Advanced Python for Finance

Lecture 2

Lecture 2

- Market Data, Initial Stock Analysis
- Working with Bar Data
- Working with Tick Data
- Case Study: Monte Carlo Simulation and Stock Price Simulation
- Case Study: MACD, etc.
- Intro to Transaction Costs

Working with Data

Types of Market Data

1. Bar Data – summary market data taken at intervals
2. Tick data – lowest level of market data
   - Quotes
   - Trades

Bar Data

- Aggregated over some time interval
- Large Scale: daily, weekly, monthly, quarterly, etc.
- Intraday: 1-minute, 5-minute, 10-minute, etc.
- Important data points:
  - Open
  - High
  - Low
  - Close
  - Volume
- Useful for analyzing trends, evaluating alpha processes of a certain scale.
- Limited utility for true microstructure work
- <u>You can always create lower resolution (i.e. larger bars) from more granular data, but not vice versa.</u>

Tick Data

- Tick data literal change-by-change data, in which each data point reflects a single (visible) update to an order book:
  - Bid size change
  - Bid price change
  - Offer size change
  - Offer price change
  - Transaction
- Data format and file layouts will vary by product and vendor, but these are general principles.
- Market data may also include different kinds of messages (market status, etc.)
- Tick Data may be **every** event or Best Bid and Offer (BBO)

Market Data Sources

Recommended for class

- https://www.alphavantage.co: free bar data
- https://www.tickdata.com/equity-data/ - tick data sample available
- Nanotick: CME Futures Tick data (tick data) – sign up for demo data

Other Data Sources

- https://stooq.com/db/h/ bar data
- Barchart – licensed bar data
- Xignite – licensed bar data

Note your data source in assignment and project credits

Market Data Examples

L2.1 Working with Bar Data.ipynb

L2.2 Working with Tick Data.ipynb

L2.3 Working with Tick Data 2.ipynb

L2.4 Working with Tick Data 3.ipynb

Modeling Security Price Dynamics: Random Walks

The Random Walk Model of Security Prices

- Before financial economists began to concentrate on the trading prices the *standard* statistical model for a security prices was the *random walk*.
- The random walk model is no longer considered to be a complete and valid description of short-term price dynamics, but it nevertheless retains an important role as a model for the fundamental security value.
- This arises from information about security *cash flows*, which are *long-lasting*, in contrast to the effects attributable to the *market organization* and *trading process* which are *transient*.
- Let p_t denote the transaction *price* at time t, where t indexes regular points of the real calendar (clock) time, for example, end-of-day, end-of-hour, and so on.
- Because it is unlikely that trade occur exactly at these times, we will approximate these observations by using prices of the last (most recent) trade, for example, the day's *closing prices*. This assumption will hold true unless we are dealing with actual tick data.

The Random Walk Model of Security Prices

The random walk model (with drift) is given by:

$$p_t = p_{t-1} + \mu + u_t \qquad \text{(EQ 1)}$$

- where the u_t, t = 0, 1, 2,… are *independently* and *identically distributed* (i.i.d.) random variables. Intuitively, they arise from *new information* that bears on the security value.
- μ is the expected price change or drift (assumed for now to be a constant).
- The units of p_t are either *levels* (e.g. dollars per share) or *logarithms*.

The logarithmic form is sometimes more convenient because price changes can be interpreted as continuously compounded returns. Some phenomena, however, require level (price) representation. Price discreteness, for example, reflects a tick size (minimum price increment) that is generally set in a level unit.

The Random Walk Model of Security Prices

In most microstructure analyses the drift is assumed to be zero and is dropped (see below). When $\mu = 0$, p_t cannot be forecast beyond its most recent value, that is:

$$E[p_{t+1} \mid p_t, p_{t-1}, p_{t-2}, \dots] = p_t \qquad \text{(EQ2)}$$

A process with this property is generally described as a *martingale.*

Definition (Martingale): A discrete time stochastic process $\{x_t\}$ is a *martingale* if $E|x_t| < \infty$ for all t, and $E[x_{t+1} \mid x_t, x_{t-1}, \dots] = x_t$.

It follows immediately from the law of iterated expectations that:

Corollary (k-period Forecast): A discrete time stochastic process $\{x_t\}$ is a martingale if and only if $E|x_t| < \infty$ for all t and $E[x_{t+k} \mid x_t, x_{t-1}, \dots] = x_t$ for all $k > 0$.

This corollary can be used as an alternative definition of a martingale process. The following theorem establishes the connection between (discrete) transaction (trade) prices of securities and (discrete time) martingale processes:

The Random Walk Model of Security Prices

Theorem (Trade Prices and Martingales): Under the assumption of *frictionless markets* and *absence of arbitrage* (or security market equilibrium), the *transaction prices* of a (non-dividend paying) security can be represented by a discrete-time *martingale* process.

Martingale behavior of asset prices is a classic result arising in many economic models with individual optimization, absence of arbitrage, or security market equilibrium. The result is generally *contingent*, however, on the assumption of frictionless trading opportunities, which are *not* appropriate in most microstructure applications.

The expectation shown above are conditioned on lagged values of x_t, that is, the *history* of the process. A more general definition involves conditioning on *broader* information sets (on the next slide):

The Random Walk Model of Security Prices

A special case of martingale is the random walk is constructed as the sum of independently and identically distributed random variables with zero-mean:

Definition (Random Walk): A *random walk* is a process whose *increments* are *independently* and *identically distributed* (i.i.d.) *zero-mean* random variables.

The transaction price in Equation 1, for example, cumulates the random variables u_t. Because u_t are i.i.d., the price process is *time-homogeneous*, that is, it exhibits the *same* behavior whenever in time we sample it. This is only sensible if the economic process underlying the security is also time-homogeneous. Stocks, for example, are claims on ongoing economic activities (of the firm, sector, overall economy etc.) and are therefore plausibly approximated in the long run by random walk.

Statistical Analysis of Security Price Series

Statistical inference in the random-walk model is straightforward. Suppose that we have a sample price series $\{p_1, p_2, \ldots, p_T\}$, generated in accordance with Equation 1. Because the u_t are i.i.d., the price changes $\Delta p_t = p_t - p_{t-1}$ should be i.i.d. with mean μ and variance $Var(u_t) = \sigma_u^2$, for which we can compute the usual estimate. When we analyze actual data samples, however, we often encounter features that suggest wariness in the interpretation and subsequent use of the estimates.

Short-run security price changes typically exhibit:

- *Means* very close to zero (zero drift assumption)

- Extreme *dispersion*(fat tails)

- *Dependence* between successive observations (autocorrelations)

We will elaborate on each of these points further below.

Near-Zero Mean Returns

In microstructure data samples μ is usually *small* relative to the *estimation error* of its usual estimate, the logarithmic mean. For this reason it is often preferable to drop the mean return from the model, implicitly setting μ to *zero*. This is illustrated in Figure 4 below.

