

# Algorithmic Transaction Cost Analysis

## INTRODUCTION

Transaction cost analysis (TCA) has regained a new found interest in the financial community as a result of the proliferation of algorithmic trading. Portfolio managers and traders are using TCA to evaluate performance of brokers and their algorithms. Furthermore, TCA is used by portfolio managers to improve performance as part of their stock selection and portfolio construction process.

Currently, there are many investors who utilize TCA to select their trading algorithms and make informed trading decisions. Those investors who are not yet utilizing TCA as a decision-making tool are missing valuable opportunities to improve portfolio performance and increase returns.

TCA has evolved significantly over the last several years, though it is still commonly conceptualized as a vague and unstructured concept. The accompanying literature and research still remains muddled due to misrepresentation by many brokers, vendors, and industry participants. We set out to shed new light below.

In order to fully assist investors' algorithmic transaction cost performance, we have developed a framework that consists of pre-, intra-, and post-trade analysis. Our framework is based on an unbundling scheme where costs are classified by ten components and categorized by where they occur during implementation. This scheme is based on the work of Perold (1988) and Wagner and Edwards (1993), and has been described in Journal of Trading, "The Expanded Implementation Shortfall: Understanding Transaction Cost Components," Kissell (2006), and in *Optimal Trading Strategies* (2003). Madhavan (2000, 2002) provides a detailed investigation of financial literature discussing transaction cost components and is considered by many as the gold standard of TCA literature review.

Content based on

## What Are Transaction Costs?

In economics, transaction costs are the fees paid by buyers, but not received by sellers, and/or the fees paid by sellers, but not received by buyers. In finance, transaction costs refers to the premium above the current market price required to attract additional sellers into the market, and the discount below the current market price required to attract additional buyers into the market. Transaction costs are described by Ronald Coase (1937) in “The Nature of the Firm” as an unavoidable cost of doing business. He was subsequently awarded the Economics Nobel Prize in 1991 for his leading edge work.

## What Is Best Execution?

The perception that best execution is an elusive concept has become severely overplayed in the industry. In reality, “best execution” is a very simple and direct concept:

Best execution (as stated in *Optimal Trading Strategies*) is the process of determining the strategy that provides the highest likelihood of achieving the investment objective of the fund. The strategy consists of managing transaction costs during all phases of the investment cycle, and determining when it is appropriate to take advantage of ever-changing market conditions.

Wayne Wagner described best execution in even simpler terms:

*It is the process of maximizing the investment idea.*

Best execution does not depend on how close the execution price occurs to an arbitrary benchmark price (such as the open, close, high, low, VWAP, etc.). Rather, it does depend on the investor’s ability to make proper trading decisions by incorporating all market uncertainties and the current market conditions. The ultimate goal of best execution is to ensure that the trading decisions are consistent with the overall investment objectives of the fund. (See Kissell and Malamut (2007) for a discussion on ensuring consistency between investing and trading consistency.)

To determine whether or not best execution has been met requires the performance evaluation to be made based on the “information set” that was available at the beginning of trading combined with the investment objective of the fund. If either the information set or the underlying investment objective is not known or is not available it is simply not possible to determine if best execution was achieved—regardless of how close the transaction prices were to any benchmark price.

goal of Best Execution

## What Is the Goal of Implementation?

Implementation is the process of determining suitable appropriate trading strategies and adaptation tactics that will result in best execution. Unfortunately, it is not possible for investors to pre-evaluate and determine the best way to execute a position under all possible scenarios, but investors can develop rules and guidelines to make these tasks quicker, easier, and more efficient during trading.

In Wayne Wagner's terminology,

*Implementation is the Journey to Best Execution.*

## UNBUNDLED TRANSACTION COST COMPONENTS

( We have identified ten distinct transaction cost components: commissions, taxes, fees, rebates, spreads, delay cost, price appreciation, market impact, timing risk, and opportunity cost. These are described below following the definitions in Kissell (2003, 2006). )

10 of  
transaction cost components.

### 1. Commission

Commission is payment made to broker-dealers for executing trades and corresponding services such as order routing and risk management. Commissions are commonly expressed on a per share basis (e.g., cents per share) or based on total transaction value (e.g., some basis point of transaction value). Commission charges may vary by:

- i. Broker, fund (based on trading volume), or by trading type (cash, program, algorithms, or DMA).
- ii. Trading difficulty, where easier trades receive a lower rate and the more difficult trades a higher rate. In the current trading arena commissions are highest for cash trading followed by programs, algorithms, and DMA.

### 2. Fees

Fees charged during execution of the order include ticket charges assessed by floor brokers, exchange fees, clearing and settlement costs, and SEC transaction fees. Very often brokers bundle these fees into the total commissions charge.

### 3. Taxes

Taxes are a levy assessed based on realized earnings. Tax rates will vary by investment and type of earning. For example, capital gains, long-term

earnings, dividends, and short-term profits can all be taxed at different percentages.

#### 4. Rebates *maker incentive, depends on Incentive.*

The rebate component is a new transaction cost component that is the byproduct of the new market environment (see Chapter 2). Trading venues charge a usage fee using a straight commission fee structure, a maker-taker model, or a taker-maker (inverted) model. In a straight commission model, both parties are charged a fee for usage of the system. In the maker-taker model, the investor who posts liquidity is provided with a rebate and the investor who takes liquidity is charged a fee. In an inverted or taker-maker model, the investor posting liquidity is charged a fee and the investor who takes liquidity is provided with a rebate. In both cases the fee charged will be higher than the rebate provided to ensure that the trading venue will earn a profit. Brokers may or may not pass this component onto their clients. In the cases when it does not pass through the component the broker will pay the fee or collect the rebate for their own profit pool. The commission rate charged to investors in these cases is likely to already have this fee and/or rebate embedded in its amount.

Since the fee amount or rebate collected is based on the trading venue and whether the algorithm posts or takes liquidity, the selection of trading venue and smart router order logic could be influenced based on the net incremental cost or rebate for the broker rather than the investor. Many questions arise (and rightly so) as to whether or not the broker is really placing orders correctly based on the needs of their investor or are looking to capture and profit from the rebates themselves. Analysts are highly encouraged to inquire about and challenge the logic of rebate-fee payment streams generated by various types of trading algorithms and smart routers in order to **confirm the logic is in their best interest.**

#### 5. Spreads

The spread is the difference between best offer (ask) and best bid price. It is intended to compensate market makers for the risks associated with acquiring and holding an inventory while waiting to offset the position in the market. This cost component is also intended to compensate for the risk potential of adverse selection or transactions with an informed investor (i.e., acquirement of toxic order flow). Spreads represent the round-trip cost of transacting for small orders (e.g., 100 share lots) but do not accurately represent the round-trip cost of transacting blocks (e.g., 10,000 + shares).

## 6. Delay Cost

Delay cost represents the loss in investment value between the time the manager makes the investment decision and the time the order is released to the market. Managers who buy rising stocks and sell falling stocks will incur a delay cost. Delay cost could occur for many reasons.

First, delay cost may arise because traders hesitate in releasing the orders to the market. Second, cost may occur due to uncertainty surrounding who are the most “capable” brokers for the particular order or trade list. Some brokers are more capable at transacting certain names or more capable in certain market conditions. Third, traders may decide to hold off the transaction because they believe better prices may occur. However, if the market moves away, e.g., an adverse momentum, then the delay cost can be quite large. Fourth, traders may unintentionally convey information to the market about their trading intentions and order size (information leakage). Fifth, overnight price change movement may occur. For example, stock price often changes from the close to the open. Investors cannot participate in this price change, so the difference results in a sunk cost or savings depending on whether the change is favorable. Investors who are properly managing all phases of the investment cycle can minimize (if not avoid completely) all delay cost components except for the overnight price movement.

## 7. Price Appreciation

Price appreciation represents how the stock price would evolve in a market without any uncertainty (natural price movement). Price appreciation is also referred to as price trend, drift, momentum, or alpha. It represents the cost (savings) associated with buying stock in a rising (falling) market or selling (buying) stock in a falling (rising) market. Many bond pricing models assume that the value of the bond will appreciate based on the bond’s interest rate and time to maturity.

## 8. Market Impact

Market impact represents the movement in the price of the stock caused by a particular trade or order. It is one of the more costly transaction cost components and always results in adverse price movement and a drag on performance. Market impact will occur due to the liquidity demand (temporary) of the investor and the information content (permanent) of the trade. The liquidity demand cost component refers to the situation where the investors wishing to buy or sell stock in the market have insufficient counterparties to complete the order. In these situations, investors will

have to provide premiums above the current price for buy orders or discount their price for sell orders to attract additional counterparties to complete the transaction. The information content of the trade consists of inadvertently providing the market with signals to indicate the investor's buy/sell intentions, which in turn the market often interprets the stock as under-or overvalued, respectively.

Mathematically, market impact is the difference between the price trajectory of the stock with the order and what the price trajectory would have been had the order not been released to the market. Unfortunately, we are not able to simultaneously observe both price trajectories and measure market impact with any exactness. As a result, market impact has been described as the "Heisenberg uncertainty principle of trading." This concept is further described and illustrated in Chapter 4, Market Impact Models.

### **9. Timing Risk**

Timing risk refers to the uncertainty surrounding the estimated transaction cost. It consists of three components: price volatility, liquidity risk, and parameter estimation error. Price volatility causes the underlying stock price to be either higher or lower than estimated due to market movement and noise. Liquidity risk drives market impact cost due to fluctuations in the number of counterparties in the market. Liquidity risk is dependent upon volumes, intraday trading patterns, as well as the aggregate buying and selling pressure of all market participants. Estimation error is the standard error (uncertainty) surrounding the market impact parameters.

