

# Transaction Cost Analytics in DRIP

Lakshmi Krishnamurthy

**v2.30**, 24 September 2016

**Optimal Execution of Portfolio Transactions**

**Overview, Scope, and Key Results**

1. Portfolio Transactions under Market Impact: Almgren and Chriss (2000) consider the execution of portfolio transactions with the aim of minimizing a combination of volatility risk and transaction costs arising from temporary and permanent market impact.
2. Efficient Frontier under Linear Cost: For a simple linear cost model, they explicitly construct an *efficient frontier* in the space of time-dependent liquidation strategies, which have the minimum expected cost for a given level of uncertainty.
3. Choice of the Utility Function: This enables one to select optimal strategies either by minimizing a quadratic utility function, or by minimizing the Value-at-Risk.
4. Liquidity Adjusted Value at Risk: The latter choice leads to the concept of liquidity-adjusted VaR, or L-VaR, that explicitly considers the best trade-off between the volatility risk and the liquidity costs.

**Motivation Background, and Synopsis**

1. Transactions Changing the Portfolio Composition: Almgren and Chriss (2000) consider the optimal execution of portfolio transactions that move a portfolio from a given staring composition to a specified final composition within a specified period of time.
2. The Bertsimas and Lo Approach: Bertsimas and Lo (1998) define the best execution as the dynamic trading strategy that provides the minimum cost of trading over a fixed period of time, and they also show that under a variety of circumstances one can find such a strategy by employing a dynamic optimization procedure; but they ignore the volatility of revenues of different trading strategies.
3. Maximization of Expected Trading Revenue: Almgren and Chriss (2000) work in the more general framework of maximizing the *expected revenue* – or equivalently minimizing the costs – with a suitable penalty for the *uncertainty* of revenue (or cost).
4. Market Microstructure Framework: This general framework arises in the market microstructure theory, but with a different purpose in mind. The *uninformed discretionary trader* trades an exogenous endowment over an exogenously specified amount of time to maximize the profits (Admati and Pfleiderer (1988)); the informed strategic trader trades over multiple periods on information not widely available, again to maximize profits (Kyle (1985)). In both cases the literature focuses on the link between the trader and the market maker, and a theory is produced to predict the market clearing price of the security at each period. Thus a trader’s optimal strategy is used as a means to study the price formation in the markets, not as an object of interest in itself.
5. Variance of the Trading Cost: Almgren and Chriss (2000) study the variance of the trading cost in optimal execution because it fits in with the intuition that the trader’s utility should figure in the definition of *optimal* in “optimal execution”.
6. Example: Trading Illiquid Volatile Securities: For example, in trading a highly illiquid, volatile security, there are two extreme outcomes; trade everything now at a known, but high, cost, or trade in equal sized packets over a fixed time at a relatively lower cost. The latter strategy has a lower expected, but this comes at the expense of greater uncertainty in the final revenue.
7. Estimation of the Trading Uncertainty: How to evaluate the above uncertainty is partly subjective, and is a function of the trader’s tolerance for risk. All that can be done is to insist that for a given level of uncertainty that the cost be minimized. This idea extends to a complete theory of optimal execution that includes an efficient frontier of optimal execution strategies.
8. Consistency with Expectations from Intuition: The framework of risk in execution yields several results that are consistent with the intuition. For example, it is evident that all else equal, a trader will choose to execute a block of illiquid security less rapidly than a liquid security.
9. Models Lacking Consistency with Intuition: While this seems obvious, Almgren and Chriss (2000) demonstrate that a model that ignores risk does not have this property; without enforcing a strictly positive penalty for risk one cannot produce models that trade differently across the spectrum of liquidity.
10. Arithmetic Brownian Motion Price Dynamics: The incorporation of risk into optimal execution does not come without cost. First, in order to be able to produce tractable analytical results, Almgren and Chriss (2000) are forced to work in largely in the framework of price dynamics that are an arithmetic walk with independent increments.
11. Use of Static Optimization Procedures: They obtain results using *static optimization* procedures which they show lead to globally optimal trading trajectories. That is, optimal trading paths may be determined in advance of trading. Only the composition of the portfolio and the trader’s utility function figure on the trading path.
12. Why does Static Optimization Work? The fact that the static strategy can be optimal even when the trader has the option to dynamically change his trading mid-course is a direct result of the assumptions of independence of returns and symmetry for the penalty functions for risk.
13. Using Non-Symmetric Penalty Functions: An interesting deviation from the symmetric penalty function was communicated by Ferstenberg, Karchmer, and Malamut at ITG Inc. They argue that the opportunity is a subjective quantity and is measures differently by different traders. Using a trader defined cost function , they define opportunity costs as the expected costs of applied to the average execution price obtained by the trader relative a benchmark price. They assume that the risk-averse traders will use a convex function that is not symmetric in the sense that there is a strictly greater penalty for underperformance than for the same level of outperformance. They show that in this setting, the optimal strategy relative to not only depends on the time remaining, but also on the performance of the strategy up to the present time, and the present price of the security. In particular, this means that in their setting, optimal strategies are dynamic.
14. Serial Correlations among Price Movements: As it is well known that price movements exhibit some serial correlations across various time horizons (Lo and MacKinlay (1988)), that market conditions change, and that some participants possess private information (Bertsimas and Lo (1998)), one may question the usefulness of results that obtain strictly in an independent-increment framework.
15. The Dynamic Nature of Trading: Moreover, as trading is known to be a dynamic process, the conclusion that optimal trading strategies can be statically determined calls for critical examination. Almgren and Chriss (2000) examine what quantitative gains are available that incorporate all the relevant information.
16. Impact of the Serial Correlations: First they consider short term serial correlations in price movements. They demonstrate that the marginal improvements available by explicitly incorporating this information into trading strategies is small, and more importantly, independent of the portfolio sizes; as portfolio sizes increase, the percentage gains possible decrease proportionately.
17. Combining “Correlated” and “Shifting” Strategies: The above is precisely true for linear transaction cost models, and is approximately true for more general models. The results of Bertsimas and Lo (1998) suggest that trading a strategy built to take advantage of serial correlation will essentially be a combination of a “correlation free” strategy and a “shifting strategy” that moves from one trade period to the next based on the information available in the last period’s return. Therefore Almgren and Chriss (2000) argue that by ignoring serial correlation, they a) preserve the main interesting features of their analysis, and b) introduce virtually no bias away from “truly optimal” solutions.
18. Impact of Scheduled News Events: Second, Almgren and Chriss (2000) examine the impact of scheduled new events on optimal execution strategies. There is ample evidence that anticipated news announcements, depending on their outcome, can have a significant temporary impact on the parameters governing price movements.
19. Scheduled News Events - Literature Review: For a theoretical treatment see Brown, Harlow, and Tinic (1988), Kim and Verrecchia (1991), Easterwood and Nutt (1999), and Ramaswami (1999). For empirical studies concerning earnings announcements, see Patell and Wolfson (1984) for changes in mean and variance of intra-day prices, and Lee, Mucklow, and Ready (1993) and Krinsky and Lee (1996) for changes in the bid-ask spread. For additional studies concerning news announcements, see Charest (1978), Morse (1981), and Kalay and Loewentstein (1985).
20. Model Incorporation of Scheduled Events: Almgren and Chriss (2000) work in a simple extension of their static framework by assuming that the security again follows an arithmetic random walk, but at a time known at the beginning of trading, an uncorrelated event will cause a material shift in price dynamics, e.g., an increase or decrease of volatility.
21. Combining Piece-Wise Static Strategies: In this context they show that optimal strategies are piece-wise static. To be precise, they show that an optimal strategy entails following a static strategy up to the moment of the event, followed by another static strategy that can only be determined once the outcome of the event in known.
22. Variation from the Original Static Strategy: It is interesting to note that the static strategy that one follows in the first leg is in general not the same strategy one would follow in the absence of information concerning the event.
23. Accommodating Unanticipated External “Sudden’ Events: Finally Almgren and Chriss (2000) note that any optimal execution strategy is vulnerable to *unanticipated events*. If such an event occurs during the course of trading and causes a material shift in the parameters of the price dynamics, then indeed a shift in the optimal trading trajectory must also occur.
24. Adaptation at Parameter Shift Edges: However if one makes a simplifying assumption that all events are either “scheduled” or “unanticipated” one then concludes that optimal execution is always a game of static trading punctuated by shifts in the trading strategies that adapt to material changes in the price dynamics.
25. Pre-determined vs. Active Approaches: If shifts are caused by events that are known ahead of time, then optimal execution benefits from a precise knowledge of the possible outcomes of the event. If not, the best approach is to be actively “watching” the market for such changes and react swiftly should they occur.
26. Simple Proxy for Unexpected Uncertainty: One approximate way to include such completely unexpected uncertainty into the model is to artificially raise the value of the volatility parameter.
27. Risk Averse Optimal Trading Strategies: As a first step, Almgren and Chriss (2000) obtain closed form solutions for trading strategies for any level of risk aversion.
28. Efficient Frontier of Optimal Strategies: They then show that this leads to an efficient frontier of optimal strategies, where an element of the frontier is represented by a strategy with a minimal level of cost for its level of variance of the cost.
29. Graphical Structure of the Frontier: The structure of the frontier is of some interest. It is a smooth convex function differentiable at its minimal point. The minimal point is what Bertsimas and Lo (1998) call the naïve strategy because is corresponds to trading equally sized packets using all available trading time equally.
30. Differential at the Minimum Point: The differentiability of the frontier at its minimum point indicates that one can obtain a first order reduction in the variance of the trading cost at the expense of only a second order in cost by trading a strategy slightly away from the globally minimal strategy.
31. Curvature at the Minimal Point: The curvature of the frontier at its minimum point is a measure of the liquidity of the security.
32. Half-Life of Optimal Execution: Another ramification of the Almgren and Chriss (2000) study is that for all levels of risk aversion except risk neutrality, optimal execution trades have a “half-life” that fall out of the calculations.
33. Independence form the Time to Complete Execution: A trade’s half-life is independent of the actual specified time to liquidation, and is a function of the security’s liquidity and volatility, and the trader’s level of risk aversion.
34. Half-Life as Execution Time: As such Almgren and Chriss (2000) regard the half-life as an idealized time to execution, and perhaps a guide to the proper amount of time over which to execute a transaction.
35. Time Lesser than Half Life: If the specified time to liquidation is short relative to the trade’s half-life, once can expect the cost of trading to be dominated by transaction costs.
36. Time Greater than Half Life: If the time to trade is long relative to the half-life, one can then expect most of the liquidation to take place well in advance of the limiting time.

**The Definition of a Trading Strategy**

1. Price Dynamics and Trade Execution: As a starting point, Almgren and Chriss (2000) define s trading strategy, and lay out the dynamics that they study. They start with a formal definition of a strategy for a sell program consisting of liquidating a single security. The definitions and results are analogous for a buy program.
2. Problem Setup - Security Liquidation: Suppose that the seller holds a block of units of a security that they want to completely liquidate before time . To keep the discussion, Almgren and Chriss (2000) speak of *units* of a security. Specifically they have in mind shares of stock, futures contract, and units of a foreign currency.
3. Trading Strategy - Price/Unit Strategy: The seller divides into units of length

and defines the discrete times

for

The *trading trajectory* is defined to be the list where is the number of units that the seller plans to hold at time .

1. Outright/Re-balanced Trajectories: The initial holdings is

and liquidation at time requires

A trading trajectory can thought of as either the ex-post realized trades resulting from some process, or as a plan concerning how to trade a block of securities. In either case one may also consider *re-balancing* trajectories by requiring

the initial position, and

the new position, but this is formally equivalent to studying trajectories of the form

and

1. Outstanding Holdings/Incremental Trade Lists: Equivalently, a strategy may be specified using the “trade list” where

is the number of units that the seller will sell between times and . Clearly, and are related by

1. Simultaneous Portfolio Buying and Selling: Almgren and Chriss (2000) also consider more general programs of buying and selling simultaneously several securities.
2. Inter-Execution Time Interval Specification: For notational simplicity they consider all the time intervals to be of equal length , but this restriction is not essential.
3. Behavior at Limits: Although they do not discuss it, in all their results it is easy to take the continuous-time limit of

and

1. Definition of a Trading Strategy: Almgren and Chriss (2000) define a “trading strategy” to be a rule for determining in terms of the information available at . Broadly speaking they distinguish between two types of trading strategies – static and dynamic.
2. Static vs. Dynamic Trading Strategy: Static strategies are determined in advance of trading, that is the rule for determining each depends only on information available at . Dynamic strategies, conversely, depend on all information up to, and including, time

**Price Dynamics**

1. Exogenous/Endogenous Price Move Factors: Suppose that the initial security price is so that the initial market value of the position is . The securities’ price evolves according to two exogenous factors – volatility and drift, and one endogenous factor – market impact.
2. Market Forces vs. Trading Impact: Volatility and drift are assumed to the result of market forces that occur randomly and independent of the trading.
3. Earlier Literature on Market Impact: Almgren and Chriss (2000) discussion s largely reflect the work of Kraus and Stoll (1972), and the subsequent works of Holthausen, Leftwich, and Mayers (1987, 1990) and Chan and Lakonishok (1993, 1995). See also Keim and Madhavan (1995, 1997).
4. Origin of the Market Impact: As the market participants begin to detect the volume that the seller (buyer) is selling (buying), they naturally adjust their bids (offers) downward (upward). Almgren and Chriss (2000) distinguish two kinds of market impact.
5. Definition of Temporary Market Impact: *Temporary* impact refers to the temporary imbalances in supply and demand caused by the seller’s trading leading to temporary price movements away from equilibrium.
6. Definition of Permanent Market Impact: *Permanent* impact refers to the changes in the “equilibrium” price due to the seller’s trading, which remain at least for the life of the liquidation.
7. Price Evolution Stochastic Difference Equation: Almgren and Chriss (2000) assume that the security price evolves according to the discrete random walk

for

1. Glossary of the Equation Terms: Here represents the volatility of the asset, ’s are draws from independent random variables each with zero mean and unit variance, and the permanent impact function is a function of the *average rate* of trading

during the interval to .

1. Lack of Explicit Drift Term: In the above equation there is no drift term. Almgren and Chriss (2000) indicate that this is due to the assumption that they have no information about the direction of the future price movements.
2. Trading Term Horizons under Consideration: Over long term investment time scales, or in extremely volatile markets, it is important to consider *geometric* rather than arithmetic Brownian motion – this corresponds to letting in

scale with . But over short term “trading” horizons of interest, the total fractional price changes are small, and the differences between arithmetic and geometric Brownian motions are negligible.

**Temporary Market Impact**

1. Intuition behind the Temporary Market Impact: The intuition behind the temporary market impact is that a trader plans to sell a certain number of units between times and , but may work the order in several smaller sizes to locate optimal points of liquidity.
2. Liquidity Reduction Impact on Price: If the total number of units is sufficiently large, the execution price may steadily decrease between and in part due to the exhaustion of the supply of liquidity at each successive price level. This effect is assumed to be short-lived, and in particular, liquidity is assumed to return back after each period, and a new equilibrium price is established.
3. The Temporary Price Impact Function: This effect is modeled by introducing a temporary price impact function , the temporary drop in the average price per share caused by trading at an average rate during one time interval.
4. Net Price Received at Execution: Give this, the actual price per share received on sale is

but the effect of does not appear in the next “market” price .

1. Choice of Market Microstructure: The functions in

and in

may be chosen to reflect any preferred model of market microstructure, subject only to certain natural convexity conditions.

**Capture and Cost of Trading Trajectories**

1. Capture across a Trading Trajectory: Almgren and Chriss (2000) then discuss the profits resulting from trading along a certain trajectory. They define the *capture* of a trajectory to be the full trading revenue upon completion of all trades. Due to the short term horizons that they consider, they do not include any notion of carry or time value of money in their discussions.
2. Full Trading Revenue across Execution: Thus, the capture is the sum of the product of the number of units sold in each time interval times the effective price per share received on that sale. It is readily computed as
3. Decomposition of the Capture Components: The first term on the RHS above is the initial market value of the position; each additional term represents a gain or a loss due to a specific market factor.
4. The Volatility Price Impact Term: The first term represents the total impact from the volatility.
5. The Permanent Market Impact Term: The permanent market impact term represents the loss in the value of the position caused by a permanent price drop associated with selling a small piece of the position.
6. The Temporary Market Impact Term: And the temporary market impact term is the price drop due to selling, acting only on the units sold during the period.
7. The Total Cost of Trading: The *total cost of trading* is the difference between the initial book value and the capture. This is the standard *ex-post* measure of the performance costs used in performance evaluations, and is essentially what Perold (1988) calls *implementation shortfall*.
8. Estimation of Implementation Short-fall: In this model, prior to trading, the implementation short-fall is a random variable. Write for the expected short-fall and for the variance of the short-fall.
9. Implementation Short-fall Mean/Variance: Given the simple nature of price dynamics, Almgren and Chriss (2000) readily compute

The units of are in dollars, and the units of are dollars squared.

1. Distribution of Implementation Short-fall: The distribution of the short-fall is Gaussian if is Gaussian, in any case if is large, it is very nearly Gaussian.
2. Almgren and Chriss Minimizer Utility: Almgren and Chriss (2000) devote much of their paper to finding trajectories that minimize for various values of . They demonstrate that for each value of there corresponds a unique trading trajectory such that is minimal.

**Linear Impact Functions**

1. Linear Temporary/Permanent Market Impact: Although Almgren and Chriss (2000) formulation does not require it, computing optimal trajectories is significantly easier if one takes the permanent and temporary impact functions to be *linear* in the rate of trading.
2. Linear Permanent Impact Market Function: For linear permanent impact, has the form

in which the constant has units of ($/share)/share.

1. Corresponding Execution Time Security Price: With this form, each units sold depresses the price per share by regardless of the time taken to sell units.

readily yields

1. Permanent Implementation Short-fall Mean: Then summing by parts, the permanent impact term in

becomes

1. Linear Temporary Impact Market Function: Similarly, for the temporary impact we take

where is the sign function.

1. Estimating the Fixed Costs of Execution: The units of are $/share, and those of are ($/share)/(share/time). A reasonable estimate for is the fixed cost of selling, such as half of bid-ask spread plus premium.
2. Estimating the Linear Impact Coefficient: It is more difficult to estimate since it depends on the internal and the transient aspects of the market microstructure. It is in this term that one would expect the on-linear terms to be most important, and the approximation

to be most doubtful.