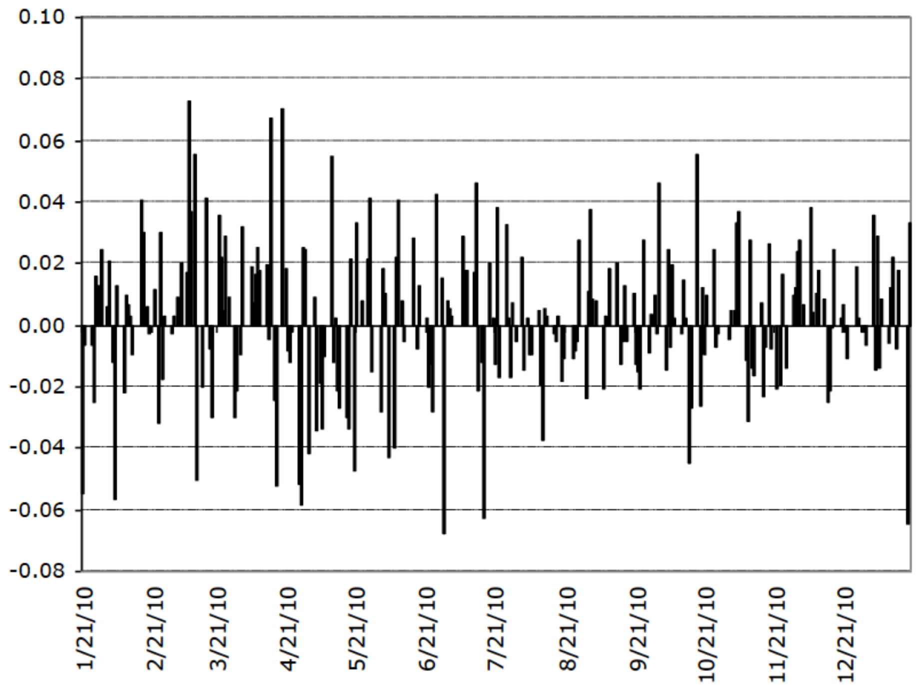


Figure 4: Daily Returns of Citigroup Inc. Stock during January 2011

Near-Zero Mean Returns

Zero, is of course, a *biased* estimate of μ, but its *estimation error* will generally be *lower* than that of arithmetic mean. For example, suppose that t indexes *days*. Consider the properties of the annual log price change implied by the *log random-walk* model:

$$p_{365} - p_0 = \sum_{t=1}^{365} \Delta p_t = \mu_{Annual} + \sum_{t=1}^{365} u_t \qquad \text{(EQ 3)}$$

where $\mu_{Annual} = 365\mu$. The *annual variance* is

$$\sigma^2_{Annual} := Var(p_{365} - p_0) = 365\sigma^2_u \qquad \text{(EQ 4)}$$

A typical U.S. stock might have an annual expected return of $\mu_{Annual} = 0.1$ (or 10%) and an annual variance of $\sigma^2_{Annual} = 0.25^2$. The implied *daily expected return* is

$$\mu_{Daily} = \mu_{Annual} / 365 = 0.1/365 = 0.000274 \qquad \text{(EQ 5)}$$

and the implied *daily variance* is given by:

$$\sigma^2_{Daily} = \sigma^2_{Annual} / 365 = 0.25^2 / 365 = 0.000171 \qquad \text{(EQ 7)}$$

Near-Zero Mean Returns

With $n = 365$ daily observations, the *standard error* of estimate for the sample *mean* is

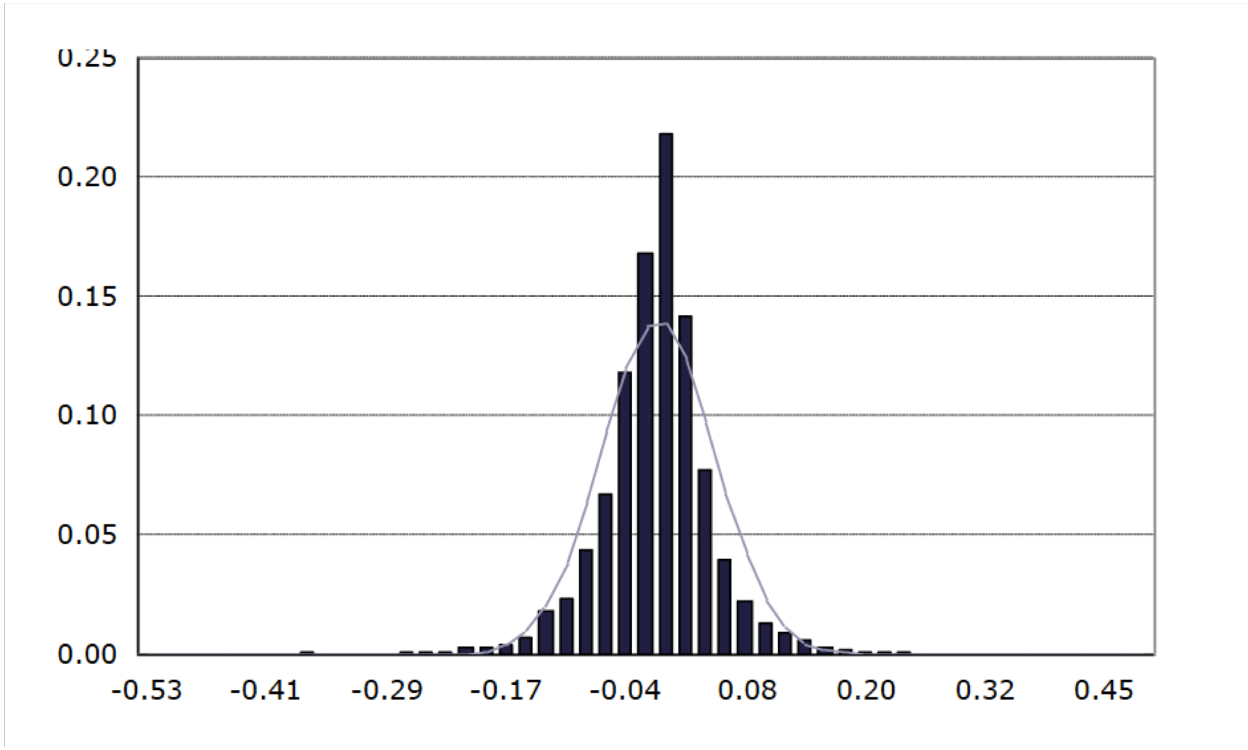
$$Stderr\left(\mu_{Daily}\right) = \sqrt{\sigma^2_{Daily} / 365} = \sqrt{\left(\sigma^2_{Annual} / 365\right) / 365} = 0.000685 \qquad \text{(EQ 8)}$$

This is about two and half times the true mean. An estimate of zero is clearly biased downward, but the standard error of estimate, being 0.000274, is a relatively small number. At the cost of a little bias, we can greatly reduce the estimation error by making the assumption that the mean is zero.

As we refine the frequency of observations from annually to monthly to daily and so on, the *number of observations* increases. More numerous observations usually enhance the precision of the estimate, but often only if it also increases the *calendar span* of the sample. Here, though, the increase in observations is not accompanied by any increase in the calendar span of the sample. Merton (1980) has shown that estimate of *second moments* (variances and covariances) are *improved* by more frequent sampling. Estimates of the mean *returns* are *not*. For this reason, the expected return will often be dropped from our microstructure models.