### **10. Opportunity Cost**

Opportunity cost is a measure of the forgone profit or avoided loss of not being able to transact the entire order (e.g., having unexecuted shares at the end of the trading period). The main reasons that opportunity cost may occur are adverse price movement and insufficient liquidity. First, if managers buy stocks that are rising, they may cancel the unexecuted shares of the order as the price becomes too expensive, resulting in a missed profit. Second, if managers cannot complete the order due to insufficient market liquidity (e.g., lack of counterparty participation) the manager would again miss out on a profit opportunity for those unexecuted shares due to favorable price movement.

## **TRANSACTION COST CLASSIFICATION**

Transaction costs can be classified into investment related, trading related, and opportunity cost components shown above.

**Investment-Related Costs** are the costs that arise during the investment decision phase of the investment cycle. They occur from the time of the investment decision to the time the order is released to the market. These costs often arise due to lack of communication between the portfolio manager and trader in deciding the proper implementation objective (strategy), or due to a delay in selecting the appropriate broker or algorithm. The longer it takes for the manager and trader to resolve these issues, the higher potential for adverse price movement and higher investment cost. Traders often spend valuable time investigating how trade lists should be implemented and what broker or trading venue to use. The easiest way to reduce investment-related transaction cost is the use of proper pre-trade analysis, alternative strategy evaluations, and algorithm selections in order for the manager and traders to work closely together to determine the strategy most consistent with the investment objective of the fund.

**Trading-Related Costs.** Trading-related transaction costs comprise the largest subset of transaction costs. They consist of all costs that occur during actual implementation of the order. While these costs cannot be eliminated, they can be properly managed based on the needs of the fund. The largest trading-related transaction costs are market impact and timing risk. However, these two components are conflicting terms and often referred to as the “trader’s dilemma,” as traders need to balance this trade-off based on the risk appetite of the firm. Market impact is highest utilizing an aggressive trading strategy and lowest utilizing a passive strategy. Timing risk, on the other hand, is highest with a passive strategy and lowest with an aggressive strategy. Market impact and timing risk are two conflicting terms.

**Opportunity Cost.** Opportunity cost, as stated above, represents the foregone profit or loss resulting from not being able to fully execute the order within the allotted period of time. It is measured as the number of unexecuted shares multiplied by the price change during which the order was in the market. Opportunity cost will arise either because the trader was unwilling to transact shares at the existing market prices (e.g., prices were too high) or because there was insufficient market liquidity (e.g., not enough sellers for a buy order or buyers for a sell order) or both. The best way to reduce opportunity cost is for managers and traders to work together to determine the number of shares that can be absorbed by the market within the manager’s specified price range. If it is predetermined that the market is not able to absorb all shares of the order within the specified prices, the manager can modify the order to a size that can be easily transacted at their price points.

## TRANSACTION COST CATEGORIZATION

Financial transaction costs are comprised of fixed and variable components and are either visible or hidden (non-transparent).

Fixed cost components are those costs that are not dependent upon the implementation strategy and cannot be managed or reduced during implementation. Variable cost components, on the other hand, vary during implementation of the investment decision and are a function of the underlying implementation strategy. Variable cost components make up the majority of total transaction costs. Money managers, traders, and brokers can add considerable value to the implementation process simply by controlling these variable components in a manner consistent with the overall investment objective of the fund.

Visible or transparent costs are those costs whose fee structure is known in advance. For example, visible costs may be stated as a percentage of traded value, as a \$/share cost applied to total volume traded, or even as some percentage of realized trading profit. Visible cost components are primarily attributable to commissions, fees, spreads, and taxes. Hidden or non-transparent transaction costs are those costs whose fee structure is unknown. For example, the exact cost for a large block order will not be known until after the transaction has been completed (if executed via agency) or until after the bid has been requested (if principal bid). The cost structures for these hidden components are typically estimated using statistical models. For example, market impact costs are often estimated via non-linear regression estimation.

Non-transparent transaction costs comprise the greatest portion of total transaction cost and provide the greatest potential for performance enhancement. Traders and/or algorithms need to be especially conscious of these components in order to add value to the implementation process. If they are not properly controlled they can cause superior investment opportunities to become only marginally profitable and/or profitable opportunities to turn bad. [Table 3.1](#) illustrates our Unbundled Transaction Costs categories. [Table 3.2](#) illustrates our Transaction Cost classification.

## TRANSACTION COST ANALYSIS

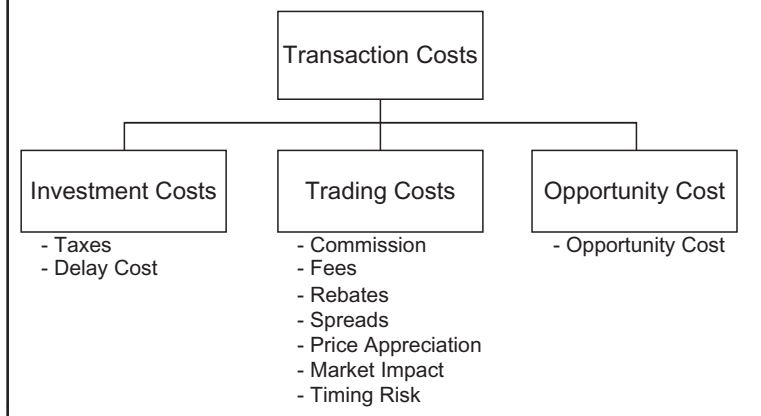
Transaction cost analysis (TCA) is the investor's tool to achieve best execution. It consists of pre-trade, intraday, and post-trade analysis.

Pre-trade analysis occurs prior to the commencement of trading. It consists of forecasting price appreciation, market impact and timing risk for the



**Table 3.1** Unbundled Transaction Costs

	Fixed	Variable
Visible:	Commission Fees Rebates	Spreads Taxes  Delay Cost
Hidden:	n/a	Price Appreciation Market Impact Timing Risk Opportunity

**Table 3.2** Transaction Cost Classification

specified strategy, evaluating alternative strategies and algorithms, and selecting the strategy or algorithm that is most consistent with the overall investment objective of the fund.

Intraday analysis is intended to ensure that the revised execution strategies will continuously be aligned with the high level trading decisions. It consists of specifying how these strategies are to adapt to the endlessly changing market conditions (e.g., price movement and liquidity conditions). The only certainty in trading is that actual conditions will differ from expected. Participants need to understand when it is appropriate to change their strategy and take advantage of these changing market conditions.

Both pre-trade and intraday analysis consist of making and revising execution strategies (in real-time) to ensure trading goals are consistent

with overall investment objectives. Best execution is determined more on decisions made pre-trade than post-trade. Most analysts are very good Monday morning quarterbacks. However, investors need a quality coach who can make and execute decisions under pressure with unknown conditions.

Post-trade analysis, on the other hand, does not consist of making any type of trading decision (either pre-trade or intraday). Post-trade analysis is used to determine whether the pre-trade models give accurate and reasonable expectations, and whether pre-trade and intraday decisions are consistent with the overall investment objectives of the fund. In other words, it is the report card of execution performance.

Post-trade analysis consists of two parts: measuring costs and evaluating performance. All too often, however, there is confusion regarding the meaning of these parts. For example, comparison of the execution price to the VWAP price over the day is not a trading cost—it is a proxy for performance. Comparison to the day's closing price is not a cost—it is a proxy for tracking error. And, comparison of execution price to the opening price on the day or the market price at time of order entry is a cost to the fund and does not give insight into the performance of the trade.

Post-trade analysis needs to provide a measurement of cost, and evaluation of performance at the broker, trader, and algorithm level. When appropriate, the post-trade report should provide universe comparisons, categorization breakdowns (large/small orders, adverse/favorable price movement, high/low volatility, market up/down, etc.) and trend analysis.

### **Measuring/Forecasting**

A cost measure is an “ex-post” or “after the fact” measure, and is determined via a statistical model. It is always a single value and can be either positive (less favorable) or negative (savings). It is computed directly from price data. A cost forecast, on the other hand, occurs “ex-ante” or “prior to trading.” It is an estimated value comprised of a distribution with an expected mean (cost) and standard deviation (timing risk).

The average or mean trading cost component is comprised of market impact and price appreciation. The forecasted market impact estimate will always be positive and indicate less favorable transaction prices. The price appreciation component, on the other hand, could be zero (e.g., no expectation of price movement), positive, indicating adverse price movement and less favorable expected transaction prices, or negative, indicating favorable

price momentum and better transaction prices. The trading cost standard error term is comprised of price volatility, liquidity risk, and parameter estimation error from the market impact model.

## Cost versus Profit and Loss

There is not much consistency in the industry regarding the terminology or sign to use when measuring and forecasting costs. Many participants state cost as a positive value while others state cost as a negative value. For example, some participants refer to a positive cost of +30 bp as underperformance and a negative cost of −30 bp as outperformance (savings). Others treat this metric in the opposite way with the +30 bp indicating better transaction prices and −30 bp indicating worse transaction prices.

To avoid potential confusion, our “Cost” and “Profit and Loss” terminology throughout the text will be as follows:

A “**Cost**” metric will always use a positive value to indicate underperformance and a negative value to indicate better performance. For example, a cost of 30 bp indicates less favorable execution than the benchmark and −30 bp cost indicates better performance than the benchmark.

A “**Profit and Loss**” or “**PnL**” metric will always use a negative value to indicate underperformance and a positive value to indicate better performance. For example, a PnL of −5 bp indicates less favorable execution than the benchmark and a PnL of +5 bp indicates better performance compared to the benchmark.

## IMPLEMENTATION SHORTFALL

Implementation shortfall (IS) is a measure that represents the total cost of executing the investment idea. It was introduced by Perold (1988) and is calculated as the difference between the paper return of a portfolio where all shares are assumed to have transacted at the manager’s decision price and the actual return of the portfolio using actual transaction prices and shares executed. It is often described as the missed profiting opportunity as well as the friction associated with executing the trade. Many industry participants refer to implementation shortfall as slippage or simply portfolio cost.