1. Total Temporary Impact Function: The linear model

is often called a *quadratic* cost because the total costs incurred by buying or selling units in a single unit of time is

1. Temporary Implementation Short-fall Mean: With both linear cost models

and

the expectation of the impact costs

becomes

in which

1. Strictly Convex Nature of : Clearly is a strictly convex function as long as

Note that if all have the same sign, as would be the case for a pure sell program or a pure buy program, then

1. and Computation Illustration: To illustrate, Almgren and Chriss (2000) compute and for linear impact functions for two of trajectory schemes at the opposite extremes: sell at a constant rate, and sell to maximize variance without regard to transaction costs.
2. Minimum Impact: Constant Execution Rate: The most obvious trajectory is to sell at a constant rate over the entire liquidation period. Thus one takes each

and

1. Minimum Impact and : From

and

one has

and from

1. Minimum Impact Limits: The trajectory minimizes total expected costs, but the variance may be large if the period is long. As the number of trading periods

remains finite, and and have finite limits.

1. Minimum Variance: One Step Execution: The other extreme is to execute the entire position in the first time step. One then takes

which results in

and

1. Minimum Variance Limits: The trajectory has the smallest possible variance –equal to zero – because of the way time has been discretized in the model above. If is large and hence is short, then on the full initial portfolio, one takes a price hit that can be arbitrarily large.
2. Trajectory between the Two Extremes: Almgren and Chriss (2000) show how to effectively compute trajectories that lie between the two extremes.

**The Efficient Frontier of Optimal Execution**

1. Computing the Optimal Execution Trajectories: Almgren and Chriss (2000) define and compute optimal execution trajectories and use that to later demonstrate a precise relationship between risk aversion and the definition of optimality.
2. Uniqueness of Optimal Execution Strategy: In particular, they show that each level of risk aversion there is a uniquely determined optimal execution strategy.

**The Definition of the Frontier**

1. Minimization of Expected Short-fall: The rational trader will always seek to minimize the expectation of short-fall for a given level of variance of the short-fall. Naturally a trader will prefer a strategy that provides minimum error in its estimate of expected costs.
2. Efficient Optimal Trading Strategy Definition: Thus a strategy is *efficient* or *optimal* if there is no other strategy that has lower variance for the same or a lower variance of the expected transaction costs, or, equivalently, no strategy which has no lower expected transaction costs for the same or lower level of variance.
3. Static vs. Dynamic Strategy Optimality: This definition of optimality of a strategy is the same whether the strategy is static or dynamic. It will be established later that under this definition and the price dynamics already stated, optimal strategies are in fact static.
4. Efficient Strategies - Constrained Optimization Formulation: One may construct efficient strategies by solving the constrained optimization problem

That is, for a given maximum level of variance

one finds a strategy that has the minimum expected levels of transaction costs.

1. Convex Objective Function and Domain: Since is convex, the set

is convex – it is a sphere – and since is strictly convex, there is a unique minimizer .

1. Sub-Optimal Trajectory Variance Cost: Regardless of the preferred balance of risk and return, every other solution which has

has higher expected costs than for the same or lower variance, and can never be more efficient.

1. Efficient Frontier of Optimal Strategies: Thus the family of all possible efficient (optimal) strategies is parametrized by a single variable representing all possible maximum levels of variance in transaction costs. This family is referred to as *the efficient frontier of optimal trading strategies*.
2. Introducing KKT Type Constraint Multipliers: The constrained optimization problem

is solved by introducing a constraint multiplier , thereby solving the unconstrained problem

1. Frontier as a Function of : If

is strictly convex, and the above minimizer has a unique solution . As varies, sweeps out the same one parameter family, and thus traces out an efficient frontier.

1. as a Risk Aversion Parameter: The Parameter has a direct financial interpretation. It is already apparent from

that is a measure of risk aversion, that is, how much the variance is penalized relative to the cost.

1. as an Efficient Frontier Curvature: In fact, is the curvature – second derivative – of a smooth utility function, as will be made more precise eventually.
2. Solution given and : For given values of the parameters, problem

can be solved by various numerical techniques depending on the functional forms chosen for and . In the special case that these are *linear* functions, we may write the solution explicitly and gain a great deal of insight into the trading strategies.

**Explicit Construction of Optimal Strategies**

1. Optimal Solution in Trajectory Space: With from

and from

and assuming that does not change sign, the combination

is a quadratic function of the control parameters ; it is strictly convex for

1. Finding the Unique Global Minima: Therefore one determines the unique global minimum by setting its partial derivatives to zero. One readily calculates

for

1. Combinations of Linear Difference Equations: Then

is equivalent to

with

1. Abstracted and Re-factored Parameter Set: Note that

is a linear difference equation whose solution may be written as a combination of the exponentials where satisfies

The tilde’s on and denote an correction; as

one has

and

1. Trading Trajectory/Trade List Solutions: The specific solution with

and

is a trading trajectory of the form

and the associated trade list is

where and are the hyperbolic sine and cosine functions, and

These solutions – although not the efficient frontier – have been constructed previously by Grinold and Kahn (1999).

1. Monotonicity of the Trading Trajectory: One has

as long as

Thus for a program of selling a large initial long position, the solution decreases *monotonically* from its initial value to zero at the rate determined by the parameter .

1. Consequence of Monotonic Trading Trajectories: For example, the optimal execution of a sell program never involves buying of securities – although this ceases to be true if there is drift or serial correlation in price movements.
2. Approximation under Small Time Step: For a small time step one has the approximate expression

Thus if the trading intervals are short is essentially the ratio of the product of volatility and the risk-intolerance to the temporary transaction cost parameter.

1. Optimal Strategy Expected Cost/Variance: The expectation and the variance of the optimal strategy for a given initial portfolio size are then

and

which reduce to

in the limits

**The Half-Life of a Trade**

1. Definition of the Half-Life: Defining

the trade’s “half-life”, and using the discussion above, it can be seen that the larger the value of and smaller the value of , the more rapidly the trade list will be depleted. The value is exactly the amount of time it takes to deplete the holdings by a factor of .

1. Half-Life Different from : The definition of is independent of the exogenously specified execution time it is determined only by the security price dynamics and the market impact factors. If the risk aversion is greater than zero, i.e., if the trader is risk-averse, then is finite and independent of .
2. Timeless Initial Portfolio Liquidation Rate: Thus, in the absence of any external time constraint, i.e.

the trader will still liquidate his position on a time scale . The half-life is the intrinsic time scale of the trade.

1. Half Life Smaller than : For a given the ratio

tells us what factors constrain the trade. If

then the intrinsic half-life of the trade is small compared to the imposed time ; this happens because temporary costs are very small, because volatility is very large, or because of high risk aversion.

1. Impact of Small Half-Life: In this case the bulk of the trading will be done well in advance of the time . Viewed on a time scale the trajectory will look like a minimum variance solution
2. Very High Half Life Limit: Conversely if

then the trade is highly constrained, and is dominated by temporary market impact costs. In the limit

one approaches the straight line minimum cost strategy

1. Trade Size Independent Execution Strategy: A consequence of this analysis is that different sized baskets of the same security will be liquidated in exactly the same fashion, on the same scale, provided the risk aversion parameter is held constant.
2. Basket Size Based Liquidity Dependence: This may seem contrary to the expectation that large baskets are effectively less liquid, and should hence be liquidated less rapidly than smaller baskets.
3. Reasons for the Counter-Intuitiveness: This is a consequence of the linear market impact assumption which has the *mathematical* consequence that both variance and market impact scale quadratically with respect to the portfolio size.
4. Higher Order Temporary Impact Function: For large portfolios it may be more reasonable to assume that the temporary impact cost function has higher-order terms, so that such costs increase *super-linearly* with the trade size. With non-linear impact functions, the general framework used here still applies, but one does not obtain explicit exponential solutions as in the linear impact case.
5. Size Dependent Temporary Impact Parameter: A simple practical solution to this problem is to choose different values of - the temporary impact parameter – depending up on the overall problem size being considered, recognizing that the model is at best only approximate.

**Structure of the Frontier**

1. Efficient Frontier and the Corresponding Trajectories: Using a specific choice for the parameters explained below, Almgren and Chriss (2000) produce a sample plot of the efficient frontier – each point on the frontier represents a distinct strategy for optimally liquidating the same basket. Their tangent line represents the optimal solution for a specified risk parameter

They also illustrate the trajectories corresponding to a few sample points on the frontier.

1. Trajectory corresponding to Positive : Their first trajectory has

– this would be chosen by a risk-averse trader who wishes to sell quickly to reduce exposure to volatility risk, despite the trading costs incurred in doing so.

1. Trajectory corresponding to Zero : Their second trajectory has

They refer to this as the naïve strategy since this represents an optimal strategy corresponding to simply minimizing expected transaction costs without regard to variance.

1. Linear Reduction of the Holdings: For a security with zero drift and linear transaction costs as defined above

corresponds to a simple linear reduction of holdings over the trading period. Since drift is generally not significant over short trading horizons, the naïve strategy is very close to the linear strategy.

1. Sub Optimality of the Strategy: As Almgren and Chriss (2000) demonstrate later, in a certain sense this is *never* an optimal strategy because one can obtain substantial reductions in variance for a relatively small increase in transaction costs.
2. Trajectory corresponding to Negative : Finally their trajectory has

it would only be chosen by a trader who likes risk. He postpones execution, thus incurring higher costs both due to rapid sales at the end, and higher variance during the extended period that he holds the security for.

**The Utility Function**

1. The Risk-Reward Trade-off: Almgren and Chriss (2000) offer an interpretation of the efficient frontier of optimal strategies in terms of the utility function of the seller. They do this in two ways – by direct analogy with modern portfolio theory employing a utility function, and by a novel approach: Value-at-risk. This eventually leads to some general observations regarding the importance of utility in forming execution strategies.
2. Utility of Risk-Averse Functions: Suppose on measure utility by a smooth convex function where is the total wealth. This function may be characterized by its risk-aversion coefficient
3. Approximation in Estimating the : If the initial portfolio is fully owned, then as the transfer of assets happens from the risky stock into the alternative riskless investment, remains roughly constant, and one may take to be a constant throughout the trading period. If the initial portfolio is highly leveraged, then the assumption of constant is an approximate one.
4. Formulation of the Optimal Execution Strategy: For short time horizons and small changes in the higher derivatives of may be neglected. Thus choosing an optimal execution strategy is equivalent to minimizing the scalar function

The units of are ; one is willing to accept an extra square of variance if it reduces the expected cost by .

1. Constructing Family of Optimal Paths: The combination is precisely the one used to construct the efficient frontier seen earlier; the parameter , introduced as a Lagrange multiplier, has a precise definition as a measure of aversion to risk. Thus, the methodology above used to construct the efficient frontier likewise produces a family of optimal paths, one for each level of risk aversion.
2. Static Nature of Optimal Path: Returning now to an important point raised earlier, the computation of optimal strategies by minimizing as measured at the initial trading time is equivalent to maximizing the utility at the outset of trading. As one trades, information arrives that could potentially alter the optimal path. The following theorem eliminates that possibility.
3. Time Homogenous Quadratic Utility Theorem: For a fixed quadratic utility function, the static strategies computed above are “time homogenous”. More precisely given a strategy that begins at a time

and ends at a time

the optimal strategy computed at

is simply a continuation from

to

of the optimal strategy computed at time

1. Proof Steps: General/Specific Functions: The proof may be seen in two ways – by the algebraic computations based on the specific solutions above, and by general valid for generic non-linear impact functions.
2. Proof Steps: Function Time Shift: First suppose that at time , where

one were to compute a new optimal strategy. The new strategy would precisely be

with replaced by , replaced by , and replaced by . Using the subscript to denote the strategy computed at time one would have

and the trade lists

1. Proof Step: Recovering Optimal Solutions: It is then apparent that if is the optimal solution from

with

then

and

where

and

are the strategies from

and

1. Proof Step: Non-linear Impact: For general non-linear impact functions and the optimality condition

is replaced by a second-order *non-linear* difference relation. The solution beginning at a given time is determined by the two boundary values and

It is then apparent that the solution does not change if we re-evaluate it at later times.

1. Origin of Time Stable Solutions: More fundamentally, the solutions are time stable because in the absence of serial correlations in the asset price movements, there is no more information about the price changes at later times than there is at the initial time.
2. Optimality over each Sub-interval: Thus, the solution which was determined to be optimal over the entire time interval is optimal as a solution over each sub-interval. This general phenomenon is well known in the theory of optimal control (Bertsekas (1976)).

**Value at Risk**

1. Motivation behind Value at Risk: The concept of value at risk is traditionally used to measure the greatest amount of money – maximum profit or loss - a portfolio will sustain over a given period of time under “normal circumstances”, where “normal” is defined by a confidence level.
2. Trading Value at Risk Definition: Given a trading strategy

the value-at-risk of defined is defined to be the level of transaction costs by the trading strategy that will not be exceeded percent of the time. Put another way, it is the percentile level of transaction costs for total costs of trading .

1. Trading Value at Risk Expression: Under the arithmetic Brownian motion assumption, the total costs – the market value minus capture – are normally distributed with known mean and variance. Thus the confidence level is determined by the number of standard deviations from the mean of the inverse of the cumulative normal distribution function, and the value at risk for the strategy is given by
2. Relation to Implementation Short-fall: That is, with a probability the trading strategy will not lose more than of its market value in trading. Borrowing from the language of Period (1988), the implementation shortfall of execution will not exceed more than a fraction of the time. A strategy is efficient if it has the minimum possible value at risk for the confidence level .
3. Execution Trajectory Optimized for VaR: Note that is a complicated non-linear function of composing ; it can be easily evaluated for any given trajectory, but finding the minimizing trajectory directly is difficult.
4. Single Parameter Efficient Frontier Solution: But once the one-parameter family of solutions that form the efficient frontier is obtained, one only needs to solve a one-dimensional problem to find the optimal solutions for the value at risk model, that is, to fund the value of corresponding to a given value of . Alternatively on may characterize the solutions by a simple graphical procedure, or may read off the confidence levels corresponding to any particular point on the curve.
5. Almgren-Chriss Optimal VaR Illustration: Almgren and Chriss (2000) produce an illustration of the above, using the square root of variance in the -axis as opposed to the variance in itself. In this co-ordinate system lines of optimal VaR have a constant slope, and for a given value of they simply find a tangent to the curve where the slope is .
6. Interim Optimal Execution Re-evaluation: The question of re-evaluation of the strategy is more complicated and subtle. If one re-evaluates the strategy half-way through the execution process, they will choose a new optimal strategy that is not the same as the original optimal one. The reason is that since is now held constant, necessarily changes.
7. General Challenges with the VaR Approach: Value at risk has many flaws from a mathematical point of view, as recognized by Artzner, Delbaen, Eber, and heath (1997). The particular issue encountered her would occur in any problem in which the time of measurement is a fixed date, rather than maintained at a fixed distance in the future. It is an open issue to formulate suitable measures of risk for general time-dependent problems.
8. Liquidity Adjusted Value at Risk: Despite this shortcoming, Almgren and Chriss (2000) the smallest possible value of as an informative measure of the possible loss associated with the initial position, in the presence of liquidity effects. This value, which they call L-VaR for Liquidity Adjusted Value at Risk, depends on the time to liquidation and the confidence level chosen, in addition to the market parameters such as the impact coefficient (Almgren and Chriss (1999)).
9. Advantages of the L-VaR Approach: The optimal trajectories determined by minimizing the value at risk do *not* have the counter-intuitive scaling behavior seen earlier; even for linear impact functions, large portfolios will be traded closer to the straight line trajectory.
10. Using L-VaR for Large Portfolios: This is because the cost assigned to uncertainty scales *linearly* with the portfolio size, while the temporary impact cost scales *quadratically* as before. Thus the latter is more important for large portfolios.

**The Role of Utility in Execution**

1. General Observations on Optimal Execution: Almgren and Chriss (2000) use the structure of the efficient frontier in the framework that they have developed to make some general observations concerning optimal executions.
2. The Naïve Strategy Benchmark: They first restrict themselves to the situation where the trader has no directional view on the security being traded. Recall that in this case, the naïve strategy is the simple straight line strategy in which the trader breaks the blocks being executed into equal sized blocks to be sold over equal time intervals. They use this strategy as a benchmark for comparison with the other strategies used throughout here.
3. Convex to Mapping: A crucial insight is that the curve defining the efficient frontier is a smooth convex function mapping the levels of variance to the corresponding minimum mean transaction cost levels.
4. Region around the Naïve Strategy: Write for the mean and variance around the naïve strategy. Regarding as a point on the smooth curve defined by the frontier, evaluated at is equal to zero. Thus for near one has

where

is positive is positive by the convexity of the frontier at the naïve strategy.

1. Special Feature of the Naïve Strategy: By definition, the naïve strategy has the property that any strategy with lower variance in cost has a greater expected cost. However a special feature of the naïve strategy is that a first-order decrease in variance can be obtained – in the sense of finding a strategy with a lower variance – while only incurring a second order increase in cost.
2. Disadvantages of Risk Neutral Strategy: From the above it follows that for small increases in variance, one can obtain much larger reductions in cost. Thus unless the trader is risk-neutral it is always advantageous to execute a strategy that is at least to some degree “to the left” of the naïve strategy. Thus one concludes that, in this framework, from a theoretical standpoint, it never makes sense to trade a strictly risk-neutral strategy.
3. The Role of a Security’s Liquidity: An intuitive proposition is that with all things being equal, a trader will execute a more liquid basket more rapidly than a less liquid one. In the extreme this is particularly clear. A broker given a small order to execute over the course of the day will execute the entire order almost immediately.
4. Executing the Highly Liquid Security: How does one explain this? The answer is that the market impact cost attributable to rapid trading is negligible compared with the opportunity cost incurred in breaking up the order over an entire day. Thus, even if the expected return on a security over the day is zero, the perception is that the risk of waiting is overweighed by any small cost of immediacy.
5. Absence of Risk Reduction Premium: Now if the trader were truly risk neutral, in the absence of any views, he would always use the naïve strategy and employ the allotted time fully. This would make sense because any price to pay for trading immediately is worthless if one places no premium on risk reduction.
6. Limitation of Risk Neutral Approach: It follows that any model that proposes optimal trading behavior should predict that more liquid baskets are traded more rapidly than less liquid ones. A model that only considers the minimization of transaction costs, like that of Bertsimas and Lo (1998), is essentially a model that excludes utility.
7. Optimal Execution Independent of Liquidity: In such a model, and under Almgren and Chriss (2000) basic assumptions, traders will trade all baskets at the same rate irrespective of the liquidity, that is unless they have an explicit directional view on the security, or the security possesses extreme serial correlation in its price movements.
8. Super Linear Market Impact Functions: Almgren and Chriss (2000) do note that their model in the case of linear transaction costs does not predict a more rapid trading for smaller versus larger baskets of the same security. However, this is a consequence of choosing linear temporary impact functions and the problem goes away when one considers more realistic super-linear functions.
9. Risk Neutral Execution Half Life: Another way of looking at this is that the half-life of all black executions, under the assumption of risk-neutral preferences, is infinite.