Extreme Dispersion

Statistical analysis of speculative price changes at all horizons generally counters sample distributions with fat tails. The incidence of extreme values is so great as to raise doubt whether population parameters like kurtosis, skewness or even the variance of underlying distributions are finite. The behavior is depicted in the chart below.



Fat-tailed Distribution of Daily Stock Returns

Extreme Dispersion

The conventional assumption that price changes are *normally distributed* is routinely *violated*. For example, from July 7, 1962 to December 31, 2004 (10,698 observations), the average daily return on the Standard & Poor's S&P 500 index is about 0.0003 (0.03%) and the standard deviation is about 0.0094. Letting $\Phi(z)$ denote the standard normal distribution function, if returns are normally distributed, then the number of days with returns below -5% is expected to be $10{,}698\, x\, \Phi((-0.05-0.0003)/0.0094) \approx 0.0005$, that is considerably less than 1. In fact, there are eight such realizations (with minimum of -20.5% occurring on October 19,1987).

Statistical analysis of this sort of dispersion falls under the rubric of *extreme value analysis*. For a random variable X the population moment of order α is defined as $E(X^{\alpha})$. The normal probability density possesses *finite* moments of *all orders*. In other distributions, though, a moment may be infinite because as X goes to $\pm\infty$ the quantity X^{α} grows faster than the probability density declines.

Extreme Dispersion

Theorem (Asymptotically Consistent Estimates of the Mean): Let $\alpha > 0$. If $E(X^\alpha)$ is finite, then the *sample estimates* $\sum_{t=1}^{T} X_t^\alpha / T$, where T is the sample size, is an asymptotically *consistent* estimate of $E(X^\alpha)$.

Proof follows from the application of the law of large numbers.

This theorem states that under the assuming $E(X^\alpha) < \infty$ we have *consistent sample estimates* of the mean $E(X^\alpha)$. *Hypothesis testing*, however, often requires existence of the *asymptotic variance* of the *sample estimate*, which *requires* existence of moments of order 2α. To get the standard error of the mean, for example, we need a consistent estimate of the variance.

Extreme Dispersion

A recent study (Gabaix et al. 2003) suggests that finite moments of daily equity return exist only up to order 3 and for daily trading volume only up to order 1.5. These findings, if correct, impose substantive *restrictions* on the sorts of models that can be sensibly estimated.

Why should one be concerned about *convergence* failures in infinite samples? The answer is that whatever one's beliefs about the properties of the distribution generating the data, the existence of extreme values in finite samples is an irrefutable fact leading to many *practical* consequences. Sample estimates may be *dominated* by a few *extreme* values. *Increasing* sample size does *not* increase precision as much as we would expect. Estimate parameters are *sensitive* to *model specification*. Finally, conclusions drawn from the model are *fragile*, which is especially disturbing in trading applications.

Dependence of Successive Observations

Time series data are *ordered*, and statistical analysis must at least allow for the probability that there is *dependence* among the ordered data. The most important summary measures of time-series dependence are *autocovariances* and *autocorrelations*.

For a real-valued time series $\{x_t\}$ the autocorrelations and autocovariance (of order k) are defined as $Cov(x_t, x_{t-k})$ and $Corr(x_t, x_{t-k})$ for $k = 0, 1, 2\ldots$.

Under the assumptions above (stationarity of time series), these quantities depend *only* on k, the *lag* (separation) between the component terms. Accordingly they can be expressed as $\gamma_k = Cov(x_t, x_{t-k})$ and $\rho_k = Corr(x_t, x_{t-k})$. When the *mean* of the series is *zero*, these quantities can be *estimated* using the sample average cross-product as

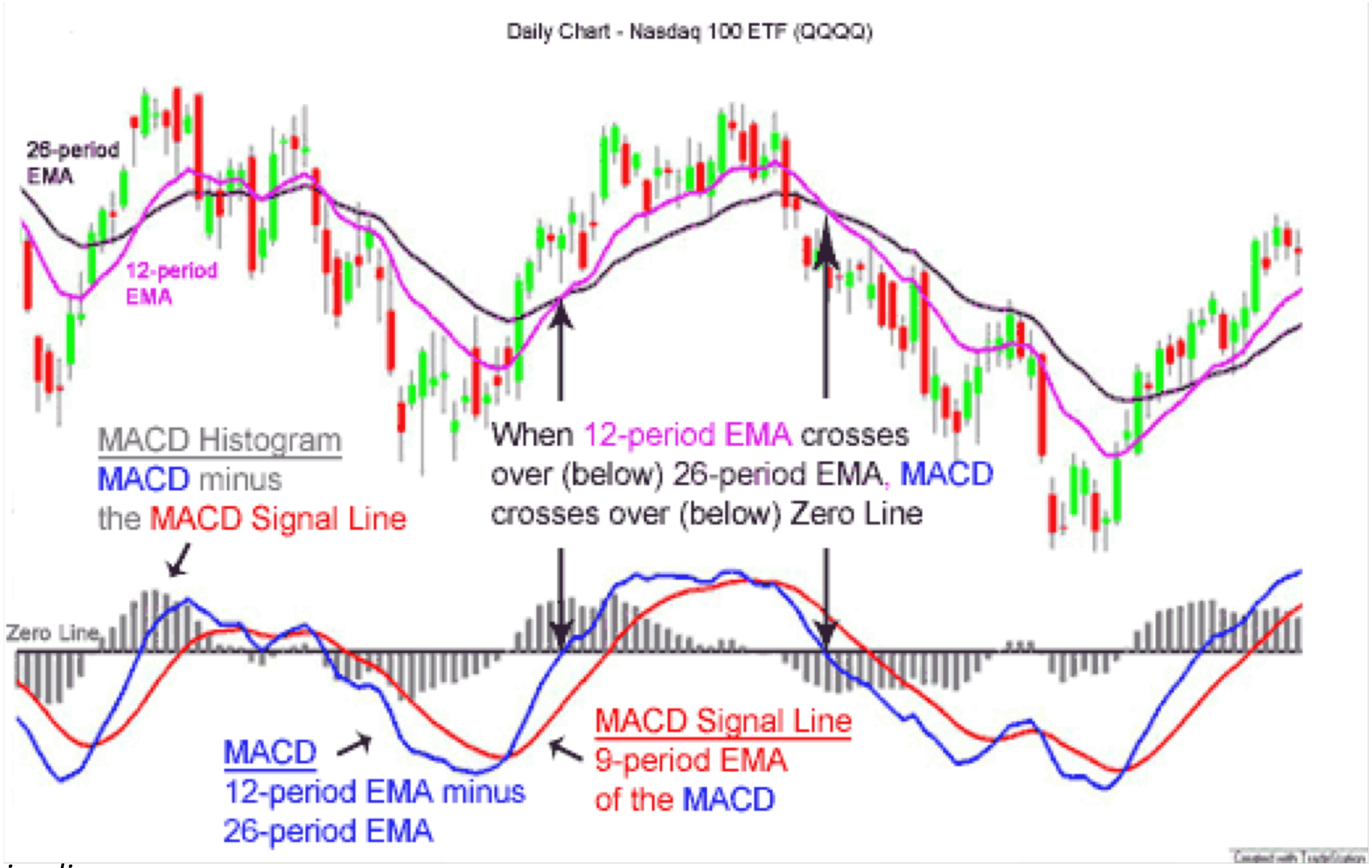
$$\hat{\gamma}_k = \sum_{t=k+1}^{T} x_t x_{t-k} / (T-k) \text{ and } \hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0 \qquad \text{(EQ 9)}$$

The incremental changes in a *random walk* are *uncorrelated*. So we would expect to find $\hat{\rho}_k \approx 0$ for $k \neq 0$. In actual samples, however, the first-order autocorrelations of short run speculative price changes are usually *negative*. The economic interpretation for this finding is the main contribution of the Roll model.

