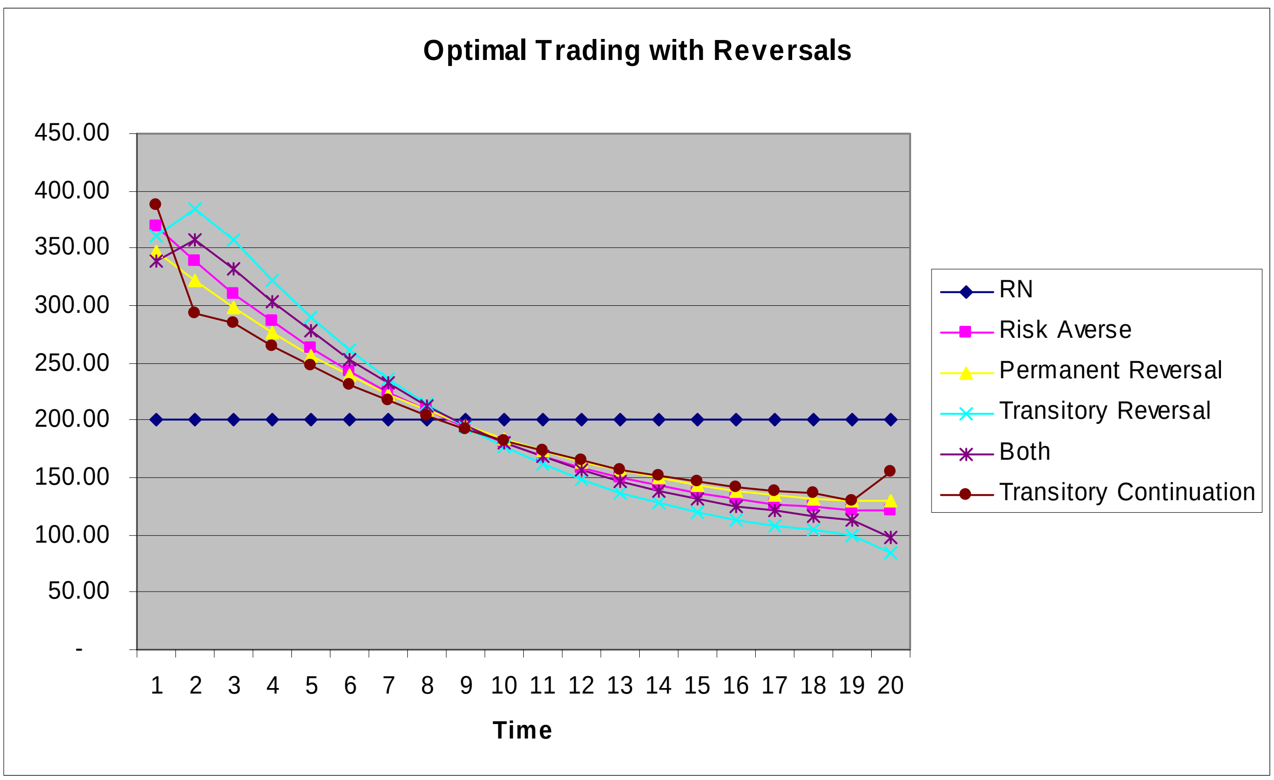


Figure 3

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

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

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

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

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 Proices around Institutional Trades", *Journal of Finance,* Vol 50, No. 4, pp1147-1174

Colacito, Riccardo and Robert Engle(2004), "Multiperiod Asset Allocation with Dynamic Volatilties", 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,