**Choice of Parameters**

1. The Asset Intrinsic Dynamics Parameters: Almgren and Chriss (2000) compute some numerical examples for the purposes of exploring the qualitative properties of the efficient frontier. Throughout the examples they consider a single stock with the current market price of

and that they initially have one million shares, for an initial portfolio size of million. The stock will have annual volatility, expected annual rate of return, a bid-ask spread of , and a median daily trading volume of million shares.

1. Stock Asset Daily Return/Volatility: With a trading year of days this gives a daily volatility of

and expected fractional return of

To obtain our absolute parameters and one must scale it by the price, so

and

The table below summarizes the information.

1. Parameter Values for the Test Case:

|  |  |  |
| --- | --- | --- |
| **Parameter Description** | **Parameter Symbol** | **Parameter Value** |
| Initial Stock Price |  | /share |
| Initial Holdings |  | shares |
| Liquidation Time |  | days |
| Number of Time Periods |  |  |
| Annual Volatility |  |  |
| Annual Growth |  |  |
| Bid Ask Spread |  |  |
| Daily Volume million shares |  |  |
| Impact at of market |  |  |
| Static Holdings shares |  |  |
| VaR Confidence |  |  |

1. Incremental and Total Execution Times: Suppose that one wants to liquidate this position in one week so that

This is divided into daily trades such that is and

1. Standard Deviation of the Trajectory: Over this period, if one holds the original position with no trading, the fluctuations in the stock value will be Gaussian with a standard deviation of

and the fluctuations in this value will have an absolute standard deviation of

As expected this is precisely the value of for the lowest point in the efficient frontier, since that point corresponds selling along a linear trajectory rather than holding a constant amount.

1. Temporary Cost Function Parameter - : One then chooses the parameters for the temporary cost function

Almgren and Chriss (2000) set

that is, the fixed part of the temporary costs will be one-half the bid-ask spread.

1. Temporary Cost Function Parameter - : For they suppose that for each of the daily volume traded they incur a price impact equal to one bid-ask spread. For example trading at a rate of daily volume incurs a one-time cost on each trade of . Under this assumption
2. Permanent Cost Function Parameter - : For permanent costs, the common rule of thumb is that price effects become significant when of the daily volume is sold. Assuming that “significant” means that the price depression is one bid-ask spread, and that the effect is linear for both smaller and larger trading rates, one has

Recall that this parameter gives a fixed cost independent of the path.

1. The Risk Aversion Parameter - : Almgren and Chriss (2000) have chosen

For these parameters, from

one has for the optimal strategy that

so that

Since this value is near in magnitude, the behavior is an interesting intermediate in-between the naïve extremes.

1. at Confidence Level: For the value at risk representation, as assumed confidence level gives

**The Value of Information**

1. Zero Drift Random Walk Assumption: The discussion carried out so far assumed that the price dynamics followed an arithmetic random walk with zero drift. Since past price paths provide no extra information on future price movements, the conclusion was that the optimal trajectories can be statically determined. There are three ways by which a random walk with zero drift may fail to represent the price process.
2. Non-zero Drift in Dynamics: First the price process may have drift. For example, if the trader has a strong directional view, the trader may want to incorporate this view into the liquidation strategy.
3. Cross Period Serial Correlation Impact: Second, the price process may exhibit serial correlation. The presence of first order serial correlation for example, implies that the price moves in a given period provide non-trivial information concerning the next period movement of the asset.
4. Incorporation of the Investor’s Private Information: Bertsimas and Lo (1998) study a general form of this assumption, wherein an investor possesses possibly private information of a serially correlated information vector that acts as a linear factor in the asset returns.
5. Exogenously Induced Material Parameter Shift: Lastly, at the start of trading, it may be known that at some specific point in time, an event will take place whose outcome will cause a material shift in the parameters governing the price process.
6. Literature Survey on Exogenous Events: Such event induced parameter shifts include quarterly and annual earnings announcements, dividend announcements, and share re-purchases. Event studies documenting these parameter shifts and providing theoretical grounding for their existence include Beaver (1968), Fama, Fisher, Jensen, and Roll (1969), Dann (1981), Patell, and Wolfson (1984), Kalay and Loewentstein (1985), Kim and Verrecchia (1991), Campbell, Lo, and MacKinlay (1997), Easterwood and Nutt (1999), and Ramaswami (1999).
7. Temporary Shifts on Dynamic Parameters: For example, Brown, Harlow, and Tinic (1988) show that events cause temporary shifts in both the risk and returns of individual securities, and the extent of these shifts depends on the outcome of the event. In general, securities react more strongly to bad news than good news.
8. Probabilistic Event Outcomes/Parameter Shifts: Almgren and Chriss (2000) study a stylized version of the events in which a known event at a known time – e.g., an earnings announcement – has several possible outcomes. The probability of each outcome is known, and the impact that a given outcome will have on the parameters of the price is also known. Clearly, optimal strategies must explicitly use this information, and Almgren and Chriss (2000) develop methods to incorporate event-specific information into their risk-reward framework.
9. Back-to-Back Static Strategies: The upshot is a piece-wise strategy that trades statically up to the event, and then reacts explicitly to the outcome of the event. Thus the burden is on the trader to determine which of the possible outcomes occurred and then trade accordingly.

**Drift**

1. Drift as a Directional View: It is convenient to regard the drift parameter in the price process as a directional view of price movements. For example, the trader charged with liquidating a single security may believe that this security is likely to rise. Intuitively it makes more sense to trade this issue more slowly to take advantage of this view.
2. Incorporating Drift into Price Dynamics: To incorporate drift into the price dynamics Almgren and Chriss (2000) modify

to

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

1. Price Expectation over Time Period: One can readily write the modified version of

as

1. Updated Objective Function Optimality Condition: The variance is still given by

The optimality condition

becomes

in which the new parameter

is the optimal level of security holding for a time independent portfolio optimization problem.

1. Drift Based Updated Execution Slice: For example, the parameters used in the example above give approximately

or of our initial portfolio. One expects this fraction to be very small, since, by hypothesis, the eventual aim is complete liquidation.

1. Drift Based Updated Optimal Solution: The optimal solution

becomes

for

with the associated trades

1. Initial Position Independent Trajectory Correction: This trading trajectory is a sum of two distinct trajectories – the zero-drift solution as computed before, plus a “correction” which profits by capturing a piece of the predictable drift component. The size of this correction term is proportional to , and thus to : it is independent of the initial portfolio size .
2. Practical Incorporation into Program Trading: To place this in an institutional framework, consider a program trading desk that sits in front of customer flow. If this desk were to explicitly generate alphas on all securities that flow through the desk in an attempt to, say, hold securities with high alphas and sell securities with low alphas more rapidly, the profit would not scale in proportion to the average size of the programs. Rather it would only scale with the number of securities that flow through the desk. An even stronger conclusion is that since the optimal strategy disconnects into a static strategy unrelated to the drift term, and a second strategy related to the drift term, there is no particular advantage to restricting trading in securities which the desk currently holds the positions in.
3. Comparison: Highly Liquid Markets Scenario: The difference between this solution and the no-drift solution in

may be understood by considering the case

corresponding to highly liquid markets. Whereas the previous one relaxed from to in a time scale of

this one relaxes instead to the optimal static portfolio size . Near the end of the trading period the trader sells the remaining holdings to achieve

at

1. Caveat: Buy-Sell Symmetry Breaking: In this case, one requires

in order for all trades to be in the same direction. This breaks the symmetry between a buy program and a sell program, if one wanted to consider buy programs it would be more logical to set

**Gain due to Drift**

1. Gain from Drift – Calculation Motivation: Now suppose that the price dynamics is given by

with

but one chooses to determine the solution as though

The situation may arise, for example, in case where the trader is trading a security with non-zero drift, but *unknowingly* assumes that the security has no drift. Almgren and Chriss (2000) explicitly calculate the loss associated with ignoring the drift term.

1. Gain adjusted and : Write for the optimal solution

with

for the sub-optimal solution

or

with

Also write and for the optimal expected cost and its variance measured by

and

with

and write and for the sub-optimal values of

and

evaluated with

1. Objective Function Gain from Drift: The corresponding objective functions are

and

One can then define the *gain due to drift* to be the difference ; this is the reduction in the cost and the variance by being aware of and taking into account of the drift term. Clearly

since is the unique optimal strategy for the model with

1. Upper Bound for the Gain: Now the value of the terms in that come from

and

is only *increased* by going from to since and not was the optimum strategy with

Therefore an *upper bound* for the gain is

1. Adjustment Applied to the Holdings: That is, in response to positive drift, one should increase the holdings throughout the trading. This reduces the net cost by the amount of the increase in the asset price one captures, at the expense of slightly increasing the transaction costs and the volatility exposure. An upper bound for the possible benefit is the amount of increase one captures.
2. Explicit Expression for the Bound: But is just the term in the square brackets in

times , which is clearly independent of . Indeed this can be explicitly evaluated to get

1. Gain Comparison against Execution Cost: Since is a positive decreasingly function, this quantity is positive and bounded above by , the amount one would gain by holding for a time . Any reasonable estimates for the parameters show that this quantity is negligible compared to the impact costs incurred in liquidating an institutional sized portfolio over a short period.

**Serial Correlation**

1. Prior Period Price Increment Component: Now one supposes that the asset prices exhibit serial correlation, so that at each period one discovers a component of predictability of the asset price in the next period.
2. Methodology behind the Price Increment Estimation: In the model

with a drift

one now supposes that the are serially correlated with period-to-period correlation

One can determine at time based on the obtained and sale

1. Optimal Strategy no more Static: With serial correlation the optimal trajectory is no longer a static trajectory determined in advance of trading; since each price movement gives some information about the immediate future price movements, the optimal trade list can be determined only one period at a time.
2. Estimation of the Realized Gain: Thus a full optimal solution requires the use of dynamic programming methods. However since the information is still roughly local in time, one can estimate the optimal gain attainable by an optimal strategy.
3. Almgren and Chriss (2000) Conclusions: Almgren and Chriss (2000) state their conclusion in advance of their estimation. The value of information contained in pure movements due to serial correlations is independent of the size of the portfolios being traded. The calculation demonstrated below lends intuition to this counter-intuitive statement.
4. Per Period Price Change Impact: Consider two consecutive periods during which the base strategy has the trader trading the same number of shares in each period. With a linear impact price model, in each period price changes by dollars/share. The trader pays this cost in each of the shares, so the total cost due of market impact per period is
5. Price Change from Serial Correlation: Suppose on has some price information due to serial correlations. If one knows at the previous period, then the predictable component of the price change is roughly .
6. Incremental Cost of the Adapted Strategy: But this adaptation increases the impact costs. After the shift in the first period the price change is while in the second period . These costs are paid on and shares respectively, so the market impact per period is now
7. Optimal Per-Period Execution Shift: To determine has many shares one should shift, one solves the quadratic optimization problem

The optimal is readily found as

and the maximum possible gain per period is . This heuristic can be confirmed by a detailed dynamic programming computation that accounts for optimal shifts across multiple periods.

1. Limitation of Optimal Gain Execution: Almgren and Chriss (2000) also explain briefly the limitation of the above approximation. When is close to zero, clearly this approximation is extremely close to accurate, because the persistence of the serial correlation effect dies down very quickly after the first period. When is too large to ignore, the approximation is too small for

That is, understates the possible gains over ignoring serial correlation. Conversely when

overstates the possible gains due to serial correlation. As

is more frequently the case Almgren and Chriss (2000) assert that is useful for bounding the possible gains in most situations available from serial correlations.

1. Position Independence of Gain/Cost: Note that both the size of the adaptation, and the resulting gain, are independent of the amount of shares that would be sold under an unadapted strategy. That is they are also independent of the size of the initial portfolio.
2. Gain/Cost Liquidity/Correlation Dependence: Instead the binding constraint is the liquidity of the security being traded, and the magnitude of the correlation coefficient. The more information available due to correlation and the more liquid the security, the more overall gain that is available due to adapting the strategy to the correlations.
3. Higher Order Impact Function Optimality: The results above are especially simple because of the assumption of linear impact functions. Almgren and Chriss (2000) also show briefly what happens in the more general case of nonlinear market impact functions

The cost per period due to market impact is

for small . Now the optimal shift and the maximal gain are given by

and respectively, where and are evaluated at the base execution rate of

The linear case is recovered by setting

This has the special property that is independent of and

1. Optimality Dependence on Impact Exponent: In general suppose

as

is required so that is increasing; selling the share always pushes the price down more. The marginal cost is

corresponds to an increasing marginal impact, and

corresponds to a decreasing marginal impact. Then the per-period cost one pays on the base strategy is

for large initial portfolios, and hence large rates of execution. The marginal gain from adapting to evolution is

in the same limit.

**Parameter Shifts**

1. Price Dynamics Parameter Set Shift: Almgren and Chriss (2000) discuss the impact on optimal execution of scheduled news earnings such as earnings and dividend announcements. Such events have two features that make them an important object of study. First the outcome of the event determines the shift in the parameters governing the price dynamics – see Brown, Harlow, and Tinic (1988), Easterwood and Nutt (1999), and Ramaswami (1999).
2. Determining an Event’ Full Impact: Second, the fact that they are scheduled increases the likelihood that one can detect what the true outcome of the event is. This situation is formalized below, and explicit formulas are given for price trajectories before and after the event takes place.
3. Scheduled Event Occurrence Time : Suppose at some time between now and the specified final time an event will occur, the outcome of which may or may not cause a shift in the parameters of price dynamics.
4. New Regime Shifted Parameter Set: The term *regime set* or *parameter set* refers to the collection

of the parameters that govern the dynamics at any particular time, and the events of interest are those that have the possibility of causing *parameter shifts*.

1. Initial to Final Regime Shift: Let

be the parameters of price dynamics at the time the execution begins. Suppose the market can shift to one of possible new sets of parameters so that is characterized by parameters for

1. Probability of a Regime Switch: One also supposes that probabilities can be assigned to these possible new states, so that is the probability that regime occurs. The probabilities are *independent* of the short term market fluctuations represented by . Of course it is possible that some has the same values as in which case is the probability that no change occurs.
2. Globally Optimal Dynamic Trading Strategy: Almgren and Chriss (2000) consider a dynamic trading strategy the yields globally optimal strategies in the presence of a parameter shift at time . Taking

one pre-computes an initial trajectory

with

Denote

1. Landscape of Switchable Trajectories: They also compute a family of trajectories

for

all of which have

and

They follow the trajectory until the time of the shift. Once the shift occurs they assume they can quickly identify the outcome of the event and the new set of parameters governing the price dynamics. With this settled, the complete trading using the corresponding trajectory is determined.

1. Key Almgren and Chriss (2000) Results: Almgren and Chriss (2000) show that it is possible to determine each trajectory using static optimization; although one cannot choose which one to use until the event occurs. Also the starting trajectory will *not be the same* as the trajectory one would use if they believed the regime would hold through the entire time .
2. Trajectory Conditional on Fixed : To determine the trajectories they reason as follows. Suppose that the common value of

is fixed. Then by virtue of the independence of the regime shift in itself from the security motions, the optimal trajectories conditional on the values of are simply those that have already been computed with a small modification to include the given non-zero final value.

1. Sequential Pair of Static Strategies: One can immediately write

where is determined from . The trajectory is determined the same way as seen before; it is the unique combination of the exponentials that has

and

Similarly

Thus one only needs to determine .

1. Principle behind the Estimation of : To determine one needs to determine the expected loss and the variance of the combined strategy. Let and denote the expectation and the loss incurred by the trajectory on the first segment

The quantities can be determined readily using

and

1. Mean Variance of the Compound Strategy: Then by virtue of the regime shift and the security motion’s independence, the expected loss of the compound strategy is

and its variance is

One can now do a one-variable optimization in to maximize . Almgren and Chriss (2000) provide a pictorial representation of the above.

**Conclusions and Further Extensions**

1. Efficient Frontier of Transaction Costs: The central feature of the Almgren and Chriss (2000) analysis has been to construct an *efficient frontier* in a two-dimensional plane whose axes are the expectation of the total cost and its variance.
2. Linear Impact Functions Analytical Solutions: Regardless of an individual’s tolerance to risk, the only strategies which are candidates for being optimal solutions are found in this one-parameter set. For linear impact functions, they give complete analytical solutions for the strategies in this set.
3. Efficient Frontier Optimal Operating Characteristic: Then, considering the details of risk aversion, they have shown how to select an optimal point on the frontier either by classical mean-variance optimization, or by the concept of value at risk. These solutions are easily constructed numerically, and interpreted graphically by examining the frontier.
4. First Conclusion: Sub-optimal Strategies: Because the set of attainable strategies, and hence the efficient frontier, are generally *smooth* and *convex*, a trader who is at all risk-averse should never trade according to the naïve strategy of minimizing expected cost. This is because in the neighborhood of that strategy, the first order reduction in the variance is attained at the expense of only a second order increase in the expected cost.
5. Second Conclusion: Custom Risk Optimization: Almgren and Chriss (2000) also observe that this careful analysis of the costs and risks of liquidation can be used to give a more precise characterization of the risk of holding the initial portfolio. As an example, they define a Liquidity-Adjusted VaR (L-VaR) to be, for as given time horizon, the minimum VaR of any static liquidation strategy.
6. Actual Gains of Dynamic Trading: Although it may seem counter-intuitive that the optimal strategies can be determined in advance of trading, Almgren and Chriss (2000) argue that only very small gains can be realized by adapting the strategy to the information as it is needed.
7. First Extension: Continuous Time Trading: The limit

is immediate in all of their solutions. Their trading strategy is characterized by a holdings function and a *trading* *rate*

Almgren and Chriss (2000) minimum variance strategy has infinite cost, but the optimal strategies for finite have finite cost and variance. However, this limit is at best a mathematical convenience, as the market model is implicitly a “coarse-grained” description of the real dynamics.