See L2.5 RandomWalks.ipynb

Case Study
Stock Price Analysis and Basic Trading Simulation

MACD



Source: Wikipedia

See L5.4 - MACD.ipynb

See L2.6 MACD.ipynb

Transaction Cost Analysis

Transaction Cost Components - Review

We can describe these transaction costs as a pyramid, with the most visible costs on the top (e.g, they can only be seen from a distance) and the lease transparent costs shown on the bottom (they can be only seen up close), as shown:

In this figure, the component costs *most visible* from a distance are those costs that contribute the *least* to the total transaction cost. Costs *least visible* (non-transparent) from a distance contribute *most* to the total transaction cost. Fortunately, these non-transparent cost components provide the greatest *opportunity* for *cost reduction* by skilled managers / traders. Unfortunately, the cost reduction of one non-transparent cost is typically at the *expense* of another non-transparent cost. Therefore, traders need to understand *all* costs and how they interact with one another. For example, as we reduce market impact by trading more passively we expose the fund to greater risk. As we trade more aggressively we reduce risk but increase market impact. It is not possible to reduce all costs

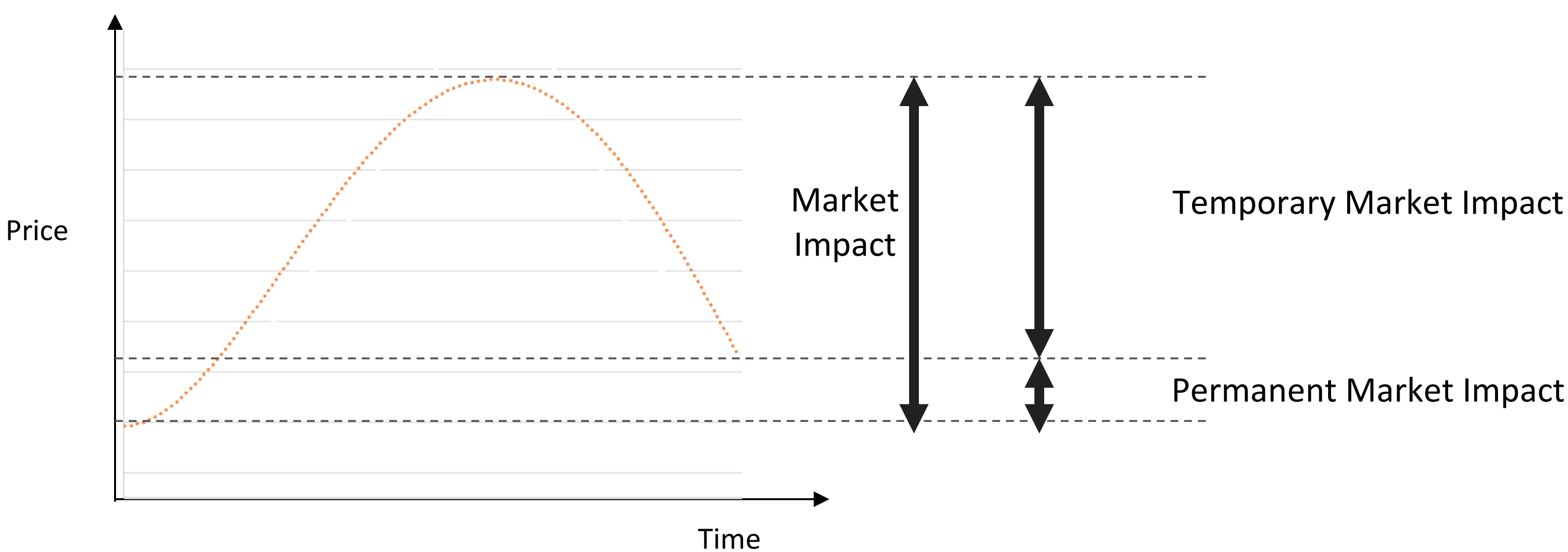
Market Impact

Market impact is perhaps the most significant topic in transaction cost analysis.

It represents the price change caused by the execution of a specific order, and typically has an adverse effect on price, driving prices up (buy) or down (sell). There are two main driving factors for market impact:

- Temporary Impact: Caused by ‘consumption’ of liquidity (orders) by ‘walking ’ the order book
- Permanent Impact: Caused by information conveyed to the market through the execution of order in the market

Market Impact = Temporary Impact + Permanent Impact



Timing Risk

Timing risk represents the uncertainty in the transaction cost estimate.

Two main sources of timing risk are the asset's price uncertainty due to volatility and the traded volume.

$$\text{Timing Risk} = \text{Volatility Risk} + \text{Liquidity Risk}$$

Volatility risk pertains to the change in asset price during the execution of an order. Price volatility is arguably the most important risk. Higher volatility may cause the price to move away and lead to high transaction costs.

Liquidity risk represents the variability of availability during the trading period. A drop in liquidity will increase the market impact while an increase will lower the market impact.

Opportunity Cost

Opportunity cost reflects the cost of not fully executing an order. This may happen when asset's price moves beyond client's defined limit price or insufficient liquidity. Opportunity cost can be measured as:

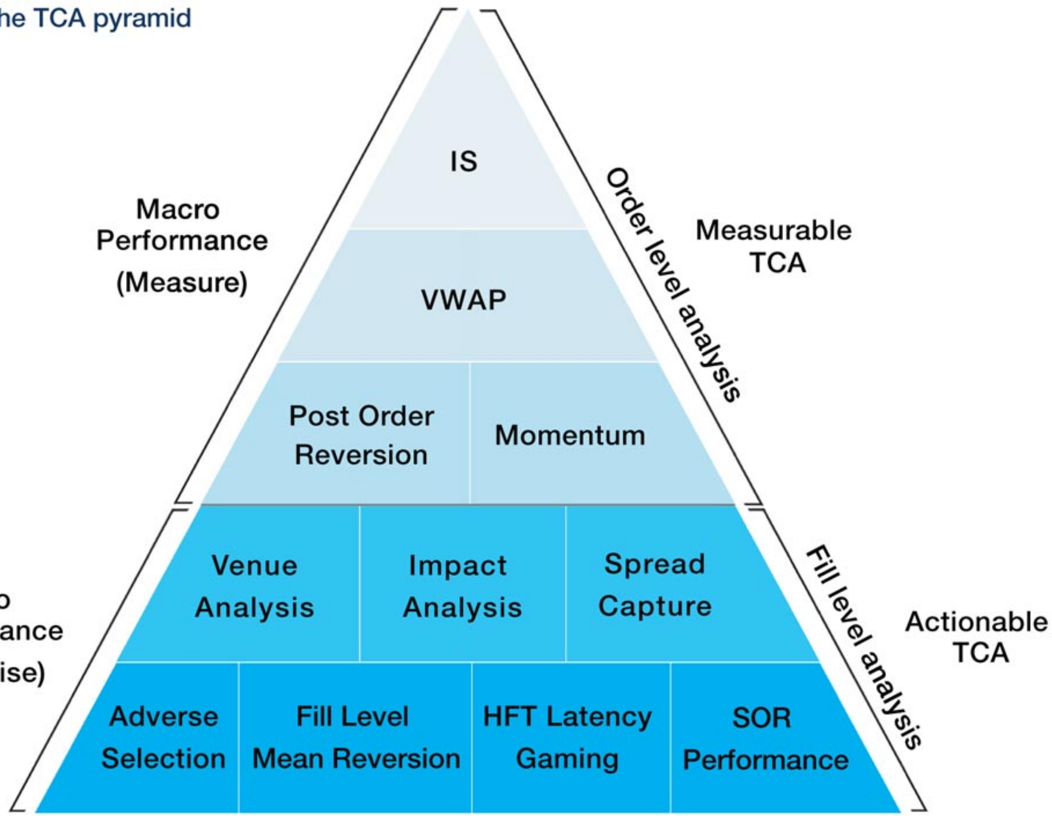
$$\text{Opportunity Cost} = (X - \sum x)(Final\ Price - Arrival\ Price)$$

where X is the Total Order Size and *x* are the individual executions (filled shares) at a given point in time.

Pre-trade analysis can help minimize opportunity cost by sizing the order correctly for market conditions.

Transaction Cost –Another Pyramid!

Figure 1: The TCA pyramid



Source: LiquidMetrix

Macro

- IS => Implementation Shortfall / Arrival Price
- VWAP => VWAP
- Other Aspects
  - Reversion
  - Momentum

Micro structure

- Venue Analysis – Slippage vs. Reversion
- Spread Capture - % of Spread
- Fill Level Adverse Selection / Reversion
- SOR Performance (price improvement)

Key Points in the Transaction Lifecycle

Lifecycle Point	Definition
Previous Close	Last price of previous trading session
Market Open	Exit price of opening auction
Order Arrival	Price at the time an order (or the desire to trade) arrives. One of: • Last price (if instrument is open and liquid market) • Market midpoint (if instrument is not open or not sufficiently liquid) • Previous close (if instrument is not open and/or current market is insufficiently developed (e.g. no bid or offer or excessively wide spread)
Average Price	Volume-weighted average execution price
Order Completion	Last price at time of order completion
Market Close	Exit price of closing auction
Next Open	Exit price of next trading session’s opening auction

Common Performance Benchmarks

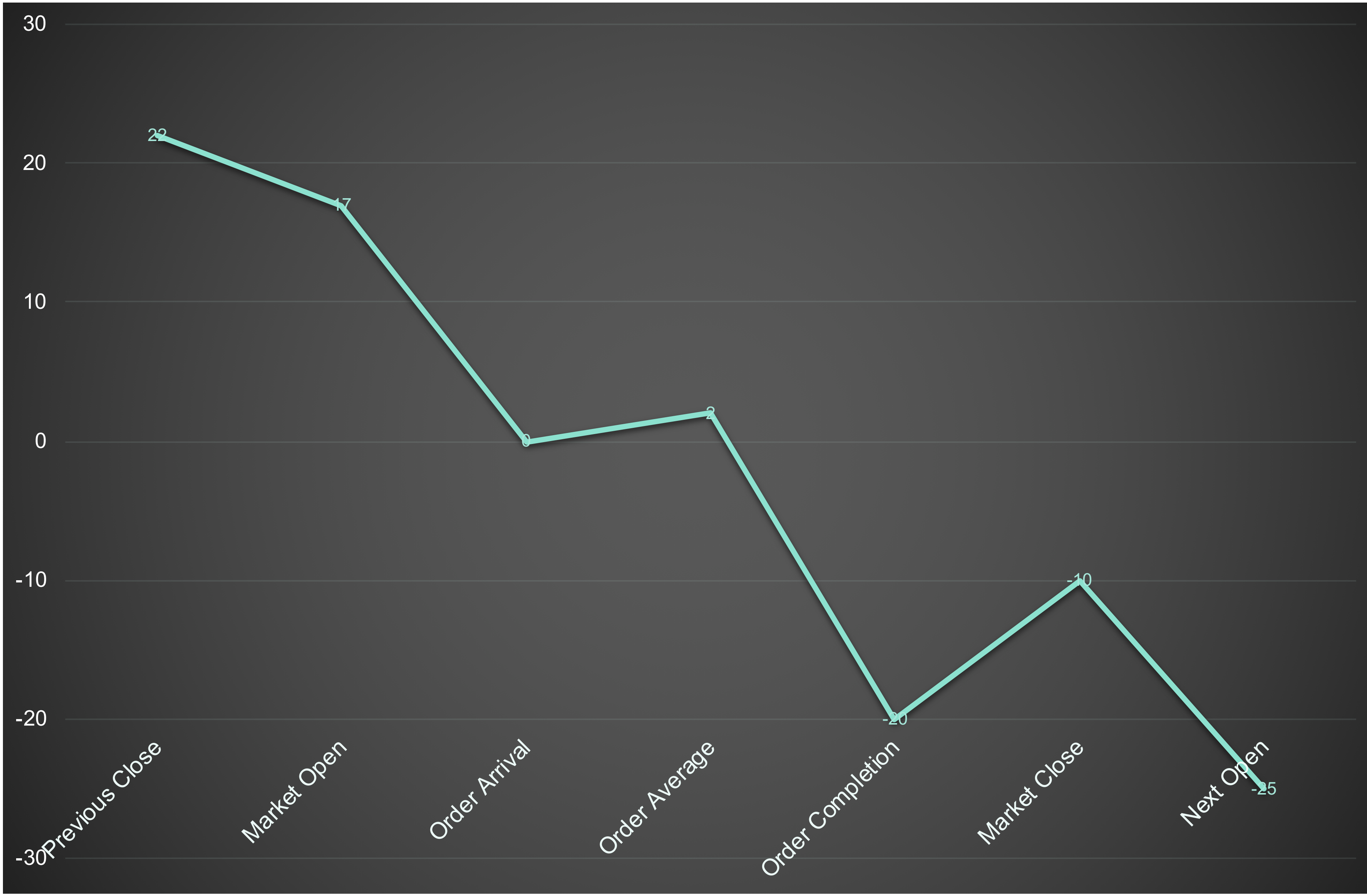
Metrics and Benchmarks	Definition
P_{avg} Volume weighted average price of *order*	Given an order of quantity *Q*, with executions of quantity *q* at price *p*, : $$\frac{\sum(q_i \times p_i)}{Q}$$
TWAP Time-weighted *market* average	1. For each bin *i*, mean($open_i$, $high_i$, low_i, $close_i$) 2. Calculate the mean across all bins
VWAP (or $VWAP_{mkt}$) Volume-weighted *market* average	$$= \frac{\sum(qty_{@price} \times price)}{total\ market\ qty}$$
Performance vs. TWAP	$= P_{TWAP} - P_{avg}$ *
Performance vs. VWAP	$= P_{VWAP} - P_{avg}$ *
Performance vs. Arrival/IS	$= P_{arrival} - P_{avg}$ *
Performance vs. Open	$= P_{OPEN} - P_{avg}$ *
Performance vs. Close	$= P_{CLOSE} - P_{avg}$ *