Mathematically, implementation shortfall is written as:

$$IS = \text{Paper Return} - \text{Actual Return} \quad (3.1)$$

**Paper return** is the difference between the ending portfolio value and its starting value evaluated at the manager's decision price. This is:

$$\text{Paper Return} = S \cdot P_n - S \cdot P_d \quad (3.2)$$

Here  $S$  represents the total number of shares to trade,  $P_d$  is the manager's decision price, and  $P_n$  is the price at the end of period  $n$ .  $S \cdot P_d$  represents the starting portfolio value and  $S \cdot P_n$  represents the ending portfolio value. Notice that the formulation of the paper return does not include any transaction costs such as commissions, ticket charges, etc. The paper return is meant to capture the full potential of the manager's stock picking ability. For example, suppose a manager decides to purchase 5000 shares of a stock trading at \$10 and by the end of the day the stock is trading at \$11. The value of the portfolio at the time of the investment decision was \$50,000 and the value of the portfolio at the end of the day was \$55,000. Therefore, the paper return of this investment idea is \$5000.

**Actual portfolio return** is the difference between the actual ending portfolio value and the value that was required to acquire the portfolio minus all fees corresponding to the transaction. Mathematically, this is:

$$\text{Actual Portfolio Return} = \left( \sum s_j \right) \cdot P_n - \sum s_j p_j - \text{fees} \quad (3.3)$$

where,

$(\sum s_j)$  represents the total number of shares in the portfolio

$(\sum s_j) \cdot P_n$  is the ending portfolio value

$\sum s_j p_j$  is the price paid to acquire the portfolio

and fees represent the fixed fees required to facilitate the trade such as commission, taxes, clearing and settlement charges, ticket charges, rebates, etc.  $s_j$  and  $p_j$  represent the shares and price corresponding to the  $j^{\text{th}}$  transaction.

For example, suppose a manager decides to purchase 5000 shares of stock trading at \$10. However, due to market impact, price appreciation, etc., the average transaction price of the order was \$10.50 indicating that the manager invested \$52,500 into the portfolio. If the stock price at the end of the day is \$11 the portfolio value is then worth \$55,000. If the total fees were \$100, then the actual portfolio return is  $\$55,000 - \$52,500 - \$100 = \$2400$ .

Implementation shortfall is then computed as the difference between paper return and portfolio return as follows:

$$IS = \underbrace{S \cdot P_n - S \cdot P_d}_{\text{Paper Return}} - \underbrace{\left( \sum s_j \right) \cdot P_n - \sum s_j p_j - fees}_{\text{Actual Portfolio Return}} \quad (3.4)$$

In our example above, the implementation shortfall for the order is:

$$IS = \$5000 - \$2400 = \$2600$$

The implementation shortfall metric is a very important portfolio manager and trader decision making metric. It is used to select stock picking ability, measure trading costs, and as we show below, measure broker and algorithmic performance.

Implementation shortfall can be described in terms of the following three examples:

1. Complete Execution
2. Opportunity Cost (Andre Perold)
3. Expanded Implementation Shortfall (Wayne Wagner)

## Complete Execution

Complete execution refers to the situation where the entire order is transacted in the market. That is  $\sum s_j = S$ . Suppose a manager decides to purchase  $S$  shares of stock that is currently trading at  $P_d$  and at the end of the trading horizon the price is  $P_n$ . Then implementation shortfall is computed following the above calculation as follows:

$$IS = (S \cdot P_n - S \cdot P_d) - \left( \left( \sum s_j \right) \cdot P_n - \sum s_j p_j - fees \right)$$

Since  $\sum s_j = S$  this equation reduces to:

$$IS = \sum s_j p_j - S \cdot P_d + fees$$

This could also be written in terms of the average execution price  $P_{avg}$  for all shares as follows:

$$IS = S \cdot P_{avg} - S \cdot P_d + fees = S \cdot (P_{avg} - P_d) + fees$$

since  $\sum s_j p_j = S \cdot P_{avg}$ . Notice that when all shares are executed the implementation shortfall measure does not depend on the future stock price  $P_n$  at all.

Example: A manager decided to purchase 5000 shares when the stock was at \$10. All 5000 shares were transacted in the market, but at an average transaction price of \$10.50. If the commission fee was \$100 then implementation shortfall of the order is:

$$IS = 5000 \cdot (\$10.50 - \$10.00) + \$100 = \$2600$$

### Opportunity Cost (Andre Perold)

The opportunity cost example refers to a situation where the manager does not transact the entire order. This could be due to prices becoming too expensive or simply a lack of market liquidity. Either way, it is essential that we account for all unexecuted shares in the implementation shortfall calculation. This process is as follows:

First, compute the paper portfolio return:

$$\text{Paper Return} = S \cdot P_n - S \cdot P_d$$

Next, compute the actual portfolio return for those shares that were executed:

$$\text{Actual Return} = \left( \sum s_j \right) P_n - \sum s_j p_j + \text{fees}$$

Then, the implementation shortfall is written as:

$$IS = (S \cdot P_n - S \cdot P_d) - \left( \left( \sum s_j \right) P_n - \sum s_j p_j + \text{fees} \right)$$

Let us now expand on this formulation. Share quantity  $S$  can be rewritten in terms of executed shares  $\sum s_j$  and unexecuted shares  $(S - \sum s_j)$  as follows:

$$S = \underbrace{\sum s_j}_{\text{Executed}} + \underbrace{\left( S - \sum s_j \right)}_{\text{Unexecuted}}$$

If we substitute the share quantity expression above into the previous IS formulation we have:

$$IS = \left( \sum s_j + \left( S - \sum s_j \right) \cdot P_n - \sum s_j + \left( S - \sum s_j \right) \cdot P_d \right) - \left( \left( \sum s_j \right) P_n - \sum s_j p_j + \text{fees} \right)$$

This equation can be written as:

$$IS = \underbrace{\sum s_j p_j - \sum s_j P_d}_{\text{Execution Cost}} + \underbrace{\left( S - \sum s_j \right) \cdot (P_n - P_d)}_{\text{Opportunity Cost}} + \text{fees}$$

This is the implementation shortfall formulation of Perold (1988) and differentiates between execution cost and opportunity cost. The execution cost component represents the cost that is incurred in the market during trading. Opportunity cost represents the missed profiting opportunity by not being able to transact all shares at the decision price.

Example: A manager decides to purchase 5000 shares of a stock at \$10 but the manager is only able to execute 4000 shares at an average price of \$10.50. The stock price at the end of trading is \$11.00. And the commission cost is \$80, which is reasonable since only 4000 shares traded in this example compared to 5000 shares in the above example. Then implementation shortfall including opportunity cost is:

$$IS = \left( \sum s_j p_j - \sum s_j P_d \right) + \left( S - \sum s_j \right) \cdot (P_n - P_d) + fees$$

It is important to note that in a situation where there are unexecuted shares then the IS formulation does depend upon the ending period stock price  $P_n$  but in a situation where all shares do execute then the IS formulation does not depend upon the ending period price  $P_n$ .

Furthermore, in situations where we have the average execution price of the order, IS further simplifies to:

$$IS = \sum s_j \cdot (P_{avg} - P_d) + \left( S - \sum s_j \right) \cdot (P_n - P_d) + fees$$

In our example we have:

$$\begin{aligned} IS &= 4000 \cdot (\$10.50 - \$10.00) + 1000 \cdot (\$11.00 - \$10.00) + \$80 \\ &= \$2000 + \$1000 + \$80 = \$3080 \end{aligned}$$

The breakdown of costs following Perold is: execution cost = \$2000, opportunity cost = \$1000, and fixed fee = \$80.

### Expanded Implementation Shortfall (Wayne Wagner)

Our third example shows how to decompose implementation shortfall based on where the costs occur in the investment cycle. It starts with opportunity cost, and further segments the cost into a delay component which represents the missed opportunity of being unable to release the order into the market at the time of the investment decision. “Expanded Implementation Shortfall” is based on the work of Wayne Wagner and is often described as Wagner’s Implementation Shortfall. This measurement provides managers with valuable insight into “who” is responsible for which costs. It helps us understand whether the incremental cost was

due to a delay in releasing the order to the market or due to inferior performance by the trader or by the algorithm. Knowing who is responsible for cost will help investors improve the process of lowering transaction costs going forward. Wagner's expanded implementation shortfall categorizes cost into delay, trading, and opportunity related cost. Perold's original formulation did not separate delay and trading related costs when they occurred during the implementation phase. Wagner's formulation of implementation shortfall is what makes it possible to measure performance across traders, brokers, and algorithms.

The derivation of the expanded implementation shortfall is as follows.

First, define two time horizons: investment and trading. The investment horizon is the time from the investment decision  $t_d$  to beginning of trading  $t_0$ . The trading horizon is the time from beginning of trading  $t_0$  to the end of trading  $t_n$ . The corresponding prices at these time intervals are  $P_d$ , which is the decision price,  $P_0$  which is the price at beginning of trading, also known as the arrival price, and  $P_n$ , which is the price at the end of trading. All prices are taken as the mid-point of the bid-ask spread if during market hours or the last traded price or official close if after hours.