1. Second Extension: Nonlinear Cost: The conceptual framework outlined by Almgren and Chriss (2000) is not limited to the linear permanent and temporary impact functions

and

though the exact exponential/hyperbolic solutions are specific to that case. For nonlinear functions and that satisfy suitable convexity conditions, optimal risk-averse trajectories are found by solving a non-quadratic optimization problem; the difficulty of the problem depends on the specific functional forms chosen.

1. Third Extension Time Varying Coefficients: Almgren and Chriss (2000) framework also covers the case in which the volatility, the market impact parameters, and perhaps the expected drift are all time-dependent; finding the optimal strategy entails solving a linear system of size equal to the number of time periods (times the number of assets, for a portfolio problem). One example in which this is useful is if the price is expected to jump up or down on a known future date – say, an earnings announcement – as long as one has a good estimate of the expected *size* of this jump.

**Numerical Optimal Trajectory Generation**

1. Varying Time Interval Cost Distribution:
2. Time Varying Interval Linear Impact:
3. Varying Interval Linear Impact Objective:

implies

1. Varying Time Interval Linear Impact Jacobian:

Extrema requires that

1. Estimation Quantities for Numerical Optimization: To carry out the numerical optimization, one needs the Jacobian (i.e., gradient) and the Hessian of the optimizer objective function in terms of

This has to be computed for each of the following quantities:

* 1. Trajectory Slice Permanent Impact Function Expectation Left Holdings
  2. Trajectory Slice Permanent Impact Function Expectation Right Holdings
  3. Trajectory Slice Permanent Impact Function Expectation Cross Holdings Jacobian
  4. Trajectory Slice Temporary Impact Function Expectation Left Holdings
  5. Trajectory Slice Temporary Impact Function Expectation Right Holdings
  6. Trajectory Slice Temporary Impact Function Expectation Cross Holdings Jacobian
  7. Trajectory Slice Permanent Impact Function Variance Left Holdings
  8. Trajectory Slice Permanent Impact Function Variance Right Holdings
  9. Trajectory Slice Permanent Impact Function Variance Cross Holdings Jacobian
  10. Trajectory Slice Temporary Impact Function Variance Left Holdings
  11. Trajectory Slice Temporary Impact Function Variance Right Holdings
  12. Trajectory Slice Temporary Impact Function Variance Cross Holdings Jacobian
  13. Trajectory Slice Core Market Function Expectation Left Holdings
  14. Trajectory Slice Core Market Function Expectation Right Holdings
  15. Trajectory Slice Core Market Function Expectation Cross Holdings Jacobian
  16. Trajectory Slice Core Market Function Variance Left Holdings
  17. Trajectory Slice Core Market Function Variance Right Holdings
  18. Trajectory Slice Core Market Function Variance Cross Holdings Jacobian
  19. Trajectory Permanent Impact Function Expectation Left Holdings
  20. Trajectory Permanent Impact Function Expectation Right Holdings
  21. Trajectory Permanent Impact Function Expectation Cross Holdings Jacobian
  22. Trajectory Temporary Impact Function Expectation Left Holdings
  23. Trajectory Temporary Impact Function Expectation Right Holdings
  24. Trajectory Temporary Impact Function Expectation Cross Holdings Jacobian
  25. Trajectory Permanent Impact Function Variance Left Holdings
  26. Trajectory Permanent Impact Function Variance Right Holdings
  27. Trajectory Permanent Impact Function Variance Cross Holdings Jacobian
  28. Trajectory Temporary Impact Function Variance Left Holdings
  29. Trajectory Temporary Impact Function Variance Right Holdings
  30. Trajectory Temporary Impact Function Variance Cross Holdings Jacobian
  31. Trajectory Core Market Function Expectation Left Holdings
  32. Trajectory Core Market Function Expectation Right Holdings
  33. Trajectory Core Market Function Expectation Cross Holdings Jacobian
  34. Trajectory Core Market Function Variance Left Holdings
  35. Trajectory Core Market Function Variance Right Holdings
  36. Trajectory Core Market Function Variance Cross Holdings Jacobian
  37. Objective Utility Function Permanent Impact Function Expectation Left Holdings
  38. Objective Utility Function Permanent Impact Function Expectation Right Holdings
  39. Objective Utility Function Permanent Impact Function Expectation Cross Holdings Jacobian
  40. Objective Utility Function Temporary Impact Function Expectation Left Holdings
  41. Objective Utility Function Temporary Impact Function Expectation Right Holdings
  42. Objective Utility Function Temporary Impact Function Expectation Cross Holdings Jacobian
  43. Objective Utility Function Permanent Impact Function Variance Left Holdings
  44. Objective Utility Function Permanent Impact Function Variance Right Holdings
  45. Objective Utility Function Permanent Impact Function Variance Cross Holdings Jacobian
  46. Objective Utility Function Temporary Impact Function Variance Left Holdings
  47. Objective Utility Function Temporary Impact Function Variance Right Holdings
  48. Objective Utility Function Temporary Impact Function Variance Cross Holdings Jacobian
  49. Objective Utility Function Core Market Function Expectation Left Holdings
  50. Objective Utility Function Core Market Function Expectation Right Holdings
  51. Objective Utility Function Core Market Function Expectation Cross Holdings Jacobian
  52. Objective Utility Function Core Market Function Variance Left Holdings
  53. Objective Utility Function Core Market Function Variance Right Holdings
  54. Trajectory Slice Core Market Function Variance Cross Holdings Jacobian

1. Permanent Impact Expectation Left Sensitivity:
2. Permanent Impact Expectation Right Sensitivity:
3. Permanent Impact Expectation Cross Jacobian:
4. Temporary Impact Expectation Left Sensitivity:
5. Temporary Impact Expectation Right Sensitivity:
6. Temporary Impact Expectation Cross Jacobian:
7. Trajectory Jacobian and Hessian Computation: In general, the trajectory Jacobian’s and the Hessian’s may be computed as a sequential, aggregate accumulations over the corresponding slices, with one very critical caveat. In the automated sensitivity generation schemes, all sensitivities to the left-most and the right-most nodes must be excluded, since these do not constitute the control nodes.
8. Power Objective Function Rationale/Formulation: A generalization of the mean-variance optimization and the Value-at-risk schemes is the power objective function formulation

corresponds to the regular mean-variance optimization scheme, and

corresponds to the liquidity based VaR formulation.

1. Liquidity VaR Control Jacobian/Hessian:

**References**

* Admati, A., and P. Pfleiderer (1988): A Theory of Intra-day Patterns: Volume and Price Variability *Review of Financial Studies* **1** 3-40.
* Almgren, R., and N. Chriss (1999): Value under Liquidation *Risk* **12 (12)** 61-63.
* Almgren, R., and N. Chriss (2000): Optimal Execution of Portfolio Transactions *Journal of Risk* **3 (2)** 5-39.
* Artzner, P., F. Delbaen, J. M. Eber, and D. Heath (1997): Thinking Coherently *Risk* **10 (11)** 68-71.
* Beaver, W. (1968): The Information Content of Annual Earnings Announcements, in: *Empirical Research in Accounting; Selected Studies*. Supplement to *Journal of Accounting Research* 67-92.
* Bertsekas, D. P. (1976): *Dynamic Programming and Stochastic Control* **Academic Press**.
* Bertsimas, D., and A. W. Lo (1998): Optimal Control of Execution Costs *Journal of Financial Markets* **1** 1-50.
* Brown, K., W. Harlow, and S. Tinic (1988): Risk Aversion, Uncertain Information, and Market Efficiency *Journal of Financial Economics* **22** 355-385.
* Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997): *Econometrics of Financial Markets* **Princeton University Press**.
* Chan, L. K. C., and J. Lakonishok (1993): Institutional Trades and Intra-day Stock Price Behavior *Journal of Financial Economy* **33** 173-199.
* Chan, L. K. C., and J. Lakonishok (1995): The Behavior of Stock Prices around Institutional Trades *Journal of Finance* **50** 1147-1174.
* Charest, G. (1978): Dividend Information, Stock Returns, and Market Efficiency – II *Journal of Financial Economics* **6** 297-330.
* Dann, L. (1981): Common Stock Re-purchases: An Analysis of Returns to Bond-holders and Stock-holders *Journal of Financial Economics* **9** 113-138.
* Easterwood, J. C., and S. R. Nutt (1999): Inefficiency in Analysts’ Earnings Forecasts: Systematic Misreaction or Systematic Optimism? *Journal of Finance* **54 (5)** 1777-1797.
* Fama, E., L. Fisher, M. Jensen, and R. Roll (1969): The Adjustment of Stock Prices to new Information *International Economic Review* **10** 1-21.
* Grinold, R. C., and R. N. Kahn (1999): *Active Portfolio Management 2nd Edition* **McGraw-Hill**.
* Holthausen, R. W., R. W. Leftwich, and D. Mayers (1987): The Effects of Large Block Transactions on Security Prices: A Cross-Sectional Analysis *Journal of Financial Economy* **19** 237-267.
* Holthausen, R. W., R. W. Leftwich, and D. Mayers (1987): Large Block Transactions, the Speed of Response, and Temporary and Permanent Stock-Price Effects *Journal of Financial Economy* **26** 71-95.
* Kalay, A., and U. Loewentstein (1985): Predictable Events and Excess Returns: The case of Dividend Announcements *Journal of Financial Economics* **14** 423-449.
* Keim, D. B., and A. Madhavan (1995): Anatomy of the Trading Process: Empirical Evidence on the Behavior of Institutional Traders *Journal of Financial Economy* **37** 371-398.
* Keim, D. B., and A. Madhavan (1997): Transaction Costs and Investment Style: An Inter-exchange Analysis of Institutional Equity Trades *Journal of Financial Economy* **46** 265-292.
* Kim, O., and R. Verrecchia (1991): Market Reaction to Anticipated Announcements *Journal of Financial Economics* **30** 273-310.
* Kraus, A., and H. R. Stoll (1972): Price Impacts of Block Trading at the New York Stock Exchange *Journal of Finance* **27** 569-588.
* Krinsky, I., and J. Lee (1996): Earnings Announcements and the Components of the Bid-Ask Spread *Journal of Finance* **51** 1523-1555.
* Kyle, A. S. (1985): Continuous Auctions and Insider Trading *Econometrica* **53** 1315-1336.
* Lee, C. M. C., B. Mucklow, and M. J. Ready (1993): Spreads, Depths, and the Impact of Earnings Information: An Intra-day Analysis *Review of Financial Studies* **6** 345-374.
* Lo, A. W., and A. C. MacKinlay (1988): Stock Market Prices do not follow Random Walks: Evidence from a Simple Specification Test *Review of Financial Studies* **1** 41-66.
* Morse, D. (1981): Price and Trading Volume Reaction surrounding Earnings Announcements: A Close Examination *Journal of Accounting Research* **19** 374-383.
* Patell, J. M., and M. A. Wolfson (1984): The Intra-day Speed of Adjustment of Stock Prices to Earnings and Dividend Announcements *Journal of Financial Economics* **13** 223-252.
* Perold, A. F. (1988): The Implementation Short-fall: Paper versus Reality *Journal of Portfolio Management* **14** 4-9.
* Ramaswami, M. (1999): Stock Volatility declines after Earnings are announced – Honest! *Global Weekly Investment Strategy* **Lehman Brothers**.

**Optimal Execution with Non-linear Impact Functions and Trading-Enhanced Risk**

**Abstract**

1. Price and Market Impact Volatility: Almgren (2003) determines optimal trading strategies for the liquidation of a large single-asset portfolio to minimize a combination of volatility risks and market impact costs.
2. Power Law Market Impact Function: The market impact cost is taken to be a power law of the trading rate with an arbitrary positive exponent. This includes, for example, the square root law that has been proposed based on market microstructure theory.
3. Holdings Size Dependent Characteristic Time: In analogy with the linear model, a *characteristic time* is defined for optimal trading, which now depends on the initial portfolio size and decreases as the execution proceeds.
4. Trade Size Dependent Liquidity Volatility: Also considered is a model in which the uncertainty of the realized price is increased by demanding rapid execution; it is shown that the optimal trajectories are defined by a *critical portfolio size* above which this effect is dominant and below which this effect may be neglected.

**Introduction**

1. Active vs. Passive Execution Strategy: In the execution of large portfolio transactions, a trading strategy must be determined that balances the risk of delayed execution against the cost of rapid execution; the choice is roughly between an *active* and *passive* trading strategy (Hasbrouck and Schwartz (1988), Wagner and Banks (1992)).
2. Construction of Optimal Execution Strategies: Several papers have constructed optimal strategies for the problem (Almgren and Chriss (1999), Grinold and Kahn (1999), Almgren and Chriss (2000), Konishi and Makimoto (2001) under the assumption that the liquidity costs per share traded are a linear function of trading rate or block size, and that the only source of volatility in execution is the price volatility of the underlying asset.
3. Price Effect of Block Trades: There is an extensive literature studying the effects of block trades on prices – see Kraus and Stoll (1972), Holthausen, Leftwich, and Mayers (1987, 1990), Chan and Lakonishok (1993, 1995), Keim and Madhavan (1995, 1997), and Koski and Michaely (2000).
4. Assumption of Linear Trading Costs: In practice linearity of trading costs is an unrealistic assumption. Perold and Salomon Jr. (1991) have argued that the liquidity premium per share demanded by the market will either be a convex or a concave function of the block size depending on whether the market’s perception is that the trader is information driven or liquidity driven, respectively.
5. Barra Market Impact Liquidity Premium: In the Barra Market Impact Model (Loeb (1983), Kahn (1993), Barra (1997), Grinold and Kahn (1999)) it is argued, based upon the detailed analysis of the risk-reward choice faced by the equity market maker that the liquidity premium per share should grow as the square root of the block size traded.
6. Block Size Dependent Liquidity Premium: Electronic trading systems such as Optimark (Rickard and Torre (1999)) have been constructed to allow traders specify precisely what liquidity premium they are willing to pay as a function of the block size, and search for clearing opportunities in the mismatch between profiles of different market participants. These effects can be captured by introducing *nonlinear impact functions* into the cost function which is minimized to determine the optimal trading strategies.
7. Approaches of Nonlinear Models: Although linear models are commonly used in empirical regression analyses for simplicity, nonlinear models can often be emulated by dividing trades into categories by size (Bessembinder and Kaufmann (1997), Huang and Stoll (1997)). In fact, Chakravarthy (2001) argues medium-sized trades have a disproportionately large effect on prices.
8. Handling Non-Deterministic Liquidity Premiums: An additional effect not considered in the theoretical strategies considered in the previous work is that the liquidity premium demanded by the market is not deterministic. In fact the premium will depend on the presence in the market at that instant of participants who are willing to take the other side of the trade.
9. Motivation for Trading Enhanced Risk: Since the presence of these counterparties cannot be predicted in advance, it represents an additional source of risk incurred by the trading profile. That is, a more complete model should include *trading-enhanced risk* representing the increased uncertainty in the execution price incurred by demanding rapid execution of large blocks.
10. Manifestation of Stochastic Liquidity Premiums: Trading-enhanced risk is an implicit feature in the model described in Rickard and Torre (1999). Chordia, Subrahmanyam, and Anshuman (2001) and Hasbrouck and Seppi (2001) argue that liquidity fluctuates due to intrinsic variations in the market activity independent of the trade size. This effect is included in the model by Almgren (2003) via the constant term but additional interest is in the *increase* in the execution price uncertainty due to larger block sizes.
11. Nonlinear Block Size Dependence: Thus Almgren (2003) extends the models of Grinold and Kahn (1999) and Almgren and Chriss (2000) in a few important ways. First, the liquidity premium, expressed as an unfavorable motion of the price per share, may be an increasing nonlinear function of the trading rate and the block size – one is considered to be a proxy for the other.
12. Power Law Premium - Closed Form: This cost is reduced by trading slowly, but it must be balanced against the volatility risk incurred by holding the initial portfolio longer than is necessary. In particular, exact solutions are provided in the case this function is a power law with an arbitrary positive exponent, which covers the range of behavior outlined above.
13. Dependence on Initial Portfolio Size: Whereas in the linear case optimal trajectories are characterized by a single *characteristic time* independent of the initial portfolio size, in the nonlinear case the characteristic time depends upon the initial portfolio size, and scales appropriately as the remaining portfolio diminishes during trading.
14. Comparison with Price Volatility Risk: The realized price per share itself is a random variable, whose variance increases ith the increased rate of trading. This introduces an additional source of risk in addition to the volatility. In contrast to the effect of market volatility, this additional risk is *decreased* by trading slowly, submitting small blocks for execution at each time.
15. Volatile Liquidity - Closed Form Solutions: Nearly explicit optimal solutions including this effect can be constructed, and an asymptotic analysis can be used to show that the effect of trading-enhanced risk is most important for large initial portfolios. Indeed for any given set of parameters there is a characteristic portfolio size above which the optimal strategy is determined by the need to reduce trading-enhanced risk, and below which this effect may be ignored.

**The Model**

1. Holdings Trade and Liquidation Time: The general framework followed is that from Almgren and Chriss (2000). At time

shares of an asset are held, which are to be completely liquidated by the time

The initial size is positive for a sell program and negative for a buy program; in the former case, there is a long exposure to the market until all the holdings have been eliminated, while in the latter case there is short exposure to the market until the purchase to which the trader has committed to at

is completed. The focus here is on the case

In the case of a portfolio trading problem may be a vector, but the consideration here is only on a single asset.

1. The Problem Trade List Determination: denotes the holdings at time with

and

The problem is to choose an optimal function so as to minimize a chosen cost functional. Later the limit

will be taken in which the natural execution time emerges as a result of the analysis, but for now the consideration is on an exogenously imposed time horizon.