**In practice these are standardized, e.g.: $((P_{TWAP} - P_{avg}) / P_{avg}) * 10000$ and expressed in basis points*

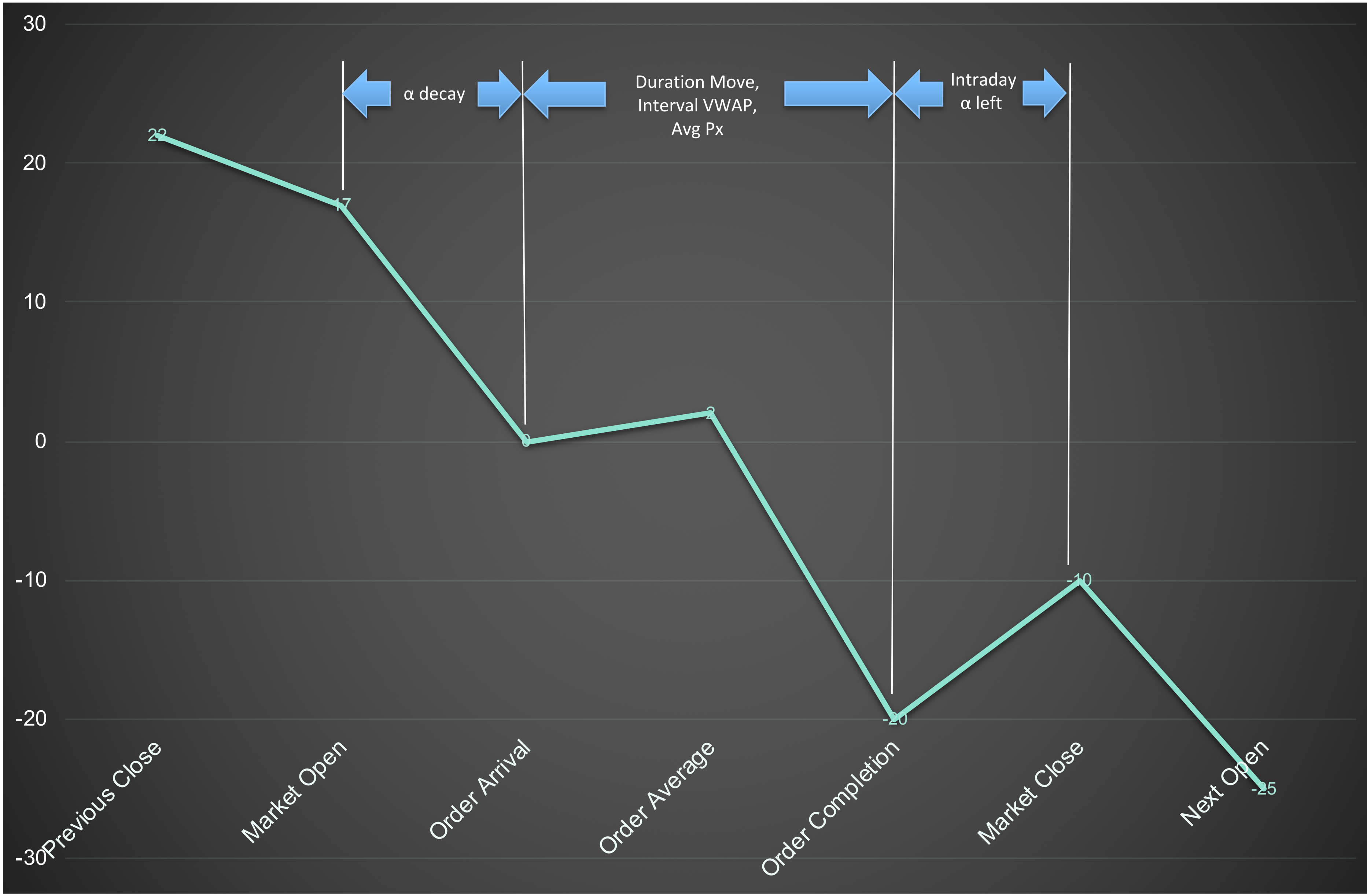
P&L Terminology

- One approach: Residual Vs. Day P&L
  - Isolates each day's trading P&L
  - Total P&L = Residual + Day where:
    - Residual = Σ $Position_{SOD}$*(MTM – previous close)
    - Day = Σ $Position_{intraday}$*(Avg Sell – Avg. Buy) + $Position_{Resid}$*(MTM – Avg Px)
- Loss Ratios (Principal Trading)
  - = Loss (always?) per unit of commission
- Total P&L = model P&L + execution P&L + fees
  - <u>Decompose P&L to determine performance drivers</u>