Next, rewrite the price change over these two intervals as follows:

$$(P_n - P_d) = (P_n - P_0) + (P_0 - P_d)$$

Now substitute this price into Perold's implementation shortfall:

$$IS = \left( \sum s_j p_j - \sum s_j P_d \right) + \left( S - \sum s_j \right) \cdot (P_n - P_d) + fees$$

This is:

$$IS = \left( \sum s_j p_j - \sum s_j P_d \right) + \left( S - \sum s_j \right) \cdot ((P_n - P_0) + (P_0 - P_d)) + fees$$

This expression can then be written based on our investment and trading horizons and is known as the Expanded Implementation Shortfall or Wagner's Implementation Shortfall. This is as follows:

$$Expanded IS = \underbrace{S(P_0 - P_d)}_{\text{Delay Related}} + \underbrace{\sum s_j p_j - \left( \sum s_j \right) P_0}_{\text{Trading Related}} + \underbrace{\left( S - \sum s_j \right) (P_n - P_0)}_{\text{Opportunity Cost}} + fees$$



This could also be written in terms of the average transaction price  $P_{avg}$  as follows:

$$Expanded\ IS = \underbrace{S(P_0 - P_d)}_{\text{Delay Related}} + \underbrace{\left(\sum s_j\right)(P_{avg} - P_0)}_{\text{Trading Related}} + \underbrace{\left(S - \sum s_j\right)(P_n - P_0)}_{\text{Opportunity Cost}} + \text{fees}$$

This is the expanded implementation shortfall metric proposed by Wayne Wagner that makes a distinction between the investment and trading horizons. It was first identified in Wagner (1975) and later explained in Wagner (1991) and Wagner and Edwards (1993). The delay related component has also been referred to as the investment related cost. The delay cost component could be caused by the portfolio manager, buy-side trader, or broker-dealer. For example, see Almgren and Chriss (2000), Kissell and Glantz (2003), or Rakhlin and Sofianos (2006).

Example: A manager decides to purchase 5000 shares of a stock at \$10. By the time the order is finally released to the market the stock price has increased to \$10.25. If the manager is only able to execute 4000 shares at an average price of \$10.50 and the stock price at the end of trading is \$11.00 what is the expanded implementation shortfall cost by components? Assume total commission cost is \$80.

The calculation of the expanded implementation shortfall is:

$$\begin{aligned} Expanded\ IS = & \underbrace{5000 \cdot (\$10.25 - \$10.00)}_{\text{Delay Related}} + \underbrace{4000 \cdot (\$10.50 - \$10.25)}_{\text{Trading Related}} \\ & + \underbrace{1000 \cdot (\$11.00 - \$10.25)}_{\text{Opportunity Cost}} + \$80 = \$3080 \end{aligned}$$

The delay related component is: \$1250

The trading related component is: \$1000

The opportunity cost component is: \$750

Fixed fee amount is: \$80

Total expanded implementation shortfall = \$3080.

Notice that Wagner's expanded IS cost is the same value as Perold's IS. However, the opportunity cost in this example is \$750 compared to \$1000 previously. The reason for this difference is that the expanded IS measures opportunity cost from the time the order was released to the market as opposed to the time of the manager's decision. In actuality, the delay related

cost component above can be further segmented into a trading related delay cost and an opportunity related delay cost. This is shown as follows:

$$Delay\ Cost = S \cdot (P_0 - P_d) = \underbrace{\left(S - \sum s_j\right) \cdot (P_0 - P_d)}_{Opportunity\ Related\ Delay} + \underbrace{\left(\sum s_j\right)(P_0 - P_d)}_{Trading\ Related\ Delay}$$

Analysts may wish to include all unexecuted shares in the opportunity cost component as a full measure of missed profitability.

It is important to point out that in many cases the analysts will not have the exact decision price of the manager since portfolio managers tend to keep their decision prices and reasons for the investment to themselves. However, analysts know the time the order was released to the market. Hence, the expanded implementation shortfall would follow our formulation above where we only analyze costs during market activity, that is, from  $t_0$  to  $t_n$ . This is:

$$Market\ Activity\ IS = \underbrace{\left(\sum s_j\right)(P_{avg} - P_0)}_{Trading\ Related} + \underbrace{\left(S - \sum s_j\right)(P_n - P_0)}_{Opportunity\ Cost} + fees$$

### **Implementation Shortfall Formulation**

The different formulations of implementation shortfall discussed above are:

$$IS = S \cdot (P_{avg} - P_d) + fees \quad (3.5)$$

$$Perold\ IS = \sum s_j \cdot (P_{avg} - P_d) + \left(S - \sum s_j\right) \cdot (P_n - P_d) + fees \quad (3.6)$$

$$Wagner\ IS = S(P_0 - P_d) + \left(\sum s_j\right)(P_{avg} - P_0) + \left(S - \sum s_j\right)(P_n - P_0) + fees \quad (3.7)$$

$$Mkt\ Act.\ IS = \left(\sum s_j\right)(P_{avg} - P_0) + \left(S - \sum s_j\right)(P_n - P_0) + fees \quad (3.8)$$

### **Trading Cost/Arrival Cost**

The trading cost component is measured as the difference between the average execution price and the price of the stock at the time the order was entered into the market (arrival price). It is the most important metric to evaluate broker, venue, trader, or algorithmic performance, because it quantifies the cost that is directly attributable to trading and these specific parties. It follows directly from the trading related cost component from the expanded implementation shortfall. The investment related and

opportunity cost components are more attributable to investment managers than to the trading party.

The trading cost or arrival cost component is:

$$Arrival\ Cost_{\$} = \sum s_j p_j - \left( \sum s_j \right) P_0 \quad (3.9)$$

$S, s_j > 0$  for buys

$S, s_j < 0$  for sells

In basis points this expression is:

$$Arrival\ Cost_{bp} = \frac{\sum s_j p_j - (\sum s_j) P_0}{(\sum s_j) P_0} \cdot 10^4_{bp} \quad (3.10)$$

In general, arrival costs can be simplified as follows:

$$Arrival\ Cost_{bp} = Side \cdot \frac{P_{avg} - P_0}{P_0} \cdot 10^4_{bp} \quad (3.11)$$

where,

$$Side = \begin{cases} 1 & \text{if Buy} \\ -1 & \text{if Sell} \end{cases}$$

## EVALUATING PERFORMANCE

In this section we describe various techniques to evaluate performance (note: we will use the profit and loss (PnL) terminology). These methods can be used to evaluate and compare trade quality for a single stock or basket of trades, as well as performance across traders, brokers, or algorithms. It can also serve as the basis for universe comparisons. In the following section we provide non-parametric statistical techniques that are being used to compare algorithmic performance.

Techniques that will be discussed in this section include: market or index adjusted cost, benchmark comparisons, various volume weighted average price (VWAP), participation weighted average price (PWP), relative performance measure (RPM), and z-score statistical measures.

### Trading Price Performance

Trading price performance or simply trading PnL is identical to the trading cost component above and is measured as the difference between

the average execution price and the price of the stock at the time the order was entered into the market (arrival price). A positive value indicates more favorable transaction prices and a negative value indicates less favorable transaction prices. Trading PnL is a measure of the cost during trading and reports whether the investor did better or worse than the arrival price. For example, a trading PnL of  $-10$  bp indicates the fund underperformed the arrival price benchmark by 10 bp. The formulation for trading PnL multiplies the arrival cost calculation above by minus 1. This is:

$$\text{Trading PnL}_{bp} = -1 \cdot \text{Side} \cdot \frac{P_{avg} - P_0}{P_0} \cdot 10^4_{bp} \quad (3.12)$$

### Benchmark Price Performance

Benchmark price performance measures are the simplest of the TCA performance evaluation techniques. These are intended to compare specific measures such as net difference and tracking error, or to distinguish between temporary and permanent impact. Some of the more commonly used benchmark prices include:

- Open—as a proxy for arrival price.
- Close—insight into end-of-day tracking error and is more commonly used by index funds that use the closing price in valuation of the fund.
- Next Day Open—as a way to distinguish between temporary and permanent market impact.
- Next Day Close or Future Day Close—also to distinguish between temporary and permanent impact.

The benchmark PnL calculation is:

$$\text{Trading PnL}_{bp} = -1 \cdot \text{Side} \cdot \frac{P_{avg} - P_B}{P_B} \cdot 10^4_{bp} \quad (3.13)$$

where  $P_B = \text{benchmark price}$ .

### VWAP Benchmark

The VWAP benchmark is used as a proxy for fair market price. It helps investors determine if their execution prices were in line and consistent with fair market prices.

The calculation is:

$$\text{VWAP PnL}_{bp} = -1 \cdot \text{Side} \cdot \frac{P_{avg} - \text{VWAP}}{\text{VWAP}} \cdot 10^4_{bp} \quad (3.14)$$

where VWAP is the volume weighted average price over the trading period. A positive value indicates better performance and a negative value indicates underperformance.

Interval VWAP comparison serves as a good measure of execution quality and does a nice job of accounting for actual market conditions, trading activity, and market movement. The interval VWAP, however, does suffer from three issues. First, the larger the order the closer the results will be to the VWAP price, as the order price will become the VWAP price. Second, actual performance can become skewed if there are large block trades that occur at extreme prices (highs or lows) in crossing venues, especially in cases where investors have limited opportunity to participate with those trades. Third, the VWAP measure does not allow easy comparison across stocks or across the same stock on different days. For example, it is not possible to determine if missing VWAP by 3 bps in one stock is better performance than missing VWAP by 10 bps in another stock. If the first stock has very low volatility and the second stock has very high volatility, missing VWAP by 10 bps in the second name may in fact be better performance than missing VWAP by 3 bps in the first name.

There are three different VWAP performance metrics used: full day, interval, and VWAP to end of day.

**Full Day VWAP:** Used for investors who traded over the entire trading day from open to close. There is currently no “official” VWAP price on the day but many different providers, such as Bloomberg, Reuters, etc., do offer one. These vendors determine exactly what trades will be included in the VWAP calculations but they may not use all the market trades. For example, some providers may filter trades that were delayed or negotiated because they do not feel these prices are indicative of what all market participants had fair access to.

**Interval VWAP:** Used as a proxy for the fair market price during the time the investor was in the market trading. The interval VWAP is a specific VWAP price for the investor over their specific trading horizon and needs to be computed from tic data. This is in comparison to a full day VWAP price that is published by many vendors.