1. Origin of the *Static* Strategy: It is a rather surprising fact that in the absence of serial correlation in the asset price movements, the optimal price may be determined *statically* at the start of the trading. Unless the market parameters change, observations of price movements in the course of trading do not convey any information that would lead to a change in the strategy.
2. Evenly Spaced Discrete Time Intervals: The analysis starts with the construction of a discrete time model. Thus for a given trading interval

for

with

and let be the holdings at time with

and

The sales between times and are

corresponding to the velocity

Thus

1. Generating the Optimal Trade List: In the discrete time model there is no assumption that the shares are traded at a uniform rate *within* each interval. Rather the assumption is that the trader achieves the optimal execution possible subject to the constraint that shares are to be traded in the next time interval . The functions introduced below are a model to describe the trader’s best efforts.
2. Temporary/Permanent Market Impact Components: On a standard manner (Stoll (1985)) the impact is divided into a permanent and a temporary component. Thus describes the price per share of the asset that is publically available in the market.
3. Discrete Arithmetic Permanent Impact Component: The price satisfies the arithmetic random walk

where are independent random variables with zero mean and unit variance, is an *absolute* (not percentage) volatility, is the *permanent impact function* representing the effect of the share price of the information conveyed by the trade. This effect is generally small, and below is taken to be a linear function, in which case it will have no effect on determining the optimal strategy.

1. Nonlinear Temporary Impact Component: The price that one actually gets on the trade is

Here is a nonlinear *temporary impact function* representing the price concession one must accept in order to trade

shares in time . The random variables are independent of each other and of with zero mean and unit variance. The new function represents the uncertainty of the trade execution as a function of the block size.

1. Liquidity Volatility Term Time Dependence: The factor in the last term of

simply represents a scaling of the parameters if is fixed and finite. When varies, for example when the continuous time

is taken, this factor is necessary to preserve the effect of the trading-enhanced risk.

1. Liquidity Volatility Incremental Time Dependence: If the above term were not present, then breaking a block into several smaller blocks would diversify away the risk due to the uncertainty of each one, regardless of the form of the risk.
2. *Capture* of the Trade Program: The *capture* of the trade program is the total cash received
3. The Trade Program *Implementation Cost*: Discounting is ignore since the trading horizon is short. The *implementation cost* is - a random variable due to uncertainties in price movements and realized prices.
4. Components of the *Implementation Cost*: Note that the implementation cost includes both the costs of finite liquidity and price uncertainty due to delayed execution. This is the *implementation shortfall* of Perold (1988) – see also Jones and Lipson (1999).
5. Implementation Cost Mean and Variance: Its expectation and variance at

depend on the free parameters of the trade strategy:

1. Mean Variance Optimal Static Strategies: A rational trader will construct his or her own strategies to minimize some combination of and . As advances the values of and change, but if and are constructed using a classic mean-variance approach, the optimal strategy continues to be the one determined initially (Almgren and Chriss (2000), Huberman and Stanzl (2005)).
2. Continuous Time Limit Trading Strategy: Now, for analytical convenience, the continuous time limit

is taken. The trade strategy becomes a continuous path and the block sizes are assumed to be well-behaved so that

with

1. Continuous Time Mean and Variance: The above expressions have finite limits

where the square brackets indicate that these are *functionals* of the entire continuous-time path .

1. Caveat behind Continuous Time Analytics: It needs to be emphasized that the continuous time limit is simply an analytical device for obtaining solutions when is reasonably small; in reality the discreteness of the trading intervals must be taken into account in order to correctly describe trading-enhanced risk.
2. Mean Variance Optimization Objective Function: Introducing the risk-aversion parameter , the combined quantity

is minimized. Whether or not mean variance optimization is appropriate for a particular case, may be considered to be a Lagrange/KKT multiplier for the constrained problem of minimizing for a given and used to construct an efficient frontier in the space of trading trajectories.

1. VaR Based Optimization Objective Functions: More general weightings of risk, including Value-at-Risk, present thorny conceptual problems for time-dependent strategies (Artzner, Delbaen, Eber, and Heath (1999), Basak and Shapiro (2001)).
2. The Calculus of Variations Approach: Minimizing is a standard problem in the calculus of variations:

with

1. Perturbation Stationarity: Euler-Lagrange Equation: Stationarity to small perturbations requires that the optimal solve the Euler-Lagrange equation

– a second order ordinary differential equation to be solved with respect to the given endpoints and

1. Integration into the First Order Form: Since does not depend explicitly on multiplying throughout by and integrating results in the first-order equation
2. Application to Optimal Trajectory Determination: In the current case one obtains

with

1. Properties of the Almgren “P” Function: The constant of integration

is the velocity with which hits

For a sell program with

and conversely for a buy program. Note that

additional assumption is that is always an *increasing* function of and hence invertible.

1. Explicit Solutions - Key Simplifying Assumptions: Almgren (2001) makes two simplifying assumptions to obtain explicit solutions.
   1. Permanent impact is linear in the trading rate.
   2. The imposed time horizon is infinite.
2. Linear Permanent Market Impact Function: A linear cost function

gives a total cost independent of the path . The first term on the right side of

vanishes, and then since appears only on the left side and itself appears only on the right side, the general solution can be written in the quadrature form as

1. Bid Ask Permanent Linear Absent: The constant is to be chosen so that

corresponds to

Note also that any constant in disappears; the bid-ask spread does not affect the optimal strategy.

1. No Extraneously Specified Liquidation Time: Since is an increasing function, so is . It is thus clear that as decreases towards zero, the liquidation time increases.
2. Invoking Longest Possible Liquidation Time: If no time horizon is exogenously imposed, the longest possible liquidation time can be obtained by setting

which leads to the quadrature problem

1. Tractability of the above Solution: Often analytic solutions to the above problem can be found when

with

would be too intractable. These solutions will still give nearly complete liquidations in finite time determined by market parameters.

**Nonlinear Cost Functions**

1. Power Law Temporary Impact Functions: Restricting the attention to the sell program, with

the temporary impact functions are taken to be

with

- for a buy program the signs will be changed in an obvious way. The linear case corresponds to

1. Temporary Impact Almgren “P” Function: As noted above a possible constant in corresponding to the bid-ask spread has been neglected. Then

which, for the case of a general finite time horizon with

leads to the quadrature problem

1. Longest Optimal Trajectory Explicit Solution: Taking

explicit solutions for the longest optimal trajectories can be obtained:

1. Characteristic Time for Optimal Execution: Here the *characteristic time* is

This is the analog of the *half-life* in the linear case. Only in the linear case

is independent of the initial portfolio size . For

the characteristic time depends on the initial size as

1. Sub Linear Power Law Exponent: For

rapid trading is *under*-penalized relative to the linear case. As the portfolio size increases, volatility risk dominates the trading costs, and the optimal trading time *decreases* since the exponent is negative.

1. Supra Linear Power Law Exponent: For

rapid trading is *over*-penalized relative to the linear case. As the portfolio size increases, the trading cost dominates the volatility risk, and the optimal trading time *increases*, since the exponent is positive. For example, if

then

1. Characteristic Time vs Half Life: As the portfolio size decreases to zero, reconciliation of the optimal trajectory would use a different starting value and hence a different time . The meaning of is thus a little less fundamental than in the linear case. However, scales in exactly the right way to make still a static solution.
2. Intuition behind the Characteristic Time: For more intuition, note that the initial rate of selling is

and is the solution to the relation

1. Characteristic Time as a Cost Balance: The left side is the risk penalty associated with holding shares for a time , and the right side, up to a factor , is , the impact cost associated with selling shares over a time (without the constant term representing the bid-ask spread, which does not impact the optimal solution).
2. The Longest Optimal Execution Time: For

the trajectory reaches

with

at a finite time

Thus these trajectories are the solution for finite imposed time if

the trajectory reaches at and stays there till

1. Self-Similar Scaling Trajectory Form: Almgren (2003) contains a graphical illustration of the optimal trajectories generated from

The form of the portfolio is independent of a particular choice of time scale and initial portfolio size ; these solutions may be easily scaled to any case.

1. Time Realization of Trajectory Differences: A sense of the differences between the solutions may be gained by noting that for short times, all optimal trajectories are fairly close to each other; but the *tail* of the trajectories are extended for small value of which strongly penalize trading at slow rates.
2. Example: Sell-Order Exponent Dependence: For example, as shown by Almgren (2003), at

the optimal trajectories reduce holdings to , , and of the initial portfolio for

and

respectively.

1. Example Sell Time Exponent Dependence: Further as demonstrated by Almgren (2003), at

the trajectory for

has reached

and remains there, the trajectory for

retains of its initial holdings, and the trajectory for

retains of the initial holdings. The relative differences become even more pronounced as time continues.

**Objective Function**

1. Determination of and : and can be explicitly computed for these solutions from

and hence the frontier can be drawn. In doing this the contributions from and the term in are neglected.

1. Closed Form for and : Then for a general
2. The Efficient Frontier Curve: As varies moves along the hyperboloid-like curve

For any positive there is a unique solution.

1. Asymptotics of and : As

one gets

and

i.e., optimizing expected cost without regard to variance leads to use of all available time. As

it can be seen that

and

i.e., uncertainty is minimized regardless of the cost.

**Almgren (2003) Example**

1. Reference Trading Rate/Market Depth: Almgren (2003) provides a sample methodology for the estimation of the parameters. The first is to choose a representative level of trading rate If a specific time period is chosen is equivalent to a certain block size

traded in the time period; it may be interpreted as the market *depth* in the sense of Kyle (1985) or Bondarenko (2001).

1. Choice of Reference Trading Rate: The examples below consider stocks that trade one million shares a day, and is taken to be of that rate, or

For time period

with this rate is equivalent to trading a block of approximately shares in each hour.

1. The Corresponding Temporary Price Impact: Next the price impact which would be incurred by the steady trading at the reference rate would be chosen. The share price is assumed to be and the assumption is that trading

incurs a price impact of or

1. Choosing the Exponent - The Rationale: Finally a choice for the value of the exponent is made that best fits the belief of how the price impact would depend upon the trading rate for rates smaller or larger than .
2. Reminder Trading Rate Dependence: The choice

corresponds to the linear dependence of price impact on rate;

means that *large* trading rates or block sizes have a disproportionately *large* effect on price; while

means that large trading rates or block sizes have a relatively *smaller* impact.

1. Based Impact Model: This impact model is then written as

or

1. Daily Volatility and Initial Portfolio: It is also supposed that the stock has an annual volatility of for the daily price change of

A portfolio of initial size

is considered, equal to of the daily volume.

1. Construction of the Efficient Frontier: There is now enough information to construct the efficient frontier

from and for any chosen describing the family of solutions as the risk aversion parameter ranges over all the possible values

To construct particular optimal solutions a specific value for needs to be set.

1. Variance/Cost Dependence on : The results are shown numerically in the table below. For any value of the natural liquidation time increases with the *risk tolerance* parameter ; as both increase the expected cost decreases and the variance increases.
2. Impact: : The table below shows that the optimal time scale , the expected cost , and the standard deviation of the cost as functions of the risk tolerance parameter and the temporary impact exponent . Market and portfolio parameters are as given in the treatment above (the initial portfolio value is ). As is varied, the reference values and are held constant; thus the coefficient varies as in

or

Time is measured in days; , , and are in thousands of dollars.

1. Table of :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 1 | 0.02 | 0.07 | 0.22 |
| 10 | 0.09 | 0.22 | 0.46 |
| 100 | 0.40 | 0.71 | 1.00 |
| 1000 | 1.84 | 2.24 | 2.15 |
| 10000 | 8.55 | 7.07 | 4.64 |
|  | 1 | 221 | 354 | 462 |
| 10 | 103 | 112 | 99 |
| 100 | 48 | 35 | 21 |
| 1000 | 22 | 11 | 5 |
| 10000 | 10 | 4 | 1 |
|  | 1 | 11 | 19 | 30 |
| 10 | 23 | 33 | 45 |
| 100 | 49 | 59 | 65 |
| 1000 | 105 | 106 | 96 |
| 10000 | 226 | 188 | 141 |

1. Large Execution Time Dependence: For large values of , optimal trajectories all execute rapidly, to reduce the volatility risk associated with the portfolio. When the trading rate is larger than costs *increase* with increasing so larger leads to slower trading.
2. and Combination Impact: Conversely for small generally trading proceeds *more slowly* than in order to minimize the total expected cost. In this regime *smaller* is more expensive and leads to relatively slower trading.
3. Cross-Over Execution Time: In the intermediate parameter regime the trajectories cross-over from one behavior to the other; larger suggests slower trading at the beginning when the rate is large, then relatively more rapid in the tail.
4. Characteristic Time Reference Rate Dependence: Note that

may be re-written as

from which it is clear that

as

regardless of the values of the other parameters.

1. Trade Speed vs. Cost Balance: In this limit, trading more rapidly than the reference rate is very strongly penalized, while trading more slowly is almost without cost, so the optimal strategy is to always trade exactly at the critical rate.
2. Estimation from Reference Parameters: Finally since is a difficult parameter to select in practice, it may be observed that it can be estimated if a time scale is chosen from

and

or

- it can be found that

1. Trading Cost vs. Variance Balance: The numerator is the price concession per share for trading at a concession rate multiplied by the total number of shares to get the expected cost; the denominator is the variance that would be incurred by holding shares for time . This ratio is multiplied by to correct for nonlinearities which are ignored by this simple description.

**Trading-Enhanced Risk**

1. Liquidity Volatility: Functional Form Considered: Now the following functional form is taken for a sell program with

The deterministic part of the temporary impact is the linear case

of the previous section.

1. Trading Rate Independent Volatility Component: The constant term in , with coefficient , represents a constant uncertainty in the realized sale price independent of the rate of selling and of the underlying process. The total risk associated with this term is minimized by splitting the sale into as many parts as possible; thus this term pushes towards the linear trajectory.
2. Trading Rate Dependent Volatility Component: The linear term, with coefficient , represents the increase in variance caused by non-zero amounts of selling. This term can even more strongly push toward the linear trajectory.
3. Liquidity Volatility Almgren “P” Function: Then with

becomes

1. Behavior of the Almgren “P” Function: The polynomial has

and is increasing for

so the graph of the trajectory is always convex, and the inverse of is well-defined. For a buy program with

the sign of the odd term in is reversed.

1. No Hard Maximum Execution Time: Since

for near zero, the integrand appearing in the quadrature formulation

behaves as as

for

and there is no “hard” maximum time as was found above for

**Constant Enhanced Risk**

1. Analytical Solution for Case: Two special cases are considered for obtaining analytical solution. The first is

With this assumption the price uncertainty on each trade is independent of the size of the trade.

1. Trading Trajectory and Execution Time: A solution can then be found for
2. Comparison with , Case: This is a pure exponential solution, except that the time constant has been increased by adding the additional variance per transaction to the impact coefficient
3. Expressions for and : The value functions are
4. and Behavior: The optimal value functions change in a more complicated way than the trajectory. As

the behavior is the same as that found earlier;

and

since there is less care about the enhanced risk.

1. Behavior: In contrast as

all quantities have finite limits;

and

Since trading itself introduces risk, risk-aversion and cost reduction both encourage spreading the trade over several periods; the minimum variance solution takes finite time and has finite cost.

**Linear Enhanced Risk**

1. Analytical Solutions for the Case: The next special case is

and

becomes

and thus

1. Trading Trajectory and Characteristic Fields: This can be integrated to obtain

in which the characteristic time and the characteristic share level are

and the nonlinear function is

where

1. Intuition behind the Characteristic Size: The characteristic time is the same as in the earlier section for

and does not depend on the new coefficient . To understand the characteristic level note that

1. Trading Enhanced Market Volatility Balance: In this expression the left side is the trading induced variance in the share price given by the model

if an initial portfolio of size were sold in a single period The right side is the variance in the share price due to the volatility in the same time interval; at the characteristic share level these two quantities are of comparable size.

1. much bigger than : To compare with the previous results note that

If this limit is attained by taking a limit of the *parameters* so that

so that

is valid uniformly over , since

- a pure exponential solution results

which, in particular, recovers the previous result with

in the limit

1. Behavior Towards ; Trajectory *Tail*: And for any fixed value of the parameters

describes the tail of the solution as

the time constant of the decay is not affected by the addition of .

1. much smaller than : For

i.e., the initial behavior when

using the expansion

gives

with

This is the same solution constructed in the earlier Section with

with

1. Almgren (2003) Asymptotic Solution Illustration:

together with

and

are illustrated in elaborate figures in Almgren (2003) Figure 5.

1. Strategy Construction Approach: Starting Trajectory: Thus the optimal strategy for construction would be as follows. Assuming

the initial trades are done using the trajectories of the temporary impact power law with

with

That is, the volatility due to trading completely dominates the intrinsic volatility

1. Strategy Construction Approach: Tail Trajectory: As reaches the level switch is done to the optimal solution in the linear case

with the other parameters taking their market values. In the tail trading-enhanced risk is a negligible quantity compared to the volatility.

**Almgren (2003) Nonlinear Example Sample**

1. Working out the Case: In this case the focus is on the previous section in which

and

so that trading enhanced risk increases linearly with block size with no constant term.

1. The Corresponding Discrete Price Equation: To estimate the coefficients, one starts with the discrete time model. With

and

the price model

becomes

1. Liquidity Risk as Price Volatility Fraction: Assuming that for a particular choice of the trading interval the standard deviation of price concession associated with trading-enhanced risk is a fraction of the deterministic impact – a plausible assumption since both quantities are linearly proportional to the block size.
2. The Corresponding Characteristic Price: That is

or

which gives

1. Dependence on Risk Aversion: At this portfolio size, the volatility risk of holding the portfolio roughly balances the risk of selling along the optimal trajectory. Al though both of these quantities are risks, involves through its influence on trading time .
2. Choice of : The market parameters are taken to be the same as in the earlier section, with

corresponding to the case where the liquidation is one a day. The trading is divided into one hour time intervals, so

and

1. Estimate of and : One obtains

and

corresponding to a portfolio size of . A liquidation problem with initial value greater than will begin in the large regime where trading-enhanced risk is dominant and end in the small regime where it is negligible.