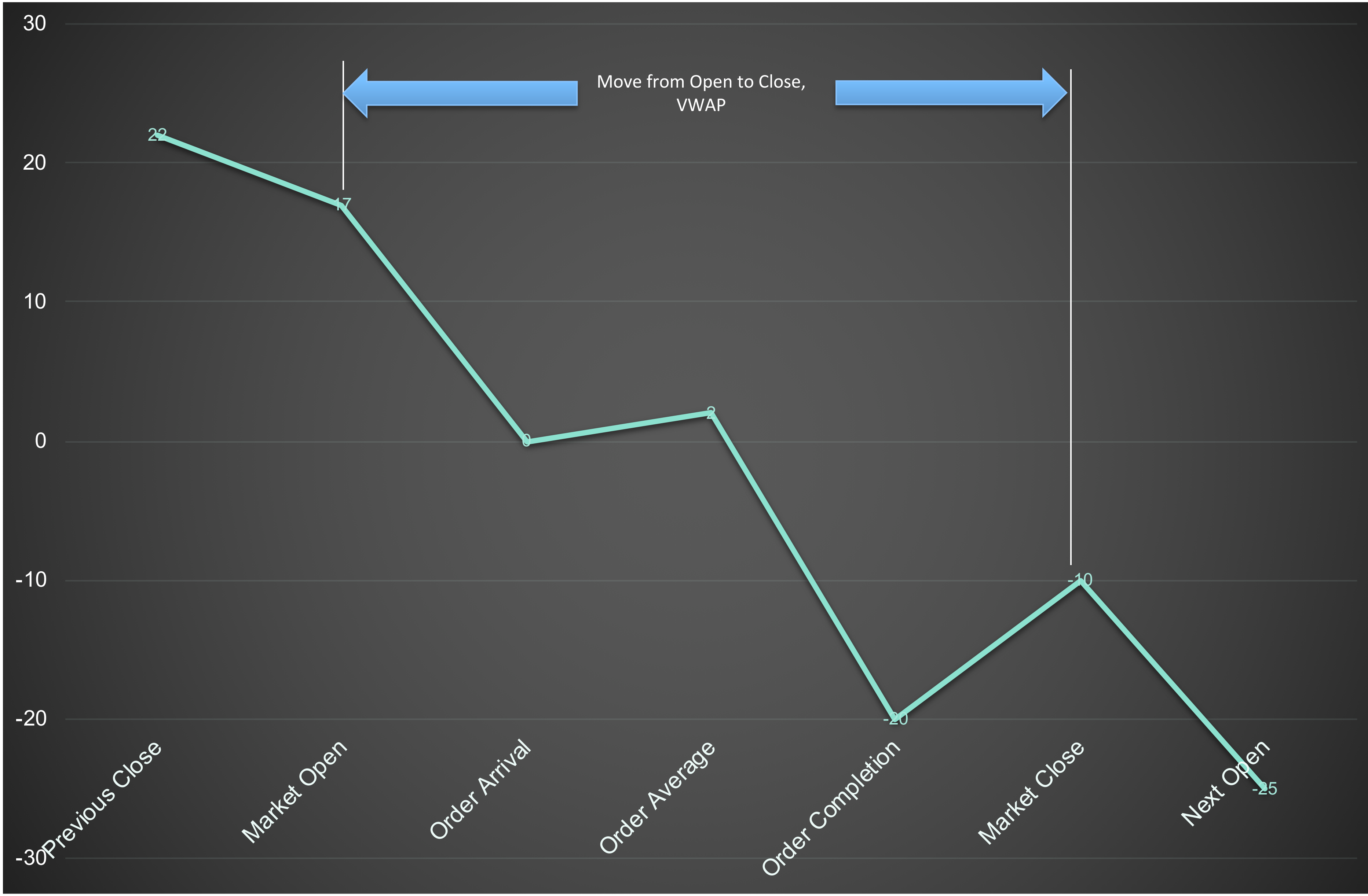
Performance vs. Points in Time



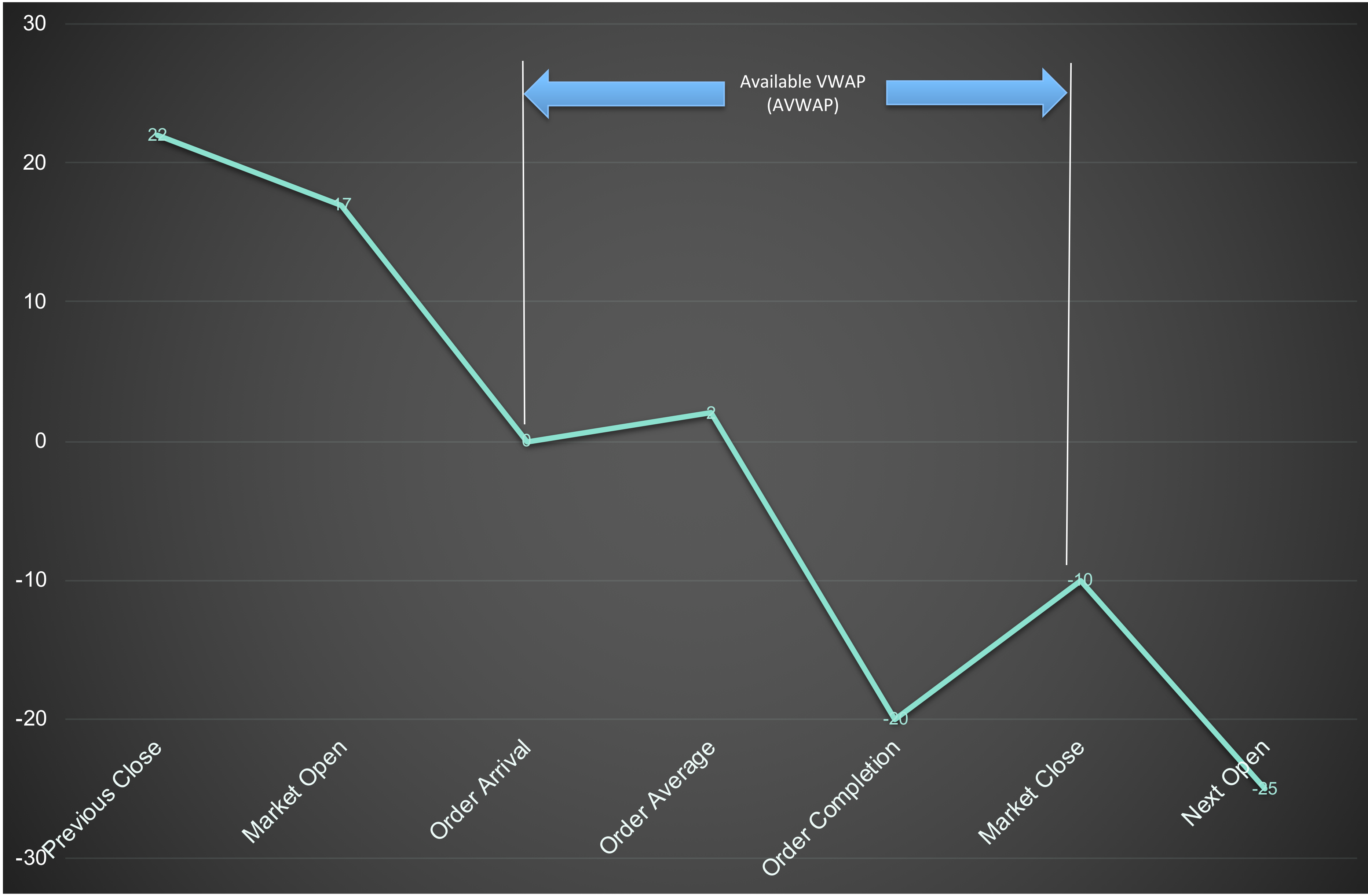
Order Life Span



Intraday Move, VWAP



Available VWAP



Post-Trade TCA – Examples

Post-Trade Analyses focus on <u>average performance of multiple trades</u> over time:

- By Algo (if applicable)
- By Benchmark
- By Trader
- By Duration
- By Sector
- Etc.