**VWAP to End of Day:** Used to evaluate those orders that were completed before the end of the day. In these cases, the broker or trader made a conscious decision to finish the trade before the end of the day. This VWAP to End of Day provides some insight into what the fair market price was including even after the order was completed. It helps determine if the decision to finish the order early was appropriate. This is a very

useful metric to evaluate over time to determine if the trader or broker is skilled at market timing. But it does require a sufficient number of observations and a large data set.

It is worth noting that some B/Ds and vendors refer to the VWAP comparison as a cost rather than a gain/loss or performance indication. For those parties, a positive value indicates a higher cost (thus underperformance) and a negative value indicates a lower cost (thus better performance) and is the complete opposite of the meaning in the formula above. Unfortunately, representation of costs, P/L, or G/L as a metric is not consistent across industry participants and investors need to be aware of these differences.

### Participation Weighted Price (PWP) Benchmark

Participation weighted price (PWP) is a variation of the VWAP analysis. It is intended to provide a comparison of the average execution price to the likely realized price had they participated with a specified percentage of volume during the duration of the order.

For example, if the PWP benchmark is a 20% POV rate and the investor transacted 100,000 shares in the market starting at 10 a.m. the PWP-20% benchmark price is computed as the volume weighted average price of the first 500,000 shares that traded in the market starting at 10 a.m. (the arrival time of the order). It is easy to see that if the investor transacted at a 20% POV rate their order would have been completed once 500,000 shares traded in the market since  $0.20 \times 500,000 = 100,000$  shares. The number of shares in a PWP analysis is equal to the number of shares traded divided by the specified POV rate.

The PWP PnL metric is computed as follows:

$$PWP \text{ Shares} = \frac{\text{Shares Traded}}{POV \text{ Rate}} \quad (3.15)$$

*PWP Price = volume weighted price of the first PWP shares starting at the arrival time  $t_0$*

$$PWP \text{ PnL}_{bp} = -1 \cdot \frac{P_{avg} - PWP \text{ Price}}{PWP \text{ Price}} \cdot 10^4_{bp} \quad (3.16)$$

The PWP benchmark also has some inherent limitations similar to the VWAP metric. First, while PWP does provide insight into fair and reasonable prices during a specified time horizon it does not allow easy comparison across stocks or across days due to different stock price volatility and

daily price movement. Furthermore, investors could potentially manipulate the PWP by trading at a more aggressive rate to push the price up for buy orders or down for sell orders, and give the market the impression that they still have more to trade. Since temporary impact does not dissipate instantaneously, the PWP price computed over a slightly longer horizon could remain artificially high (buy orders) or artificially low (sell orders) due to temporary impact cost. Participants may hold prices at these artificially higher or lower levels waiting for the non-existent orders to arrive. The end result is a PWP price that is more advantageous to the investor than what would have occurred in the market if the order had actually traded over that horizon.

### Relative Performance Measure (RPM)

The relative performance measure (RPM) is a percentile ranking of trading activity. It provides an indication of the percentage of total activity that the investor outperformed in the market. For a buy order, it represents the percentage of market activity that transacted at a higher price and for a sell order it represents the percentage of market activity that transacted at a lower price. The RPM is modeled after the percentile ranking used in standardized academic tests and provides a descriptive statistic that is more consistent and robust than other measures.

The RPM was originally presented in *Optimal Trading Strategies* (2003) and Kissell (2007) and was based on a volume and trade metric. That original formulation, however, had at times small sample size and large trade percentage limitations bias. For example, the original formulation considered all of the investor's trades at the average transaction price as outperformance. Therefore, in situations where the investor transacted a large size at a single price all the shares were considered as outperformance and the end result would overstate the actual performance. Leslie Boni (2009) further elaborates on this point in her article "Grading Broker Algorithms," *Journal of Trading*, Fall 2009, and provides some important insight and improvements.

To help address these limitations, we revised the RPM formulation as follows:

The RPM is computed based on trading volume as follows:

$$\begin{aligned} \text{RPM} = & \frac{1}{2} \cdot ((\% \text{ of volume traded at a price less favorable or equal to } P_{avg}) \\ & + (1 - \% \text{ of volume traded at a price less favorable or equal to } P_{avg})) \end{aligned} \quad (3.17)$$

This metric can also be formulated for buy and sell orders as follows:

$$RPM_{Buy} = \frac{1}{2} \cdot \left( \frac{\text{Total Volume} + \text{Volume at Price} > P_{avg} - \text{Volume at Price} < P_{avg}}{\text{Total Volume}} \right) \quad (3.18)$$

$$RPM_{Sell} = \frac{1}{2} \cdot \left( \frac{\text{Total Volume} + \text{Volume at Price} < P_{avg} - \text{Volume at Price} > P_{avg}}{\text{Total Volume}} \right) \quad (3.19)$$

This formulation of RPM is now the average of the percentage of volume that traded at our execution price or better and 1 minus the average of the percentage of volume that traded at our price or worse. Thus, in effect, it treats half of the investor's orders as better performance and half the order as worse performance. As stated, the original formulation treated all of the investor's shares as better performance and inflated the measure.

The RPM is in many effects a preferred measure to the VWAP metric because it can be used to compare performance across stocks, days, and volatility conditions. And it is not influenced to the same extent as VWAP when large blocks trade at extreme prices.

The RPM will converge to 50% as the investor accounts for all market volume in the stock on the day similar to how the VWAP converges to the average execution price for large orders.

Brokers achieving fair and reasonable prices on behalf of their investors should achieve an RPM score around 50%. RPM scores consistently greater than 50% are an indication of superior performance and scores consistently less than 50% are an indication of inferior performance. The RPM measure can also be mapped to a qualitative score, for example:

RPM	Quality
0–20%	Poor
20–40%	Fair
40–60%	Average
60–80%	Good
80–100%	Excellent

### Pre-Trade Benchmark

The pre-trade benchmark is used to evaluate trading performance from the perspective of what was expected to have occurred. Investors compute the difference between actual and estimated to determine whether performance was reasonable based on how close they came to the expectation.



Actual results that are much better than estimated could be an indication of skilled and quality execution, whereas actual results that are much worse than estimated could be an indication of inferior execution quality.

The difference between actual and estimated, however, could also be due to actual market conditions during trading that are beyond the control of the trader—such as sudden price momentum, or increased or decreased liquidity conditions. (These are addressed below through the use of the z-score and market adjusted cost analysis.)

The pre-trade performance benchmark is computed as follows:

$$\text{Pre-Trade Difference} = \text{Estimated Arrival Cost} - \text{Actual Arrival Cost} \quad (3.20)$$

A positive value indicates better performance and a negative value indicates worse performance.

Since actual market conditions could have a huge influence on actual costs, some investors have started analyzing the pre-trade difference by computing the estimated market impact cost for the exact market conditions—an ex-post market impact metric. While this type of measure gives reasonable insight in times of higher and lower volumes, on its own it does not give an adequate adjustment for price trend. Thus investors also factor out price trend via a market adjusted performance measure.

## Index Adjusted Performance Metric

A market adjusted or index adjusted performance measure is intended to account for price movement in the stock due to the market, sector, or industry movement. This is computed using the stock's sensitivity to the underlying index and the actual movement of that index as a proxy for the natural price appreciation of the stock (e.g., how the stock price would have changed if the order was not released to the market).

First compute the index movement over the time trading horizon:

$$\text{Index Cost}_{bp} = \frac{\text{Index VWAP} - \text{Index Arrival Cost}}{\text{Index Arrival Cost}} \cdot 10^4_{bp} \quad (3.21)$$

Index arrival is the value of the index at the time the order was released to the market. Index VWAP is the volume weighted average price for the index over the trading horizon. What is the index volume weighted price over a period? Luckily there are many ETFs that serve as proxies for various underlying indexes such as the market (e.g., SPY), or sectors, etc., and thus provide easy availability to data to compute volume weighted average index prices.

If the investor's trade schedule sequence followed a different weighting scheme than volume weighting, such as front-or back-loaded weightings, it would be prudent for investors to compute the index cost in each period. In times where the index VWAP is not available, it can be approximated as  $\text{Index VWAP} = 1/2 \cdot R_m$ , where  $R_m$  is the total return in basis points of the underlying index over the period. The  $1/2$  is the adjustment factor to account for continuous trading Kissell (2008).

The index adjusted cost is then:

$$\text{Index Adjusted Cost}_{bp} = \text{Arrival Cost}_{bp} - \hat{b}_{KI} \cdot \text{Index Cost}_{bp} \quad (3.22)$$

$\hat{b}_{KI}$  is the stock k's sensitivity to the index. It is determined via linear regression in the same manner we calculate beta to the market index. Notice all we have done is subtract out the movement in the stock price that we would have expected to occur based only on the index movement. The index cost is not adjusted for the side of the trade.

### Z-Score Evaluation Metric

The z-score evaluation metric provides a risk adjusted performance score by normalizing the difference between estimated and actual by the timing risk of the execution. This provides a normalized score that can be compared across different stocks and across days. (A z-score measure is also used to measure the accuracy of pre-trade models and to determine if these models are providing reasonable insight to potential outcomes cost.)

A simple statistical z-score is calculated as follows:

$$Z = \frac{\text{Actual} - \text{Expected}}{\text{Standard Deviation}}$$

For transaction cost analysis, we compute the normalized transaction cost as follows:

$$Z = \frac{\text{Pre-Trade Cost Estimate} - \text{Arrival Cost}}{\text{Pre-Trade Timing Risk}} \quad (3.23)$$

Notice that this representation is opposite the statistical z-score measure ( $z = (x-u)/\text{sigma}$ ). In our representation a positive z-score implies performance better than the estimate and a negative value implies performance worse than the estimate. Dividing by the timing risk of the trade normalizes for overall uncertainty due to price volatility and liquidity risk. This ensures that the sign of our performance metrics are consistent—positive indicates better performance and negative indicates lower quality performance.