**Conclusions: Summary and Extensions**

1. Summary: Power Law Temporary Component: The treatment seen above obtains explicit analytical solutions for certain cases of the impact model. First, it neglects the effects of trading-enhanced risk, and takes the impact function to be a simple power law. The solutions in this case are straightforward nonlinear extensions of Almgren and Chriss (2000); the exponential solutions obtained there are a particular dividing case of these power law solutions.
2. Summary: Constant Trading-Enhanced Risk: With trading-enhanced risk, two particular cases with linear impact functions were considered. If the price uncertainty per transaction is independent of the transaction size, then the optimal trajectories are given by the previous results, simply augmenting the impact coefficient by the additional variance. A risk-averse trader lengthens his trade program, diversifying some variance away by spreading the execution over more different transactions at the expense of slightly higher volatility risk.
3. Summary: Linear Trading-Enhanced Risk: If price uncertainty per transaction is linearly proportional to the transaction size, then a characteristic portfolio size emerges, above which reduction of the added variance is the dominant effect. In this regime trade trajectories are equivalent to the previous power law solutions with exponent equal to 3. For portfolios smaller than this size the new effect may be neglected compared to the deterministic impact costs and volatility.
4. Extension #1: Optimal Numerical Trajectories: Throughout this treatment, the focus has been on obtaining explicit solutions for the sake of analytical insight. Numerical solutions would be quite straightforward, and allow lifting the restrictions described above, and consideration of a more general class of models.
5. Extension #2: Linear Impact Portfolios: Portfolios of assets are an interesting extension. Already in the linear case (Almgren and Chriss (2000)), to obtain explicit solutions it is necessary to make simplifying assumptions about cross-impacts, for example, that trading in each asset affects only the price of that asset.
6. Extension #3: Nonlinear Impact Portfolios: Even with that assumption, the nonlinear formulation opens a wide class of possible models; for example, should the exponent be the same for each asset? Determination and characterization of optimal trajectories in this case is a topic for future work.

**References**

* Almgren, R. F., and N. Chriss (1999): Value under Liquidation *Risk* **12 (12)** 61-63.
* Almgren, R. F., and N. Chriss (2000): Optimal Execution of Portfolio Transactions *Journal of Risk* **3 (2)** 5-39.
* Almgren, R. F. (2003): Optimal Executions with Nonlinear Impact Functions and Trading-Enhanced Risk *Applied Mathematical Finance* **10 (1)** 1-18.
* Artzner, P., F. Delbaen, J. M. Eber, and D. Heath (1999): Coherent Measures of Risk *Mathematical Finance* **9** 203-228.
* Barra (1997): *Market Impact Model Handbook*.
* Basak, S., and A. Shapiro (2001): Value-at-risk Based Risk management: Optimal Policies and Asset Prices *Review of Financial Studies* **14** 371-405.
* Bessembinder, H., and H. M. Kaufmann (1997): A Comparison of Trade Execution Costs for NYSE and NASDAQ-Listed Stocks *Journal of Financial and Quantitative Analysis* **32** 287-310.
* Bondarenko, O. (2001): Competing Market Makers, Liquidity Provisions, and Bid-Ask Spreads *Journal of Financial Markets* **4 (3)** 269-308.
* Chakravarthy, S. (2001): Stealth Trading: Which Traders’ Trades moves Prices? *Journal of Financial Economics* **61** 289-307.
* Chan, L. K. C., and J. Lakonishok (1993): Institutional Trades and Intra-day Stock Price Behavior *Journal of Financial Economy* **33** 173-199.
* Chan, L. K. C., and J. Lakonishok (1995): The Behavior of Stock Prices around Institutional Trades *Journal of Finance* **50** 1147-1174.
* Chordia, T., A. Subrahmanyam, and V. R. Anshuman (2001): Trading Activity and Expected Stock Returns *Journal of Financial Economics* **59** 3-32.
* Grinold, R. C., and R. N. Kahn (1999): *Active Portfolio Management 2nd Edition* **McGraw-Hill**.
* Hasbrouck, J., and R. A. Schwartz (1988): Liquidity and Execution Costs in Equity Markets *Journal of Portfolio Management* **14** 10-16.
* Hasbrouck, J., and D. J. Seppi (2001): Common Factors in Prices, Order Flows, and Liquidity *Journal of Financial Economics* **59** 383-411.
* Holthausen, R. W., R. W. Leftwich, and D. Mayers (1987): The Effects of Large Block Transactions on Security Prices: A Cross-Sectional Analysis *Journal of Financial Economy* **19** 237-267.
* Holthausen, R. W., R. W. Leftwich, and D. Mayers (1990): Large Block Transactions, the Speed of Response, and Temporary and Permanent Stock-Price Effects *Journal of Financial Economy* **26** 71-95.
* Huang, R. D., and H. R. Stoll (1997): The Components of the Bid-ask Spread: A General Approach *Review of Financial Studies* **10 (4)** 995-1034.
* Huberman, G., and W. Stanzl (2005): Optimal Liquidity Trading *Review of Finance* **9 (5)** 165-200.
* Jones, C. M., and M. L. Lipson (1999): Execution Costs of Institutional Liquidity Orders *Journal of Financial Intermediation* **8** 123-140.
* Kahn, R. N. (1993): How the Execution of Trades is best Operationalized, in: *Execution Techniques, True Trading Costs, and the Microstructure of Markets*, K. F. Sherrerd (Editor) **AIMR**.
* Keim, D. B., and A. Madhavan (1995): Anatomy of the Trading Process: Empirical Evidence on the Behavior of Institutional Traders *Journal of Financial Economy* **37** 371-398.
* Keim, D. B., and A. Madhavan (1997): Transaction Costs and Investment Style: An Inter-exchange Analysis of Institutional Equity Trades *Journal of Financial Economy* **46** 265-292.
* Konishi, H., and N. Makimoto (2001): Optimal Slice of a Block Trade **3 (4)** 33-51.
* Koski, J. L., and R. Michaely (2000): Price, Liquidity, and the Information Content of Trades *Review of Financial Studies* **13** 659-696.
* Kraus, A., and H. R. Stoll (1972): Price Impacts of Block Trading at the New York Stock Exchange *Journal of Finance* **27** 569-588.
* Kyle, A. S. (1985): Continuous Auctions and Insider Trading *Econometrica* **53** 1315-1336.
* Loeb, T. F. (1983): Trading Costs: The Critical Link between Investment Information and Results *Financial Analysts Journal* **39** 39-44.
* Perold, A. F. (1988): The Implementation Short-fall: Paper versus Reality *Journal of Portfolio Management* **14** 4-9.
* Perold, A. F., and R. S. Salomon Jr. (1991): The Right Amount of Assets under Management *Financial Analysts Journal* **47** 31-39.
* Rickard, J. T., and N. G. Torre (1999): Information Systems for Optimal Transaction Implementation *Journal of Management Information Systems* **16** 47-62.
* Stoll, H. R. (1989): Inferring the Components of the Bid-Ask Spread: Theory and Empirical tests *Journal of Finance* **44** 115-134.
* Wagner, W. H., and M. Banks (1992): Increasing Portfolio Effectiveness via Transaction Cost Management *Journal of Portfolio Management* **19** 6-11.

# Market Impact Function/Parameters Estimation

### Introduction, Overview, and Background

1. Power Law Temporary Impact Function: The impact of large trades on prices is very important and widely discussed, but rarely measured. Using a large data set from a major bank and a simple but theoretical model, Almgren, Thum, Hauptmann, and Li (2005) propose that the impact is a law of the block size, with specific dependence on trade duration, daily volume, volatility, and shares outstanding.
2. Incorporation into Scheduling/Cost Estimation Algorithms: The results can be directly incorporated into an optimal trade scheduling algorithms and into pre- and post-trade estimation systems.
3. Performance Impact of Transaction Costs: Transaction costs are widely recognized as an important determinant of investment performance (see, for example, Freyre-Sanders, Guobuzaite, and Byrne (2004)). Not only do they affect the realized results of an active investment strategy, they also control how rapidly assets can be converted into cash should the need arise.
4. Direct Fixed Transaction Cost Component: Such costs generally fall into two categories. First are the direct costs such as commissions and fees that are explicitly stated and easily measured. These are important and should be minimized, but are not the focus here.
5. Indirect Controllable Transaction Cost Component: Indirect costs are not explicitly stated. For large trades the most important component of these is the impact of the traders’ own actions on the market. These costs are notoriously difficult to measure but they are most amenable to careful trade management and execution.
6. Calibration of Market Impact Costs: Almgren, Thum, Hauptmann, and Li (2005) present a quantitative analysis of the market impact costs based on a large of Citigroup US brokerage executions. A simple theoretical model that brings in the very important role of execution is used.
7. Out-of-Sample Cross Validation: The model and its calibration are constructed to satisfy two criteria. First, the predicted costs are quantitatively accurate, as determined by direct fit and out-of-sample back testing, as well as extensive consultations with the traders and the other market participants.
8. Deployability with External Execution Schedulers: The results may be directly used as an input into optimal portfolio scheduling systems, although the scheduling algorithm itself may be non-trivial.
9. Use in Citigroup’s BECS System: The results of this study have been incorporated into Citigroup’s Best Execution Consulting Services (BECS) software for use internally at all desks as well as the clients of the equity division. While this work has focused on US markets, it has been extended to global equities. BECS is the delivery platform for Citigroup’s next generation of trading analytic tools, both pre- and post- execution.
10. Extension to the Standard Trading Model: This pre-trade analyzer is an extension to the market standard existing model that has been delivered through the Stock Facts Pro software for the past 25 years (Sorensen, Price, Miller, Cox, and Birnbaum (1998)).
11. Solid Empiricals and Real-Data Verifications: This model is based on better developed empirical foundations; it is based on real trading data taking time into consideration while verifying the results through post-trade analysis. The table below summarizes some of the advantages/disadvantages of this approach.
12. Distinguishing Features of the Model:
    1. Advantages
       1. Calibrated from Real Data
       2. Includes Time Component
       3. Incorporates Intra-day Profiles
       4. Uses non-linear Impact Functions
       5. Confidence Levels for Coefficients
    2. Disadvantages
       1. Based only on Citigroup Data
       2. Little Data for Small-Cap Stocks
       3. Little Data for very large Trades
13. Academic Industrial Market Data Quantification: Much work in both the academic and industrial communities has been devoted to understanding and quantifying market impact costs. Many academic studies have only worked with publically available data in Trade and Quote (TAQ) tick-record from the New York Stock Exchange (NYSE).
14. Buy-Sell Market Imbalance Analysis: Breen, Hodrick, and Korajczyk (2002) regress the net markets movements over five-minute and half-hour time periods against a net buy-sell impact during the same period, using a linear impact model. A similar model is developed in Kissell and Glantz (2003).
15. Impact Cost Function Dependence Analysis: Rydberg and Shephard (2003) develop a rich econometric foundation for describing price motions; Dufour and Engle (2000) investigate the key role of waiting period between successive trades. Using techniques from statistical physics, Lillo, Farmer, and Mantegna (2003) look for a power law scaling in the impact cost function, and find significant dependence on total market capitalization as well as daily volume, and Bouchaud, Gefen, Potters, and Wyart (2004) discover non-trivial serial correlation ion volume and price data.
16. Limitations of Public Data Sets: The publically available data sets lack the reliable classification of individual trades as buyer- or seller- initiated. Even more significantly, each transaction exists in isolation; there is no information on the sequence of trades that form part of the larger transaction.
17. Transaction Trade Sequence Incorporation Studies: Some academic studies have used limited data sets made available by asset managers that do have this information, where the date, but not the time duration of the trade is known (Holthausen, Leftwich, and Mayers (1990), Chan and Lakonishok (1995), and Keim and Madhavan (1996)).
18. Permanent and Temporary Impact Costs: The transaction cost model embedded in this analysis is based on the model presented by Almgren and Chriss (2000) with non-linear extensions from Almgren (2003). The essential features of this model, as described below, is that it explicitly divides the market impact costs into a permanent component associated with information and a temporary component arising from the liquidity demands made by the execution in a short time.

**Data Description and Filtering Rules**

1. Data Generation Period and Universe: The data on which the analysis was based on contains, before filtering, almost 700,000 US stock orders executed by the Citigroup equity trading desks for a 19-month period from December 2001 to June 2003.
2. Orders, Transactions, and Resulting Executions: Each order is broken down into one or more transactions, each of which may generate one or more executions. The information presented below is available for each order.
3. Symbol, Size, and Order Type: The stock symbol, the requested order size, and the sign (buy/sell) of the entire order – the client information is removed.
4. Order Submission Time and Method: The times and the methods by which the transaction was submitted by the Citigroup trader to the market. The time of the first transaction is taken to be the start of the order. Some of the transactions are sent as market orders, some as limit order, and some are submitted to the Citigroup’s automated VWAP server. Except for the starting time and except to exclude the VWAP orders, no use is made of this transaction information.
5. Execution Times, Sizes, and Prices: The times, the prices, and the sizes corresponding to the execution of each transaction is used. Some transactions are cancelled or inly partly executed; only completed sizes and prices are used. The execution times re denoted by , sizes by , and prices by .
6. Order Completion Finish Time Frame: All orders are completed within one day, though not necessarily filled.
7. Additional Types of Information Available: In addition, various additional pieces of information – such as instructions given by the client to the trader for the order – e.g., ‘market on close’, ‘market on open’, ‘over the day’, VWAP or blank – are available.
8. Sample Subset Filtering Criteria: the total sample contains 682,582 orders, but only a subset is used for the data analytics.
9. S & P 500 Constituent Stocks Only: To exclude small and thinly traded stocks, only orders on the Standard and Poor’s Index are considered, which represent about half of the total number of orders, but a large majority of the dollar value. Even within this universe there is enough diversity to explore the dependence on the market capitalization, as there are both NYSE and OTC stocks.
10. Exclusion of Highly Volatile Stocks: Also excluded are approximately 400 orders for which the stocks exhibit more than 12.5% daily volatility – 200% annual volatility.
11. Orders that Match the Analysis Objectives: Furthermore, only those orders that are reasonably representative of the actual scheduling strategies that are the ultimate goal are considered.
12. Excluding Market-on-Open/Close: The orders for which the client requested market-on-open or market-on-close executions are excluded. These orders are likely to be executed with strongly non-linear profiles that do not satisfy the modeling assumption – there are only a few hundred of these.
13. Excluding Client Requested VWAP Orders: The orders for which the client requested VWAP execution are excluded. These orders have consistently long execution times and represent very small rates of trading relative to the market volume. These are about 16% of the total number of orders.
14. Excluding Later-in-the-Day Transactions: Also excluded are orders for which any executions are recorded after 16:10 EST, approximately 10% of the total. In many cases these use Citigroup’s block desk for some or all of the transactions, and the fills are reported sometime after the order is completed. Therefore the time information is not reliable.
15. Caveat - Impact of Orders Exclusion: This exclusion, together with the use of fill size in place of originally requested size, could be a source of significant bias. For example, if clients and traders consistently used limit orders, orders may be filled only if the prices moved in a favorable direction. Analysis of the data set suggests that this effect is not significant – for example the same coefficients are obtained with or without partially filled orders – and informal discussions with the traders confirm the belief that partial fills are not the result of a limit order strategy.
16. Other Minimum Cut off Criteria: Most significantly, small order are excluded since the goal is to estimate transaction costs in the range where they are significant. Specifically only the following orders are included:
    1. The order has at least two completed transactions.
    2. Orders are at least 1,000 shares.
    3. Orders are at least 0.25% of average daily volume in that stock.
17. Range of Execution Sizes/Times: The results of the model are reasonably stable to changes in the above criteria. After this filtering there are 29,509 orders in the data set; the largest number of executions for any order is 548, and the median is around 5. The median time is around 5 minutes.
18. Order/Volume Ratio Range: The table below shows some descriptive statistics for the sample. Most of the orders constitute only a few percent of the typical market volume, and the model is designed to work within this range of values. Orders greater than a few percent of daily volume have substantial sources of uncertainty that are not modeled here, and the model does not represent them.
19. Summary Statistics of the Sample Orders:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Minimum** | **Q1** | **Median** | **Q3** | **Maximum** |
| Total Cost % | 0.04 | -3.74 | -0.11 | 0.03 | 0.19 | 3.55 |
| Permanent Cost % | 0.01 | -3.95 | -0.17 | 0.01 | 0.19 | 2.66 |
| Temporary Cost % | 0.03 | -3.57 | -0.11 | 0.02 | 0.17 | 2.33 |
| Shares/ADV % | 1.51 | 0.25 | 0.38 | 0.62 | 1.36 | 88.62 |
| Time Days | 0.39 | 0.00 | 0.10 | 0.32 | 0.65 | 1.01 |
| Daily Volatility % | 2.68 | 0.70 | 1.70 | 2.20 | 3.00 | 12.50 |
| Mean Spread % | 0.14 | 0.03 | 0.03 | 0.11 | 0.16 | 2.37 |

**Data Model - Variables**

1. Market Impact Input Dependence Estimation: The goal of the study is to determine the market impact in terms of a small number of input variables. Below is a list of precisely which market impacts are measured, and what primary and auxiliary variables will be used to model them.
2. Pre- and Post- Market Prices: Let be the price of the asset being traded. For each order the following price points of interest are defined: is the market price before this order begins executing; is the market price after this order is completed; and is the average realized price on the order.
3. The Transaction Weighted Average Price: The realized price

is calculated from the transaction data set. The market price and are the bid-ask mid points from .

1. First Transaction Pre- Trade Price: The pre-trade price is the price before the impact on the trade begins to be felt (this is an approximation, since some information may leak before any record enters the system). is computed from the latest quote just preceding the first transaction.
2. Post-trade Price Capture - Caveat: The post-trade price should capture the permanent effects of the trade program. That is, it should be long enough after the last execution that any effects of temporary liquidity have dissipated.
3. Accounting for the Permanent Impact: In reportedly performing the fits, Almgren, Thum, Hauptmann, and Li (2005) have found that 30 minutes after the last execution is enough to achieve this. For shorter time intervals, the regressed values depend on the time lag, and about this level the variation stops. That is, they define
4. Delay Date Roll Over: The price is taken from the first quote following . If is after the market close, it carries over to the next morning. This risks distorting the results by including excessive overnight volatility, but Almgren, Thum, Hauptmann, and Li (2005) have found this to give more consistent results than truncating at the market close.
5. Permanent Realized Impact Variables Definition: Based on these prices the following dimensionless impact variables are defined. The dimensionless permanent impact is

and the dimensionless realized impact is

1. Conversion into Observed Market Impacts: The “effective dimensionless impact is the quantity of most interest, since it specifies the actual cash spent or received on the trade. In the model below the temporary impact will be defined to be minus a suitable fraction of and this temporary impact will be the quantity described by the theory.
2. Signs of the Impact Variables: On any individual order, the signs of , can be positive or negative. In fact since volatility is a very large contributor to either values, they are almost likely to have either sign. They are defined so that positive cost is experienced if , have the same sign as the total order ; for a buy order with

positive cost means that the price moves upwards. The average values of , taken across many orders is expected to have the same sign as .