If the pre-trade estimates are accurate, then the z-score statistic should be a random variable with mean zero and variance equal to one. That is,  $Z \sim (0, 1)$ . There are various statistical tests that can be used to test this joint hypothesis.

There are several points worth mentioning with regards to trading cost comparison. First, the test needs to be carried out for various order sizes (e.g., large, small, and mid-size orders). It is possible for a model to overestimate costs for large orders and underestimate costs for small orders (or vice versa) and still result in  $Z \sim (0, 1)$  on average. Second, the test needs to be carried out for various strategies. Investors need to have a degree of confidence regarding the accuracy of cost estimates for all of the broker strategies. Third, it is essential that the pre-trade cost estimate be based on the number of shares traded and not the full order. Otherwise, the pre-trade cost estimate will likely overstate the cost of the trade and the broker being measured will consistently outperform the benchmark giving the appearance of superior performance and broker ability. In times where the order was not completely executed, the pre-trade cost estimates need to be adjusted to reflect the actual number of shares traded. Finally, analysts need to evaluate a large enough sample size in order to achieve statistical confidence surrounding the results, as well as conduct cross-sectional analysis in order to uncover any potential bias based on size, volatility, market capitalization, and market movement (e.g., up days and down days).

It is also important to note that many investors are using their own pre-trade estimates when computing the z-score measure. There is a widespread resistance to using a broker's derived pre-trade estimate to evaluate their own performance. As one manager stated, everyone looks great when we compare their performance to their cost estimate. But things start to fall into place when we use our own pre-trade estimate. Pre-trade cost comparison needs to be performed using a standard pre-trade model to avoid any bias that may occur with using the provider's own performance evaluation model.

## Market Cost Adjusted Z-Score

It is possible to compute a z-score for the market adjusted cost as a means of normalizing performance and comparing across various sizes, strategies, and time periods similar to how it is used with the trading cost metric. But in this case, the denominator of the z-score is not the timing risk of the trade since timing risk accounts in part for the uncertainty in total price movement (adjusted for the trade schedule). The divisor in this case has to be the tracking error of the stock to the underlying index (adjusted for

the trading strategy). Here the tracking error is identical to the standard deviation of the regression equation:

$$\text{Index Adj Cost} = \text{Arrival Cost} - \hat{b}_{kl} \cdot \text{Index Cost} + \varepsilon$$

Where the adjusted tracking error to the index is  $\sqrt{\sigma_\varepsilon^2}$

Here we subtract only estimated market impact cost (not total estimated cost) for the market adjusted cost since we already adjusted for price appreciation using the stock's underlying beta and index as its proxy.

$$\text{Mkt Adj Z-Score} = \frac{\text{Pre-Trade Estimate} - \text{Mkt Adj Cost}}{\text{Adj Tracking Error to the Index}} \quad (3.24)$$

### Adaptation Tactic

Investors also need to evaluate any adaptation decisions employed during trading to determine if traders correctly specify these tactics and to ensure consistency with the investment objectives. For example, many times investors instruct brokers to spread the trades over the course of the day to minimize market impact cost, but if favorable trading opportunities exist then trading should accelerate to take advantage of the opportunity. Additionally, some instructions are to execute over a predefined period of time, such as the next two hours, but with some freedom. In these situations, brokers have the opportunity to finish earlier if favorable conditions exist, or extend the trading period if they believe the better opportunities will occur later in the day.

The main goal of evaluating adaptation tactics is to determine if the adaptation decision (e.g., deviation from initial strategy) was appropriate given the actual market conditions (prices and liquidity). That is, how good a job does the broker do in anticipating intraday trading patterns and favorable trading opportunities.

The easiest way to evaluate adaptation performance is to perform the interval VWAP and interval RPM analyses (see above) over the time period specified by the investor (e.g., a full day or for the specified two hour period) instead of the trading horizon of the trade. This will allow us to determine if the broker actually realized better prices by deviating from the initially prescribed schedule and will help distinguish between skill and luck.

As with all statistical analyses, it is important to have a statistically significant sample size and also perform cross-sectional studies where data points are grouped by size, side, volatility, market capitalization, and market

movement (e.g., up days and down days) in order to determine if there is any bias for certain conditions or trading characteristics (e.g., one broker or algorithm performs better for high volatility stocks, another broker or algorithm performs better in favorable trending markets, etc.).

## COMPARING ALGORITHMS

One of the biggest obstacles in comparing algorithmic performance is that each algorithm trades in a different manner, under a different set of market conditions. For example, a VWAP algorithm trades in a passive manner with lower cost and more risk compared to an arrival price algorithm which will trade in a more aggressive manner and have higher cost but lower risk. Which is better?

Consider the results from two different algorithms. Algorithm A has lower costs on average than algorithm B. Can we conclude that A is better than B? What if the average cost from A and B are the same but the standard deviation is lower for A than for B. Can we now conclude that A is better than B? Finally, what if A has a lower average cost and also a lower standard deviation? Can we finally conclude that A is better than B? The answer might surprise some readers. In all cases the answer is no. There is simply not enough information to conclude that A is a better performing algorithm than B even when it has a lower cost and lower standard deviation. We need to determine whether or not this is a statistical difference or due to chance.

One of the most fundamental goals of any statistical analysis is to determine if the differences in results are “true” differences in process or if they are likely only due to chance. To assist with the evaluation of algorithms we provide the following definition:

*Performance from two algorithms is equivalent if the trading results are likely to have come from the same distribution of costs.*

There are two ways we can go about comparing algorithms: paired observations and independent samples.

A paired observation approach is a controlled experiment where orders are split into equal pairs and executed using different algorithms over the same time periods. This is appropriate for algorithms that use static trading parameters such as VWAP and percentage of volume (POV). These are strategies that will not compete with one another during trading and are likely to use the exact same strategy throughout the day. For example, trading 1,000,000 shares using a single broker’s VWAP algorithm will

have the same execution strategy as trading two 500,000 share orders with two different VWAP algorithms (provided that the algorithms are equivalent). Additionally, trading 1,000,000 shares with one broker's POV algorithm (e.g.,  $POV = 20\%$ ) will have the same execution strategy as using two different broker POV algorithms at one-half the execution rate (e.g.,  $POV = 10\%$  each). A paired observation approach ensures that identical orders are executed under identical market conditions. Analysts can also choose between the arrival cost and VWAP benchmark as the performance metric. Our preference for the paired sample tests is to use the VWAP.

An independent sampling approach is used to compare orders that are executed over different periods of time using different algorithms. This test is appropriate for algorithms such as implementation shortfall that manage the trade-off between cost and risk and employ dynamic adaptation tactics. In these cases we do not want to split an order and trade in algorithms that adapt trading to real-time market conditions because we do not want these algorithms to compete with one another. For example, if a 1,000,000 shares order is split into two orders of 500,000 shares and given to two different brokers, these algorithms will compute expected impact cost based on their 500,000 shares not on the aggregate imbalance of 1,000,000 shares. This is likely to lead to less than favorable prices and higher than expected costs since the algorithms will likely transact at an inappropriately faster or slower rate. The algorithm may confuse the incremental market impact from the sister order with short-term price trend or increased volatility, and react in a manner inappropriate for the fund, resulting in higher prices. Our preference is to use the arrival cost as our performance metric in the independent sample tests.

A paired observation approach can use any of the static algorithms providing that the underlying trade schedule is the same across brokers and algorithms, e.g., VWAP and POV. An independent sampling approach needs to be used when we are evaluating performance of dynamic algorithms that adapt to changing market conditions.

## Non-Parametric Tests

We provide the outline of six non-parametric tests that can be used to determine if two algorithms are equivalent. They are based on paired samples (Sign Test, Wilcoxon Signed Rank Test), independent samples (Median Test, Mann-Whitney U Test) and evaluation of the underlying data distributions (Chi-Square and Kolmogorov-Smirnov goodness of fit). Readers who are interested in a more thorough description of these tests as

well as further theory are referred to Agresti (2002), De Groot (1986), Green (2000), and Mittelhammer, Judge and Miller (2000). Additionally, Journal of Trading's "Statistical Methods to Compare Algorithmic Performance" (2007) gives additional background and examples for the Mann-Whitney U test and the Wilcoxon signed rank test. We follow the mathematical approach presented in the JOT article below.

Each of these approaches consists of: (1) devising a hypothesis, (2) the calculation process to compute the test statistic, and (3) comparing that test statistic to a critical value.

### ***Paired Samples***

For paired samples the analysis will split the order into two equal pieces and trade each in a different algorithm over the same exact time horizon. It is important in these tests to only use algorithms that do not compete with one another such as VWAP, TWAP, or POV. A static trade schedule algorithm could also be used in these tests since the strategy is predefined and will not compete with another. The comparison metric used in these tests can be either arrival cost or VWAP performance.

### ***Sign Test***

The sign test is used to test the difference in sample medians. If there is a statistical difference between medians of the two paired samples we conclude that the algorithms are not equivalent.

Hypothesis:

$H_0$ : Medians are the same ( $p = 0.5$ )

$H_1$ : Medians are different ( $p \neq 0.5$ )

Calculation Process:

1. Record all paired observations.

$(X_i, Y_i)$  = paired performance observations for algorithms X and Y.

Let  $Z_i = X_i - Y_i$ .

$k$  = number of times  $Z_i > 0$ .

$n$  = total number of pairs of observations.

2.  $T$  is the probability that  $z \geq k$  using the binomial distribution

$$T = \sum_{j=k}^n \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j} = \sum_{j=k}^n \binom{n}{j} \cdot (0.5)^j \cdot (0.5)^{n-j}$$

For a large sample the normal distribution can be used in place of the binomial distribution.