1. Intra-day Volume Weighted Time: The level of market activity is known to vary substantially and consistently over different periods of the trading day; this intra-day variation affects both the volume profile and the variance of prices. To capture this effect, all computations are performed in volume time which represents the fraction of the average day’s volume that has executed up to the time .
2. Intra-day Volume Weighted Trajectory: Thus a constant rate trajectory in the variable corresponds to a VWAP execution in real time. The relationship between and is independent of the daily trading volume; it is scaled so that

at market open and

at market close.

1. Intra-day Volume Weighted Times: Each of the clock times in the data set is mapped to the corresponding volume time .Since the stocks in the sample are heavily traded a non-parametric estimator that directly measures the differences in is used; the shares traded during the period correspond to the execution of each order.
2. Time Volume vs. Price Volatility: Almgren, Thum, Hauptmann, and Li (2005) display an illustration of the empirical profiles. The fluctuations in each time period in these illustrations correspond to the approximate size of statistical error in the volume calculation for a 15-minute trade; these errors are typically less than 5%, and are smaller for longer periods.
3. The “Dimensional” Parametric Explanatory Variables: The impacts and are to be described in terms of the following quantities.
   1. Total executed size in shares
   2. Volume Duration of Active Trading:
   3. Volume Duration of the Impact:
4. Caveats around the Explanatory Variables: As noted above, is positive for a buy order, and negative for a sell order. Explored defining using a size weighted average of execution times, but the results are not substantially different. The intermediate execution times are not used, and the execution sizes are not used either except in calculating the order size and the mean realized prices.
5. Fixed Trade - Optimal Time Nodes: In the eventual application for trajectory optimization, the size will be assumed given, and the execution schedule here represented by will be optimized.
6. Execution Time as Optimizing Parameter: In general the solution will be a complicated time dependent trajectory parametrized by a time scale . For the purposes of data modeling the trajectory optimization is ignored and the schedules are taken to be determined only by a single number .
7. Market Core Empirical Parametric Inputs: Although the goal is to explain the dependence of the impact costs , on order size and trade time , other market variables will influence the solution. The most important of these are: – which is the average daily volume in shares, and – the daily volatility.
8. Daily Volume/Volatility “Wander” Scale: is a day moving average. For volatility, an intra-day estimator that makes use of every transaction in the day is used. It is important to track changes in these variables not only between different stocks but also across time for the same stock.
9. Order Size/Daily Volume Normalization: These values serve primarily to “normalize” the active variables across the stocks with widely varying properties. It seems natural that order size should be measured as a fraction of the average daily volume : is a more natural variable than itself.
10. Intrinsic Notion of Volume Time: In the model presented below, the order size as a fraction of the average volume traded during the time of execution will also be seen to be important. is estimated directly by taking the average volume that executed between the times and over the previous days. In fact, since in the model the trade duration appears only in the combination this avoids the need to measure directly.
11. “Wander” Scale of Market Impact: The volatility is used to scale the impacts – a certain level of participation in the daily volume should cause a certain level of participation in the ‘normal’ motion of the stock. Empirical investigation by Almgren, Thum, Hauptmann, and Li (2005) shows that volatility is the most important scale factor for cost impact.

**Trajectory Cost Model**

1. Constant Volume Time Trading Rate: The model used is based on the framework developed by Almgren and Chriss (2000), and Almgren (2003), with simplifications made to facilitate data fitting. The main simplification is that the rate of trading is constant (in volume time). In addition cross impact is neglected, since the data has no information about the effect of trading one stock on the price of the other.
2. The Permanent Impact Market Component: The price impact is decomposed into two components. First is a permanent component that reflects the information transmitted to the market by the buy/sell imbalance. This component is believed to be roughly independent of trade scheduling; ‘stealth’ trading is not admitted by this construction. In the data fit this component will be independent of the execution time .
3. The Temporary Market Impact Component: A temporary component reflects the price concession needed to attract counterparties within a specified short time interval. This component is highly sensitive to trade scheduling; here it will strongly depend on .
4. Other Elaborate Market Impact Frameworks: More detailed conceptual frameworks have been developed (Bouchaud, Geffen, Potters, and Wyart (2004)), but this easily understood model has become standard in industry and academic literature (Madhavan (2000)).
5. Decomposition of the Realized Market Impact: The realized price impact is a combination of the above two effects. In terms of the realized and the permanent impact defined above and observed from the data, the model may be summarized as

with suitable coefficients and scaling depending upon . Thus the temporary impact is obtained as a difference between the permanent impact and the realized impact; it is not directly observed, although there is a direct model for it.

1. Uniform Rate of Order Liquidation: The starting point is the initial order demand of shares. This is assumed to be completed by a uniform rate of trading over a volume interval . That is, the trade rate in volume units is

and is held constant until the program is completed.

1. Sign of the Trade Rate: Constant rate in these units is equivalent to VWAP execution during the time of execution. Note that has the same sign as ; thus

for a buy order and

for a sell order. Market impact will move the price in the same direction as .

**Permanent Impact**

1. Volatility/Permanent Impact Price Change: The model postulates that the asset price follows an arithmetic Brownian motion with a drift term that depends on the trade rate term . That is

where is a standard Brownian motion (or a Bachelier process); and is the permanent impact function; the only assumptions made are that is increasing and has

1. Integrated Form of Price Change: As noted above, is volume time, representing the fraction of an average day’s volume that has executed so far. This expression can be integrated in time taking to equal for

to obtain the permanent impact

where

is a standard Gaussian variable.

1. Linearity of the Permanent Impact Function: Note that if is a linear function, then the accumulated drift at time is equal to , the number of shares executed to time , and the permanent impact is proportional to the total order size independently of the time scale .

**Temporary Impact**

1. Temporary Impact Price Change Realization: The actual price received from the trade is

where is the temporary impact function. For convenience, it has been scaled by the market price at the start of trading, since the time intervals involved are all less than one day.

1. Discretization of the Price Impact: This expression is a continuous time approximation to a discrete process. A more accurate description would be to imagine that the time intervals would be broken down into intervals such as, say, one hour or minute intervals. Within each interval the average price realized on the trade during that interval would be less favorable that the average price that an unbiased observer would measure during that time interval.
2. Unbiased Price Plus Liquidity Concession: The unbiased price is affected by the previous trades that have been executed before this interval (as well as the volatility) but not on their timing. The additional concession during this time interval is strongly dependent on the number of shares executed in this interval.
3. Closed Form Temporary Impact Expression: At a constant liquidation rate, calculating the time average of the execution price gives the temporary impact expression

where

is independent of . The term reflects the effect on the later execution prices of permanent impact caused by the earlier parts of the program.

1. Estimate of the Heteroscedastic Corrections: The rather complicated error expression reflects the fluctuations on the middle part of the Brownian motion on relative to the end point at . It is only used for the heteroscedastic corrections for the regression fits below.
2. Fluctuations and Error Residuals Estimation: The equations and provide explicit expressions for the permanent and the temporary impact components , in terms of the values of the functions and at known trade rates, together with the estimates of the magnitude of the error coming from the volatility.
3. Regression Based Impact Form Estimation: The data fitting procedure above is in principle straightforward, the impacts and are computed from the transaction data, and those values are regressed against order sizes and times as indicated to directly extract the functions and .

**Choice of the Functional Form**

1. Permanent/Temporary Impact Function Structure: The next question that needs to be addressed is what should the structure of the permanent impact function and the temporary impact function be. Even with a large sample it is not possible to extract these functions purely from data, so a hypothesis must be made about their structure.
2. Power Law Impact Functional Forms: The postulate is that these functions are power laws, that is, that:

and

where the numerical values of the dimensionless coefficients and and the exponents and are to be determined by linear and non-linear regressions on the data. The sign is to be chosen so that and have the same sign as .

1. Range of Power Law Representation: This class of power law is extremely broad. It includes concave functions (exponent < 1), convex functions (exponent > 1), and linear functions (exponent = 1). It is the functional form that is implicitly assumed by fitting straight lines on a log-log plot as is very common in physics, and has been used in this context, for example, by Lillo, Farmer, and Mantegna (2003).
2. Order Type/Exchange Parameter Independence: The same coefficients are taken for buy orders

and sell orders

It would be a trivial modification to introduce different coefficients and for the two sides, but the exploratory analysis by Almgren, Thum, Hauptmann, and Li (2005) has not indicated a string need for this. Similarly it would be possible to use different coefficients for stocks traded on different exchanges, but this does not appear to be necessary either.

1. Quasi Arbitrage Permanent Impact Elimination: There is reason to be specific in the choice of the exponents. For the permanent impact function there is a strong reason to prefer the linear model with

This is the only value for which the model is free from quasi-arbitrage (Huberman and Stanzl (2004)).

1. Linearity of the Permanent Impact: Furthermore, the linear function is the only one for which the permanent price impact is independent of the trading time. Of course this substantial conceptual simplification must be supported by the data.
2. Concave Nature of the Temporary Exponents: For temporary impacts, there is ample evidence indicating that the function should be concave, that is

This evidence dates back to Loeb (1983) and is strongly demonstrated by the fits in Lillo, Farmer, and Mantegna (2003). In particular theoretical arguments (Barra (1997)) suggest that the particular value of

is especially plausible, resulting in a square root impact function.

1. Verification of Power Exponent Values: The approach is then as follows. Unprejudiced fits to the power law functions shall be made to the entire data set to determine the best estimates for the exponents and . The validity of the values

and

will then be tested to validate the linear and the square root candidate functional forms.

1. Determination of the Impact Coefficients: Once the exponents have been selected, simple linear regression is adequate to determine the coefficients. In this regression heteroscedastic weightings are used, with the error magnitudes from

and

The result of this regression is not only the values for the coefficients, but also a collection of the error residuals and which must be tested for normality as the theory supposes.

**Cross-Sectional Description**

1. Motivations for the Properties Normalization: The above analysis has assumed an ‘ideal’ asset, all of whose properties remain constant in time. For any real asset, the parameters that determine the market impact will vary with time. For example one would that the execution of a given number of shares would incur higher impact costs on a day with unusually low volume or unusually high volatility.
2. Basis for the Normalizer Choice: That is, the impact of the cost functions should be expressed in terms of the dimensionless quantity rather than itself, where is the average number of shares per day defined above.
3. Normalization of the Price Moves: Furthermore, the motion of the price should not be given as a raw percentage figure, but it should be expressed as a fraction of ‘normal’ daily motion of the price, as expressed by the volatility .
4. Normalized Expressions for , : With these assumptions, the equations

and

can be modified to

and

respectively, where is the error expression depending on the volatility.

1. Dimensionless Permanent Temporary Function Inputs: Now and are dimensionless functions of a dimensionless variable. They are assumed to be constant in time for a single stock across days when and vary. The next step is to investigate these functions for their dependence on cross-stock variables.

**Model Determination**

1. Specification of Extraneous Model Regressors: To bring the full size of the data into play, one must address the more complex and the less precise question of how the impact functions vary across the stocks, that is, how much they depend variables such as market capitalization, shares outstanding, bid-ask spread, or other quantities. Temporary and permanent impact must be considered separately.
2. Permanent Impact Function Liquidity Regressor: A ‘liquidity factor’ is inserted into the permanent cost function , where depends on the market parameters characterizing each stock (in addition to daily volume and liquidity). There are several candidates for inputs into .
3. Liquidity Regressor Candidate - Inverse Turnover: The form of is constrained to be

where is the total number of shares outstanding, and is the exponent to be determined. The dimensionless quantity is the inverse of the ‘turnover’ – the fraction of the company’s value traded each day. This is a natural explanatory variable, and has been used in empirical studies such as Breen, Hodrick, and Korajczyk (2002).

1. Liquidity Regressor Candidate – Bid-Ask: Almgren, Thum, Hauptmann, and Li (2005) did not find any consistent dependence on the bid-ask spread across the sample, so it is not included in .
2. Liquidity Regressor Candidate - Market Capitalization: This differs from the shares outstanding by the price per share, so including this factor is equivalent to including a ‘price effect’. The study by Almgren, Thum, Hauptmann, and Li (2005) found that there is a persistent price effect, as also found by Lillo, Farmer, and Mantegna (2003), but that the dependence is weak is enough that it may be neglected in favor of the conceptually simpler quantity .
3. Temporary Impact Function - Regressor Candidate: In further extensive preliminary exploration, it was found that the temporary cost function does not require any stock-specific modification; liquidity costs as a fraction of volatility only depends upon the fraction of the shares traded as a fraction of the average daily volume.
4. Revised , Functional Forms: After assuming the functional form defined above, the model is validated and the exponent is determined by performing a non-linear regression of the form

and

where is the again the heteroscedastic error term from

and is the sign function.

1. Estimates and Residuals of Exponents: A modified Gauss-Newton optimization algorithm was used to determine the values of , , and that minimized the normalized residuals. The results are:
2. Errors Represented as One Sigma Amounts: Here, as throughout this chapter, the error bars are expressed with are one standard deviation, assuming Gaussian error model. Thus the ‘true’ value can be expected to be within this range with probability, and within a range twice as large with probability.
3. Choice of Linear Permanent Impact: From these values the following conclusions can be drawn. First the value

for linear impact cannot be reliably rejected. In view of enormous practical simplification of the linear permanent impact

is chosen.

1. Permanent Impact Liquidity Exponent Estimation: The liquidity factor is very approximately
2. Temporary Impact Power Law Exponent: For temporary impact, the analysis confirms the concavity of the function with strictly less than one. This confirms the fact that the bigger the trades made by the fund managers on the market, the less additional cost they experience per share traded. At confidence level, the square root model

is rejected. The temporary cost exponent is therefore fixed at

In comparison with the square root model, this gives slightly smaller costs for smaller trades, and slightly larger costs for large trades.

1. Permanent Dependence on Shares Outstanding: Note that because

for fixed values of number of shares in the order, and the average daily volume , the cost increases with , the total number of shares outstanding. In effect a large number of outstanding shares means that a smaller fraction of the company is traded each day, so a given fraction of that flow has a greater impact.

1. Linear Permanent Concave Temporary Impact: Therefore the results confirm empirically the theoretical arguments of Huberman and Stanzl (2004) for permanent impact that is linear in the block size, and the concavity of the temporary impact has been widely described in the literature for both theoretical and empirical reasons.

**Determination of the Coefficients**

1. Estimation of the Impact Coefficients: After fixing the exponent values, the values of and are determined by linear regression of the models.

and

using the heteroscedastic error estimates given in

and

It is found that

with

and

with

1. Interpretation of the statistic: The statistic is calculated assuming that the Gaussian model expressed in

and

is valid; the error estimates are the values divided by the statistic. Although the actual residuals are fat-tailed as discussed below, these estimates indicate that the coefficient values are highly significant.

1. Permanent Impact Signal Contribution: The values are typically less than indicating that only a small part of the dependent variables and is explained by the model in terms of the independent variables. This is precisely what is expected given the small size of the random impact term relative to the random motion of the price due to the volatility arising from the trade execution.
2. Importance of the Permanent Cost: The permanent persistent cost, though small, is of major importance since it is on the average the cost incurred by the fund managers while trading. Furthermore since most orders are part of large portfolio trades, the volatility costs experienced on the portfolio level is considerably lower than exhibited in the stock-level analysis, increasing the significance of the fraction of the impact cost estimated. As previously mentioned the non-linear optimization of the volatility versus impact cost trade-off would reveal additional profitable strategies.
3. Universal Coefficients of Market Impact: The dimensionless numbers and are the universal coefficients of market impact. According to the model, they apply to every order and every asset in the entire data set.
4. Interpretation Caveat vs. : To summarize, they are to be inserted into the equations

and

giving the expectation of the impact costs; in any particular order the realized values will vary greatly due to the volatility. To reiterate, does not signify the total cost, but is simply the net price motion from pre-trade to post-trade. The actual cost experienced by the trade is signified by .

1. Sub Group Impact Parameters Determination: Almgren, Thum, Hauptmann, and Li (2005) have chosen the simple form above to have a single model that applies reasonably well across the entire data set which consists entirely of large cap stocks in the US market. More detailed models could be constructed to capture more limited data or assets, or to account for variations across global markets. In practice, it is expected that the coefficients, perhaps even the exponents, or maybe even the functional forms, will be continually updated to reflect the most recent data.
2. Example of Impact Cost: The table below shows the impact cost functions and the numerical examples for two large cap stocks when the customer buys 10% of the average daily volume. The permanent cost is independent of the time of execution, the temporary cost depends on the time of execution, but across different asset it is the same fraction of the daily volatility. is written as
3. Example of Impact Costs Table:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | **IBM** | | | **DRI** | | |
| Average Daily Volume | Million |  | 6.561 | | | 1.929 | | |
| Shares Outstanding | Million |  | 1728 | | | 168 | | |
| Inverse Turnover |  |  | 263 | | | 87 | | |
| Daily Volatility | % |  | 1.57 | | | 2.26 | | |
| Normalized Trade Rate |  |  | 0.1 | | | 0.1 | | |
| Normalized Permanent Impact |  |  | 0.126 | | | 0.096 | | |
| Permanent Price Impact | bp |  | 20 | | | 22 | | |
| Trade Duration | Days |  | 0.1 | 0.2 | 0.5 | 0.1 | 0.2 | 0.5 |
| Normalized Temporary Impact |  |  | 0.142 | 0.094 | 0.054 | 0.142 | 0.094 | 0.054 |
| Temporary Impact Cost | bp |  | 22 | 15 | 8 | 32 | 21 | 12 |
| Realized Cost | bp |  | 32 | 25 | 18 | 43 | 32 | 23 |

1. Example of Impact Cost – Analysis: In the above example, because DRI turns over of its float each day, whereas IBM turns over only , trading of the day’s volume causes a permanent price move of only 0.1 times volatility for DRI, but 0.13 times for IBM; half of this is experienced as cost. Because the permanent cost function is linear, the permanent cost numbers are independent of the times of execution.

**Residual Analysis**

1. Results of Impact and Market: The results of the above analysis are not simply the values of the coefficients presented. In addition, the error formulation provides specific predictions for the nature of the residuals and for the permanent and the temporary impact respectively from

and

1. Validating the Independence of the “Wanderers”: Under the assumption that the asset price is a Brownian motion with the drift caused by the impact, these two variables should be independent standard Gaussians. This assumption has already been used in heteroscedastic regression, now it needs to be verified.
2. Residual Mean and Covariance: Almgren, Thum, Hauptmann, and Li (2005) demonstrate histograms and plots of and . The means are quiet close to zero, the variances a r reasonably close to , and the correlation is reasonably small.
3. Fat-Tailed Nature of the Distribution: But the distribution is extremely fat-tailed, as is normal for returns distributions on short-time intervals (see Rydberg (2000)), and hence does not indicate that the model is poorly specified. Nonetheless, the structure of the residuals confirms that the model is close to the best that can be obtained within the Brownian framework.

**References**

* Almgren, R. F., and N. Chriss (2000): Optimal Execution of Portfolio Transactions *Journal of Risk* **3 (2)** 5-39.
* Almgren, R. F. (2003): Optimal Executions with Nonlinear Impact Functions and Trading-Enhanced Risk *Applied Mathematical Finance* **10 (1)** 1-18.
* Almgren, R. F., C. Thum, E. Hauptmann, and H. Li (2005): Equity Market Impact *Risk* **18 (7)** 57-62.
* Barra (1997): *The Market Impact Handbook*.
* Bouchaud, J. P., Y. Gefen, M. Potters, and M. Wyart (2004): Fluctuations and Responses in Financial Markets: The “Subtle” Nature of Random Price Changes *Quantitative Finance* **4 (2)** 176-190.
* Breen, W., L. Hodrick, and R. Korajczyk (2002): Predicting Equity Liquidity *Management Science* **48 (4)** 470-483.
* Chan, L. K. C., and J. Lakonishok (1995): The Behavior of Stock Prices around Institutional Trades *Journal of Finance* **50** 1147-1174.
* Dufour, A. and R. Engle (2000): Time and the Price Impact of a Trade *Journal of Finance* **55 (6)** 2467-2498.
* Freyre-Sanders, A., R. Guobuzaite, and K. Byrne (2004): A Review of Trading Cost Models *Journal of Investing* **13** 93-115.
* Holthausen, R. W., R. W. Leftwich, and D. Mayers (1990): Large Block Transactions, the Speed of Response, and Temporary and Permanent Stock-Price Effects *Journal of Financial Economy* **26** 71-95.
* Huberman, G., and W. Stanzl (2004): Price Manipulation and Quasi-arbitrage *Econometrica* **72 (4)** 1247-1275.
* Keim, D. B., and A. Madhavan (1995): Anatomy of the Trading Process: The Upstairs Market for Large Block Transactions; Analysis and Measurement of Price Effects *Review of Financial Studies* **9** 1-36.
* Kissell, R., and M. Glantz (2003): *Optimal Trading Strategies* **Amacom**.
* Lillo, F., J. Farmer, and R. Mantegna (2003): Master Curve for Price-Impact Function *Nature* **421** 129-130.
* Loeb, T. (1983): Trading Cost: The Critical Link between Investment Information and Results *Financial Analysts Journal* **39 (3)** 39-44.
* Rydberg, T. (2000): Realistic Statistical Modeling of Financial Data *International Statistical Review* **68 (3)** 233-258.
* Rydberg, T., and N. Shephard (2003): Dynamics of Trade-by-trade Price Movements; Decomposition and Models *Journal of Financial Economics* **1 (1)** 2-25.
* Sorensen, E., L. Price, K. Miller, D. Cox, and S. Birnbaum (1998): *The Solomon Smith Barney Global Equity Impact Cost Model* Technical Report **Solomon Smith Barney***.*

**Competitive Bids for Principal Program Trades**

**Introduction**

1. The Definition of the Program Trade: The program trade involves the sale or the purchase of a basket of stocks that is too large to be traded immediately in the market. When such trades are brokered, they take on of two forms.
2. The Agency Program Trade Type: In the *agency trade* type the broker executes the trade on behalf of the client on a commission basis, and all the risk of the trade is borne by the client.
3. The Principal Program Trade Type: In a *principal trade*, also called a principal basket, principal bid, or risk bid, the broker directly purchases the entire basket for a fixed price, usually expressed as a discount to the fair market value. By design, principal trades transfer all of the risk from the client to the broker in exchange for a single price, which therefore proxies for the risk of the market portfolio.
4. Program Trades Share of NYSE: Program trading represents an increasing percentage of the overall stock market volume. In 2002 program trades averaged over 30% of the New York Stock Exchange trading volume, up from approximately 20% in 19999 and 2000.
5. Share of the Principal Trades: Overall 50% of all trading volume took place in the NYSE, and of this 30-40% was done on a principal basis (NYSE (2002)).
6. Attributes of a Principal Trade: Almgren and Chriss (2003) begin by making two key observations about the program trading business. First the principal trade consists of two attributes – a basket of stocks, and a price.
7. The Basket Components and Price: The basket is determined by the client, but the price – usually expressed as a per share discount to the current market value – is agreed upon by the client and the broker, and the potential profitability of the trade depends upon the precise price that can be secured for trading the basket. Throughout this chapter, the trade will be treated as the logical unit which consists of both the basket and its price.
8. Corporate Finance View of the Principal Trade: The second observation is that since program trading represents an investment of the firm’s capital, the correct way to view performance is on an annualized basis.
9. First Result - The “Optimal” Execution: The treatment here constructs a mathematical framework for pricing and trading of principal baskets, yielding two main results. First it shows that for a broad class of measures of annualized risk-adjusted return there is a unique optimal way to trade the basket.
10. Second Result - The Information Ratio: Second, the measure called the *information ratio* of a trade is introduced, which is the ratio of the annualized expected profit to the annualized standard deviation of the profit.
11. Information Ratio as a Relative Value Metric: Given a proposed trade one can calculate the information ratio from the known information about the basket. Because the information ratio is annualized by the expected time to completion of the liquidation, it provides a way to compute the profitability of trades of different sizes and levels of liquidity. This yields a powerful tool for analyzing principal businesses.
12. Accommodating Various Market Impact Models: One does not need to know all of the constituents of a basket to compute its information ratio – just the volatility, the liquidity, and a proposed price for the basket. In particular, one must know the effect of trading a basket on the prices of its constituents, that is, the market impact. The methods in this treatment fit in with a wide variety of existing models.
13. The “Hurdle Rate” Minimum Price: The information ratio is primarily a pricing tool. By specifying a hurdle rate – a minimum information ratio that every principal basket must exceed – one can determine for a given principal basket what the minimum price will be that exceeds the hurdle rate.
14. Viability of the “Going Price”: Alternatively it may be used as an evaluation tool. If one knows the “going price” of a trade, one can compute the information ratio of the trade based on that price to determine whether it is worthwhile to submit a winning bid for the business.

**Efficient Frontier Pricing of Program Trades**

1. *Ex-Ante* Risk Adjusted Return: The program trade is essentially the use of capital by a trading desk. In this treatment, the use of this capital is evaluated using an *ex-ante* risk adjusted return ratio – the ratio of the predicted profit to the standard deviation of the predicted profit. This turns out to the familiar information ratio, analogous to the familiar Sharpe ratio of the Information Theory.
2. The Investment Horizon Cost/Variance: Almgren and Chriss (2003) argue that for a principal desk engaged in an ongoing business, the correct approach is to *annualize* the cost and the variance, placing them in the context of other investment opportunities.
3. Discount Based Optimal Execution Point: Remarkably for each value of the discount received for trading the basket, there is then a *single* optimal point, independent of the risk preferences. This point corresponds to the single best value for the overall information ratio.
4. “Put Out To Bid” Call: The value of this ratio is therefore a potential tool to be used in evaluating whether to accept a certain piece of business at a certain price. In many situations the program trading business is “put out to bid”. That is, the portfolio manager contacts multiple desks about a particular portfolio.
5. Information Ratio as a Decision Tool: Each business responds with a certain bid – the discount to the fair market value required to do the business. Program trading desks often know the level of bid required to win the business, and therefore the information ratio can be used as a hurdle or an evaluation tool to decide whether or not to bid at a level to win the business.

**The Efficient Frontier Including Discount**

1. Incorporating the Execution Discount Premium: This section connects the problem of liquidating a basket to the price of a basket. The aim is to eventually compute the information ration of liquidating a basket incorporating the value of the discount to the fair value received in the transaction.
2. Units of the Execution Discount: Thus the assumption is that the trader will receive a discount of dollars per share for a basket in the principal trade, and explicitly calculate the cash received in trading out the basket and its variance. For example, if the program trading desk were to be able to dump the entire portfolio onto the market without any market impact, it would earn a profit of dollars.
3. Execution Profit Under market Impact: In general, because of the market impact, the total expected profit of a trade would be and the variance is the same as stated in other publications (Almgren and Chriss (2000), Almgren (2003)).
4. Execution Profit Expectation and Variance: The expected profit and its variance can be explicitly calculated as functions of the execution time :
5. Market Impact Reduction of Profit: Shortly and will be used to construct the information ratio of a trade, but for now some of its properties are examined. Clearly it is seen that the expected profit of a trade is its total discount reduced by a temporary impact and a permanent impact amount.
6. Difference between the Permanent and the Temporary Impact Costs: The effect of the permanent impact is to reduce the expected profit per share, as reflected in the size of the discount , by an amount equal to the portfolio size, while temporary impact is proportional to a *per share* reduction in the expected profit of .
7. Time Horizon Dependence of Profit: It is worth noting the dependence of and on . Short liquidation times correspond to and All profit is dissipated in impact costs, but no variance is incurred.
8. Intuitive Interpretation of Long Times: Long times correspond to and Temporary impact costs are avoided completely by essentially holding the portfolio forever, but at the expense of any certainty of profit.
9. Elimination of Permanent Impact Costs: In principle, if the portfolio were held forever without trading, then the permanent impact costs will also be avoided; is not the same as
10. The Zero Net Profit Trade: Assuming that the discount is at least enough to compensate for the permanent impact, there is an intermediate point at which - the zero profit trade; impact costs are exactly compensated by the discount on average, but risk is taken to achieve this.

**References**

* Almgren, R. F., and N. Chriss (2000): Optimal Execution of Portfolio Transactions *Journal of Risk* **3 (2)** 5-39.
* Almgren, R. F. (2003): Optimal Executions with Nonlinear Impact Functions and Trading-Enhanced Risk *Applied Mathematical Finance* **10 (1)** 1-18.
* Almgren, R. F., and N. Chriss (2003): Bidding Principals *Risk* **16 (6)** 97-102.
* NYSE (2002): Press Release.

1. Tight Skew: 
2. Loose Skew: 
3. Tight Width: 
4. Loose Width: 
5. Algorithmically generated Ideal Mid Cash Price: 
6. Position: (expressed in cumulative net position per unit under consideration – firm/desk/trader)
7. Position Pivot: . Dimensionless ontological view of the scaling position metric – roughly equivalent to the Reynolds’ number of market making position units. Expressed in currency units.
8. Risk: (expressed in cumulative net risk per unit under consideration – firm/desk/trader)
9. Risk Pivot: . Dimensionless ontological view of the scaling risk metric – roughly equivalent to the Reynolds’ number of market making risk units. Expressed in PV01 currency units.

### Framework Glossary

1. Equilibrium quantity: Quantity that only changes with the macro drivers/factors, and not the technical factors. Typically stable, but jumpy and undergoes changes when drivers shift – and introduces perturbations on the disequilibrium quantities.
2. Disequilibrium quantity: Quantity that changes with the technical, transient factors.

## Width/Skew/Size Estimation Models

1. Tight Models:

* Tight models estimate the market making quantities on a trader/firm/desk independent manner.
* They estimate the “secular” market making parameters – width, skew, and size for either the Market Making Outputs or the Axe Outputs – estimate them based on classes of input parameters.
* For each input parameter class, the following are needed:
  + - * 1. A proxy that serves as a quantitative estimate of the desired parameter class.
        2. Segmentation of the proxy over the sub-classified parameter set.

1. Input Class => Risk Profile:

* Captures all the cumulative risk components => the credit/solvency, market, and liquidity risk behind the issue.
* Proxy => CDS Spread, rating, bond basis
* Sub-classification => Issue, issuer, and sector.

1. Input Class => Liquidity:

* Captures the frequency and volume of the trade flow of a given issue, and the ease of getting in and getting out at the given side.
* Proxy:
  + 1. Aggregated periodic (e.g., daily) volume for each side (buy/sell).
    2. Aggregated periodic (e.g., daily) notional for each side (buy/sell).
* Sub-classification => Issue, issuer, sector, and the instrument universe.

1. Firm/Desk/Trader level parameters: These provide aggregated controls for trading.

* Net Position => vital metric for inventory control.
* Risk limits => to control/manage exposure to specific granules – issue, issuer, tenor, sector, unit etc.

1. Monitor Mobility: Certain measures such as PV01 based risk, inventory, etc. are more easily human-monitored, so they are done daily. Others (such as tenor 01s) are less easily monitored, so they are done infrequently.

## Market Making System SKU

* 1. Intra day Curve Generation Scheme
  2. Mid Price Estimation Models
     1. Accommodate different mid price estimation models, and their respective parameters
  3. Algorithmic Quote Construction => used for generating venue/ECN independent width/skew/size [composed of tight/loose components]. Broadly speaking achieves the following:
     1. Specific parameters to control skew for targeted alpha generation strategies
     2. Accommodate different width and size estimation models, and their respective parameters
     3. Venue-independent base quote synthesis/construction
     4. Circuit breaker heuristics
     5. Policy driven/policy enforcement/policy control applied at this level
  4. Quote Management: Publishing/tailoring the constructed quote towards specific venues (possibly with order routing applied at this stage).
     1. Venue specific rules (and thereby external vendor incorporations, like Broadway etc. at this stage.

## Market Making Parameter Types

1. Model Parameters: Parameters for generation of algorithmic generation of width, skew, and size.
2. Quote Generation Control Parameters
3. Quote Heuristics Control
4. Quote Management Control

## Intra-day Pricing Curve Generation Schemes

1. Issue Benchmark Bonds: The following set of threshold criteria are used to determine the issuer specific benchmark bonds:

* + 1. Threshold of daily TRACE volume/number of trades
    2. Threshold of outstanding notional
    3. Only senior obligations
    4. Some combination of the following threshold of the ratios:
* 
* 

1. Benchmark bonds basis tracking: Track the bid side and ask side credit basis of the benchmark bonds from each TRACE print, using EMA VWAP/TWAP from the intra-day rates/credit curves. This will be the attempt to estimate the mid credit basis for the, and it is generally well behaved.

* Need to find a way to accommodate the institutional closing CDS mid marks and the benchmark bonds into the credit curve construction – these are highly valid points.

1. Liquid vs. illiquid: Typical liquid securities’ quote may be proxied out of print (or at least EMA’d). Intra-day quote generation, however, is materially important for illiquid securities.
2. Intra-day credit curve generation inputs: Need a way to generate the credit curve from

The CDS marks

The basis-adjusted benchmark bonds

It always needs to be used in conjunction with tension splines.

Also need intra-day TRACE series to update the basis (direct or EMA) – will use this to establish the intra-day relationship between the CDS nodes and the TRACE cut-off threshold).

1. Intra-day credit curve updating:
   * + - 1. Use the relationship grid between CDS 5Y, the off-tenors, and the benchmark bonds
         2. Any change in any of them automatically re-adjusts using the set relationships.
         3. CDS Curves are trader set; bond basis are EMA’d from the TRACE series using the prior credit curve
         4. Relationships are either reviewed daily EOD
2. Live updating of bond prices: Use the live curve (either pure CDS, or a mixture of CDS/bond instruments) to extract the basis of each print, and then EMA that to generate the bond live prices.

## Mid Price Models

* + 1. Definition: Computed theoretical mid-price, as to where the next print should be – assuming zero transaction costs, zero position/risk constraints, and infinite liquidity. Mid Price is an ***Equilibrium Quantity***.
    2. Estimation parameters: Typical mid price estimation parameters are: the IR curve, the survival curve, and the recovery curve. The other possible drivers are: funding curve – typically for long position, and repo curve – typically for shorts.

## Width Models

1. Tight Width: Computed theoretical width, after accounting for the issue liquidity and the issue riskiness. Tight width is the first in the set of disequilibrium quantities. Tight width is:
   1. Proportional to issue risk (combination of credit and market risk – not counter party risk).
   2. Inversely proportional to liquidity

## Skew Models

* 1. Tight Skew: This measure how far the last print has been OFF from the theoretical mid price. Thus Tight Skew is representative of the alpha potential – for a theoretical mid price that chases the print in a sequence, the tight skew is zero.
  2. Tight Bid Skew and Tight Ask Skew: This is an alternative SKU – instead of tight width and tight skew cognitive view, tight bid/ask skew parameters are determined only from their corresponding liquidity and flow metrics (i.e., bid/ask liquidity metrics).
  3. Loose Skew: Simply put, loose skew is:



* 1. Heuristic Checks on Loose Skew: Following checks applied to round out quoting:

1. Ceiling/floor applied
2. Maximum cutoff for width
3. Best right skew – bid becomes ask.
4. Best left skew – ask becomes bid.

## Size Models

1. Tight bid size/tight ask size: Basically, tight size is inversely proportional to tight width, to within normalized bounds.

## Heuristics Control

* 1. Can Buy/Can Short: Can But/Can Short => whether the bid/ask stays within the LONGABLE/SHORTABLE cutoff.
  2. ECN Threshold Cross: Check to see if there is a cross between the published bid/ask and a given ECN’s bid/ask.

## Published Market Quote Picture

1. Bid/Ask Sizes: Truncated to their appropriate rounding.
2. Bid Price: 
3. Ask Price: 
4. Bid/Ask Prices rounded downward/upward to their appropriate increments.

## Flow Analysis

1. Dimensionless flow classifier: If the metric (ADV etc.) is greater than a specific threshold, then the flow becomes “turbulent”, else it is “laminar”.
2. Flow Potential: Skew of all kinds is related to the flow driver/equilibration strength.