Comparison to Critical Value:

- $\alpha$  is the user specified confidence level, e.g.,  $\alpha = 0.05$ .
- Reject the null hypothesis if  $T \geq \alpha$  or  $T \geq (1 - \alpha)$ .

### **Wilcoxon Signed Rank Test**

The Wilcoxon signed rank test determines whether there is a difference in the average ranks of the two algorithms using paired samples. This test can also be described as determining if the median difference between paired observations is zero. The testing approach is as follows:

Hypothesis:

$H_0$ : Sample mean ranks are the same

$H_1$ : Sample mean ranks are different

Calculation process:

1. Let  $(A_i, B_i)$  be the paired performance results.  
Let  $D_i = A_i - B_i$  where  $D_i > 0$  indicates algorithm A had better performance and  $D_i < 0$  indicates algorithm B had better performance.
2. Sort the data based on the absolute values of differences  $|D_1|, |D_2|, \dots, |D_n|$  in ascending order.
3. Assign a rank  $r_i$  to each observation. The smallest absolute value difference is assigned a rank of 1, the second smallest absolute value difference is assigned a rank of 2, ..., and the largest absolute value difference is assigned a rank of  $n$ .
4. Assign a signed rank to each observation based on the rank and the original difference of the pair. That is:

$$S_i = \begin{cases} +r_i & \text{if } A_i - B_i > 0 \\ -r_i & \text{if } A_i - B_i < 0 \end{cases}$$

5. Let  $T_n$  be the sum of all ranks with a positive difference. This can be determined using an indicator function  $W_i$  defined as follows:

$$W_i = \begin{cases} 1 & \text{if } S_i > 0 \\ 0 & \text{if } S_i < 0 \end{cases}$$

$$T_n = \sum_{i=1}^n r_i \cdot W_i$$

Since the ranks  $r_i$  take on each value in the range  $r_i = 1, 2, \dots, n$  (once and only once)  $T_n$  can also be written in terms of its observation as follows:

$$T_n = \sum_{i=1}^n i \cdot W_i$$



- If the results are from the same distribution then the differences  $D_i$  should be symmetric about the point  $\theta = 0 \rightarrow P(D_i \geq 0) = 1/2$  and  $P(D_i \leq 0) = 1/2$ .
  - If there is some bias in performance then differences  $D_i$  will be symmetric about the biased value  $\theta = \theta^* \rightarrow P(D_i \geq \theta^*) = 1/2$  and  $P(D_i \leq \theta^*) = 1/2$ .
  - Most statistical texts describe the Wilcoxon signed ranks using a null hypothesis of  $\theta^* = 0$  and alternative hypothesis of  $\theta^* \neq 0$ . This book customizes the hypothesis test for algorithmic comparison.
6. If performance across algorithms is equivalent then there is a 50% chance that  $D_i > 0$  and a 50% chance that  $D_i < 0$ . The expected value and variance of our indicator function  $W$  is as follows:

$$E(W) = 1/2 \cdot 1 + 1/2 \cdot 0 = 1/2$$

$$V(W) = E(W^2) - [E(W)]^2 = 1/2 - (1/2)^2 = 1/4$$

7. This allows us to easily compute the expected value and variance of our summary statistic  $T_n$ . This is as follows:

$$E(T_n) = \sum_{i=1}^n i \cdot E(W_i) = \frac{1}{2} \cdot \sum_{i=1}^n i = \frac{n(n+1)}{4}$$

$$V(T_n) = \sum_{i=1}^n i^2 \cdot V(W_i) = \frac{1}{4} \cdot \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{24}$$

As  $n \rightarrow \infty$ ,  $T_n$  converges to a normal distribution and we can use the standard normal distribution to determine our critical value:

$$Z_n = \frac{T_n - E(T_n)}{\sqrt{V(T_n)}}$$

Comparison to Critical Value:

- Reject the null hypothesis if  $|Z_n| > C_{\alpha/2}$  where  $C_{\alpha/2}$  is the critical value on the standard normal curve corresponding to the  $1 - \alpha$  confidence level.
- For a 95% confidence test (e.g.,  $\alpha = 0.05$ ) we reject the null hypothesis if  $|Z_n| > 1.96$ .
- Above we are only testing if the distributions are different (therefore we use a two-tail test).
- This hypothesis can also be constructed to determine if A has better (or worse) performance than B based on whether  $D_i > 0$  or  $D_i < 0$  and using a one-tail test and corresponding critical values.

### ***Independent Samples***

The independent samples can be computed over different periods, used for different stocks. The total number of observations from each algorithm can also differ. As stated above, it is extremely important for the analyst to randomly assign trades to the different algorithms, ensure similar trading characteristics (side, size, volatility, market cap) and market conditions over the trading period. Below are two non-parametric tests that can be used to compare algorithms using independent samples. It is best to compare like algorithms in these tests such as arrival price, IS, aggressive-in-the-money, etc. Since the orders are not split across the algorithms, they can be dynamic and will not compete with one another.

### **Mann-Whitney U Test**

The Mann-Whitney U test compares whether there is any difference in performance from two different algorithms. It is best to compare “like” algorithms in this case (e.g., IS to IS, ultra-aggressive to ultra-aggressive, etc.). The arrival cost metric is the performance metric in this test.

Hypothesis:

- $H_0$ : Same performance
- $H_1$ : Different performance

Calculation Process:

1. Let  $m$  represent the number of orders transacted by broker A.  
Let  $n$  represent the number of orders transacted by broker B.  
Total number of orders =  $m + n$ .
2. Combine the samples into one group.
3. Order the combined data group from smallest to largest cost.  
For example, the smallest value receives a rank of 1, the second smallest value receives a rank of 2, ..., the largest value receives a rank of  $m + n$ .  
Identify each observation with an “A” if the observation was from algorithm A and “B” if it was from algorithm B.
4. The test statistic  $T$  is the sum of the ranks for all the observations from algorithm A.

This can be computed using help from an indicator function defined as follows:

$$W_i = \begin{cases} 1 & \text{if the observation was from algorithm A} \\ 0 & \text{if the observation was from algorithm B} \end{cases}$$

Then the sum of the ranks can be easily computed as follows:

$$T = \sum_{i=1}^n r_i \cdot W_i$$

- If the underlying algorithms are identical the actual results from each sample will be evenly distributed throughout the combined grouping. If one algorithm provides better performance results its sample should be concentrated around the lower cost rankings.
- In the situation where the null hypothesis is true the expected rank and variance of  $T$  are:

$$E(T) = \frac{m \cdot (m + n + 1)}{2}$$

$$V(T) = \frac{mn \cdot (m + n + 1)}{12}$$

As with the Wilcoxon signed rank test, it can be shown that as  $n, m \rightarrow \infty$  the distribution of  $T$  converges to a normal distribution. This property allows us to test the hypothesis that there is no difference between broker VWAP algorithms using the standard normal distribution with the following test statistic:

$$Z = \frac{T - E(T)}{\sqrt{V(T)}}$$

Comparison to Critical Value:

- Reject the null hypothesis  $H_0$  if  $|Z| > C_{\alpha/2}$ .
- $C_{\alpha/2}$  is the critical value on the standard normal curve corresponding to the  $1 - \alpha$  confidence level.
- For example, for a 95% confidence test (i.e.  $\alpha = 0.05$ ) we reject the null hypothesis if  $|Z| > 1.96$ . Notice here we are only testing if the distributions are different (therefore a two tail-test).
- The hypothesis can also be constructed to determine if A has better (or worse) performance than B by specifying a one-tail test. This requires different critical values.

Analysts need to categorize results based on price trends, capitalization, side, etc. in order to determine if one set of algorithms performs better or worse for certain market conditions or situations. Many times a grouping of results may not uncover any difference.

An extension of the Mann Whitney U test used to compare multiple algorithms simultaneously is the Kruskal-Wallis one way analysis of

variance test. This test is beyond the scope of this reference book, but readers interested in the concept can reference Mansfield (1994) or Newmark (1988).

**Median Test**

The median test is used to determine whether or not the medians of two or more independent samples are equal. If the medians of the two samples are statistically different from one another then the algorithms are not equivalent. This test is as follows:

Hypothesis:

- $H_0$ : Same medians
- $H_1$ : Different medians

Calculation Process:

1. Use arrival cost as the performance measure.
2. Choose two algorithms that are similar (e.g., arrival, IS, etc.). This experiment can be repeated to compare different algorithms.
3. Use a large enough number of orders and data points in each algorithm so that each has a representative sample size. Make sure that the orders traded in each algorithm are similar: size, volatility, market cap, buy/sell, and in similar market conditions.
4. Let X = set of observations from algorithm A.  
Let Y = set of observations from algorithm B.
5. Determine the overall median across all the data points.
6. For each sample count the number of outcomes that are less than or equal (“LE”) to the median and the number of outcomes that are greater than (“GT”) the median. Use the table below to tally these results.

	Sample A	Sample B	Subtotal
LE overall median	a	b	(a + b)
GT overall median	c	d	(c + d)
Subtotal	(a + c)	(b + d)	(a + b + c + d) = n

7. Compute the expected frequency for each cell

$$ef_{ij} = \frac{\text{total observations in row } i + \text{total observations in column } j}{\text{overall total number of observations}}$$

8. Compute test statistic  $\chi^2$

$$\chi^2 = \sum \frac{(\text{number of observations} - ef)^2}{ef}$$

$$\chi^2 = \frac{(a - ef_{11})^2}{ef_{11}} + \frac{(b - ef_{12})^2}{ef_{12}} + \frac{(c - ef_{21})^2}{ef_{21}} + \frac{(d - ef_{22})^2}{ef_{22}}$$

David M. Lane (Rice University) provided an alternative calculation of the test statistic  $\chi^2$  that makes a correction for continuity. This calculation is:

$$\chi^2 = \frac{n \left( |ad - bc| - \frac{n}{2} \right)^2}{(a + b)(c + d)(a + c)(b + d)}$$

Comparison to Critical Value:

- $df = (\text{number of columns} - 1) \cdot (\text{number of rows} - 1) = 1$ .
- Reject the null hypothesis if  $\chi^2 \geq \chi^2_{*}(df = 1, \alpha = 0.05) = 3.84$ .

## Distribution Analysis

Distribution analysis compares the entire set of performance data by determining if the set of outcomes could have been generated from the same data generating process (“DGP”). These tests could be based on either pair-samples or independent samples. Analysts need to categorize results based on price trends, capitalization, side, etc. in order to determine if one set of algorithms performs better or worse for certain market conditions or situations. Many times a grouping of results may not uncover any difference in process.

### Chi-Square Goodness of Fit

The chi-square goodness of fit test is used to determine whether two data series could have been generated from the same underlying distributions. It utilizes the probability distribution function (pdf). If it is found that the observations could not have been generated from the same underlying distribution then we conclude that the algorithms are different.

Hypothesis:

- $H_0$ : Data generated from same distribution
- $H_1$ : Data generated from different distributions

Calculation Process:

1. Use the arrival cost as the performance measure.
2. Choose two algorithms that are similar (e.g., arrival, IS, etc.).
3. Trade a large enough number of orders in each algorithm in order to generate a representative sample size. Ensure that the orders traded in each algorithm have similar characteristics such as side, size, volatility, trade time, and market cap, and were traded in similar market conditions.
4. Let  $X$  = set of results from algorithm A.  
Let  $Y$  = set of results from algorithm B.
5. Categorize the data into groups of buckets.  
Combine the data into one series. Determine the bucket categories based on the combined data. We suggest using from ten to twenty categories based on number of total observations. The breakpoints for the category buckets can be determined based on the standard deviation of the combined data or based on a percentile ranking of the combined data. For example, if using the standard deviation method use categories such as  $< -3\sigma$ ,  $-3\sigma$  to  $-2.5\sigma$ , ...,  $2.5\sigma$  to  $3\sigma$ ,  $3\sigma$  +. If using the percentile ranking method order all data points from lowest to highest and compute the cumulative frequency from  $1/n$  to 100% (where  $n$  is the combined number of data point). Select break points based on the values that would occur at 10%, 20%, ..., 100% if ten groups, or 5%, 10%, ..., 95%, 100% if twenty buckets. Count the number of data observations from each algorithm that fall into these bucket categories.
6. Compute the test statistic  $\chi^2$

$$\chi^2 = \sum_{k=1}^m \frac{(\text{observed sample X in bucket k} - \text{observed sample Y in bucket k})^2}{\text{observed sample Y in bucket k}}$$

$m$  = number of buckets.

Comparison to Critical Value:

- Reject the null hypothesis if  $\chi^2 \geq \chi^{2*}(\text{df} = m - 1, \alpha = 0.05)$

### Kolmogorov-Smirnov Goodness of Fit

The Kolmogorov-Smirnov goodness of fit test is used to determine whether two data series of algorithmic performance could have been generated from the same underlying distributions. It is based on the cumulative distribution function (cdf). If it is determined that the data samples could not have been generated from the generating process then we conclude that the algorithms are different.

Hypothesis:

$H_0$ : Data generated from same distribution

$H_1$ : Data generated from different distributions

Calculation Process:

1. Use the arrival cost as the performance measure.
2. Choose two algorithms that are similar (e.g., arrival, IS, etc.).
3. Trade a large enough number of orders in each algorithm in order to generate a representative sample size. Ensure that the orders traded in each algorithm have similar characteristics such as side, size, volatility, trade time, and market cap, and were traded in similar market conditions.
4. Let  $X$  = set of results from algorithm A— $n$  observations in total.  
Let  $Y$  = set of results from algorithm B— $m$  observations in total.
5. Construct empirical frequency distributions for each data series by ranking the data from smallest to lowest. Let  $F_A(x)$  be the cumulative probability for data series A at value  $x$  and Let  $F_B(x)$  be the cumulative probability for data series B at value  $x$ . That is, these functions represent the number of data observations in each respective data series that are less than or equal to the value  $x$ .
6. Compute the maximum difference between these cumulative functions over all values. That is:

$$D_n = \left( \frac{mn}{m+n} \right)^{1/2} \max_x |F_A(x) - F_B(x)|$$

Mathematicians will often write this expression as:

$$D_n = \left( \frac{mn}{m+n} \right)^{1/2} \sup_x |F_A(x) - F_B(x)|$$

Comparison to Critical Value:

- The critical value is based on the Kolmogorov distribution.
- The critical value based on  $\alpha = 0.05$  is 0.04301.
- Reject the null hypothesis if  $D_n \geq 0.04301$ .

## EXPERIMENTAL DESIGN

There are five concerns that need to be addressed when performing the statistical analyses described above. These are: (1) Proper Statistical Test; (2) Small Sample Size; (3) Data Ties; (4) Categorization of Data; and (5) Balanced Sample Set.

### Proper Statistical Tests

In statistical testing, the preferred process is a controlled experiment so that the analyst can observe the outcomes from two separate processes under identical market conditions (e.g., Wilcoxon signed rank test). While this is an appropriate technique for static strategies such as VWAP and POV algorithms, it is not an appropriate technique for those algorithms with dynamic trading rates and/or those that employ real-time adaptation tactics. Employing a controlled experiment for dynamic algorithms will likely cause the algorithms to compete with one another and will lead to decreased performance. For dynamic algorithms (e.g., implementation shortfall and ultra-aggressive algorithms) it is recommended that investors utilize the two sample non-pair approach and the Wilcoxon-Mann-Whitney ranks test.

In theory, it is appropriate to compare algorithms with static strategies (e.g., VWAP and POV) with the Wilcoxon-Mann-Whitney ranks test. However, doing so causes more difficulty with regards to robust categorization and balanced data requirements. It is recommended that algorithms with static parameters be compared via the Wilcoxon signed rank test approach.

### Small Sample Size

In each of these statistical techniques it is important to have a sufficiently large data sample in order to use the normal approximation for hypothesis testing. In cases where the sample sizes are small (e.g.,  $n$  and/or  $m$  small) the normal distribution may not be a reasonable approximation methodology and analysts are advised to consult statistical tables for the exact distributions of  $T_n$  and  $T$ . We recommend using at least  $n > 100$  and  $m > 100$  for statistically significant results.

### Data Ties

It is assumed above the results are samples from a continuous distribution (i.e., statistically there will never be identical outcomes). Due to finite precision limitations, analysts may come across duplicate results, inhibiting a unique ranking scheme. In these duplicate situations, it is recommended that the data point be included in the analysis twice. In the case that algorithm “A” is the better result for one data point and algorithm “B” is the better result for the second data point, a unique ranking scheme will exist. If the tail areas of the results are relatively the same this approach should not affect the results. If the tail areas are different this may be a good indication that the data is too unreliable and further analysis is required. Analysts with strong statistical training may choose alternative ranking schemes in times of identical results.



## Proper Categorization

When analyzing algorithmic performance it is important to categorize trades by side (buy/sell/short), size, market conditions (such as up and down days), company capitalization (large, mid, and small cap), etc. Categorization allow analysts to determine if one algorithm works statistically better or worse in certain situations. For example, if VWAP algorithm “A” makes market bets by front-loading executions and VWAP algorithm “B” makes market bets by back-loading, “A” will outperform “B” for buys on days with a positive drift and for sells on days with a negative drift. Conversely, algorithm “B” will outperform “A” for buys on days with a negative drift and for sells on days with a positive drift. A statistical test that combines executions from a large array of market conditions may miss this difference in performance especially if we are comparing averages or medians. It is essential that analysts perform robust statistical hypothesis testing for all performance testing techniques.

## Balanced Data Sets

It is imperative that analysts utilize a random selection process for submitting orders to algorithms and ensure that the data sets are balanced across the specified categorization criteria, e.g., size, side, capitalization, market movement, etc. This basically states that the percentage breakdown in the categorization groups described above will be similar. Otherwise, the statistical results may fall victim to Simpson’s Paradox (e.g., dangers that arise from drawing conclusions from aggregate samples).

## FINAL NOTE ON POST-TRADE ANALYSIS

One final note on post-trade analysis is the following. Consider the possibility that performance is equivalent across all families of algorithms. For example, there is no difference across VWAP algorithms, IS algorithms, ultra-aggressive algorithms, etc. Subsequently, two important issues arise. First, can brokers still add value to the trading process? Second, is there any need for third party post-trade services? The answer to both these questions is yes.

Brokers can still add value to the process by providing appropriate pre-trade analysis to ensure proper selection of algorithms and algorithmic parameters. Furthermore, brokers can partner with investors to customize algorithms to ensure consistency across the investment and trading decisions. For example, see Kissell and Malamut (2006) and Engle and Ferstenberg (2006). Most

importantly, however, broker competition propels innovation and advancement that continue to benefit investors.

Third party consultants also serve as an essential service to the industry. Not only can they be used by the buy-side to outsource numerical analysis, but more importantly, these consultants have access to a larger universe of trades for various investment styles and algorithms, both robust and balanced, and are thus positioned to provide proper insight into performance and trends. Comparatively, brokers typically only have access to trades using their algorithms and investors only have access to their trades. Access aside, the statistical testing procedure of these consultants cannot remain a black box; transparency is crucial in order for the industry to extract value from their services. Transaction cost analysis remains an essential ingredient to achieve best execution. When administered properly, improved stock selection and reduced costs have proven to boost portfolio performance. As such, advancement of TCA models is an essential catalyst to further develop the algorithmic trading and market efficiency space.