Single Stock TCA

- Outlier analysis (i.e. what went wrong!)
- Not as common for general trades

Trading Summary

Value Traded:	€3,053,204,896
(Bought):	€1,513,902,868
(Sold):	€1,539,302,028
Number Of Orders:	8,204
(Fills):	162,334
Average Order Size:	€372,160.52
(Average Fill Size):	€18,808.17
Book Spread (depth):	31.18 BPS
Book Spread (touch):	20.70 BPS
Average Duration:	0h 50m
Distinct Instruments:	519

Execution Style:

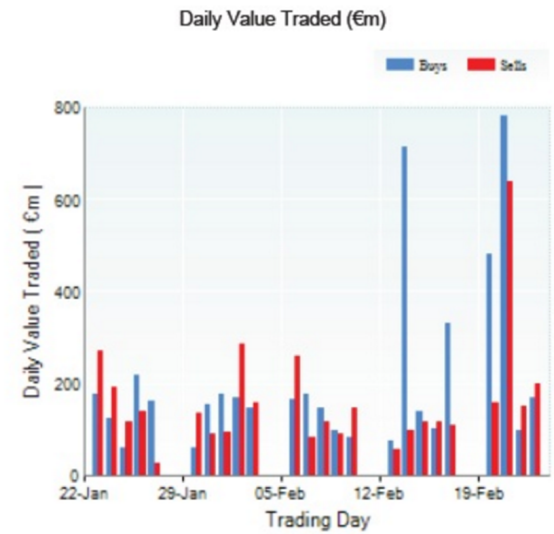
Aggr	Pass	Mid
99.3%	0.7%	0.1%

Venue Types:

Prim	MTF	Other
94.3%	0.0%	5.7%

Venues Traded:

XMAD: 71.51% XPAR: 7.02% XETR: 5.53% XHKG: 2.81% XNYS: 2.34% XSWX: 2.32% XAMS: 1.90% XMIL: 1.90% LSE2: 1.86% XLIS: 0.62% XSTO: 0.62% XHEL: 0.52% XNMS: 0.38% XOSL: 0.28% XBRU: 0.17% XSES: 0.11% XCSE: 0.03% XIOB: 0.02% XVIE: 0.02% XDUB: 0.01%

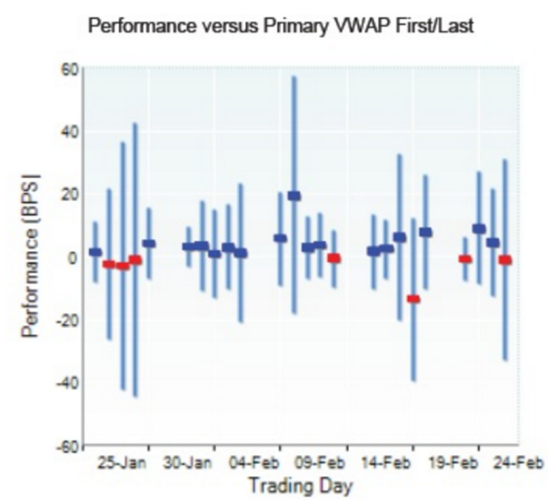


Performance Summary

Performance Benchmark	All	Buys	Sells
Market VWAP First/Last	1.88	-1.55	5.26
Primary VWAP First/Last	1.93	-1.40	5.22
Market Daily VWAP	7.46	-8.31	22.86
Primary Daily VWAP	1.93	-1.40	5.22
First Fill (mid)	-38.66	-93.43	26.46
Previous Closing Price	22.94	14.66	31.04
Closing Price	1.41	4.72	-1.82

Performance versus Primary VWAP First/Last

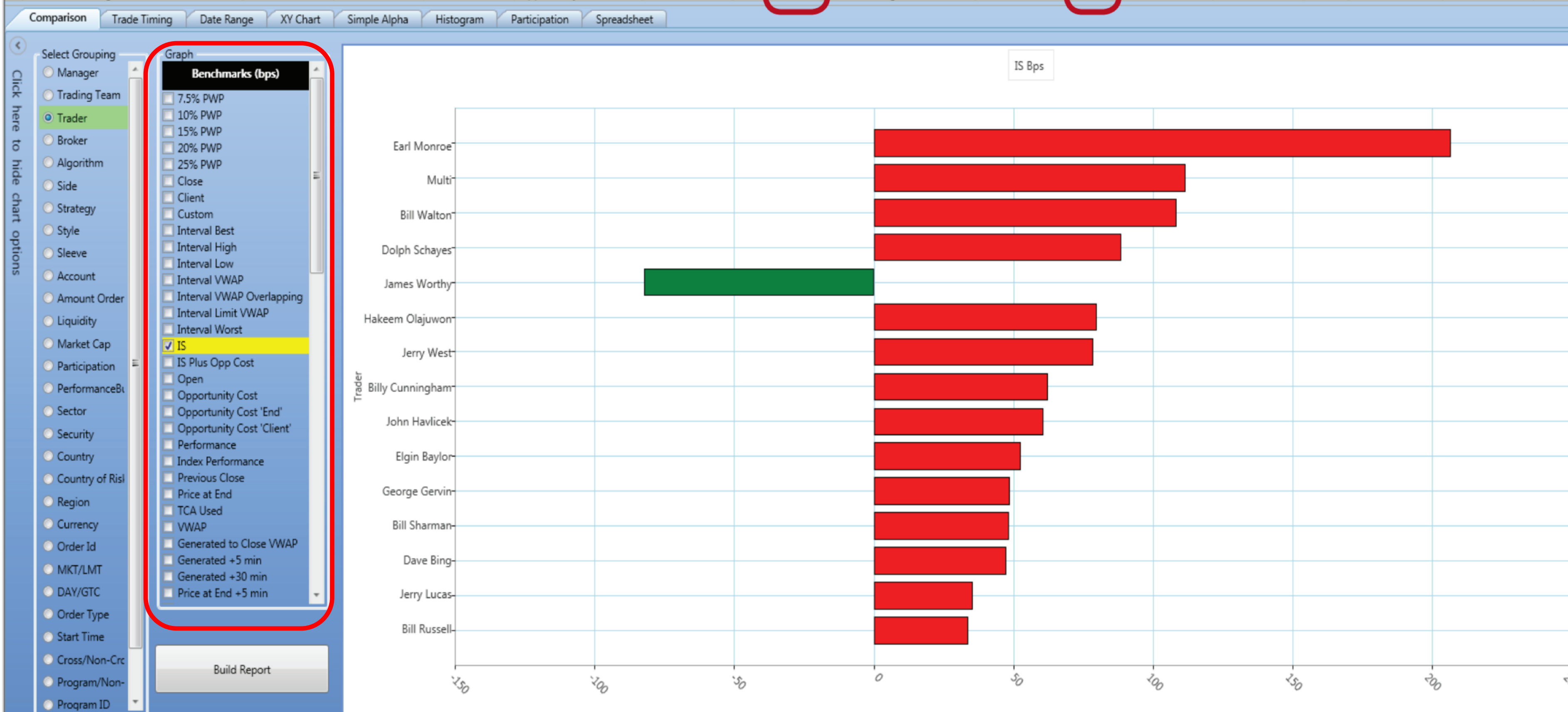
Average Improvement/Shortfall	1.93BPS
Standard Deviation	23.12BPS



Source: LiquidMetrix

Post-Trade TCA - Examples

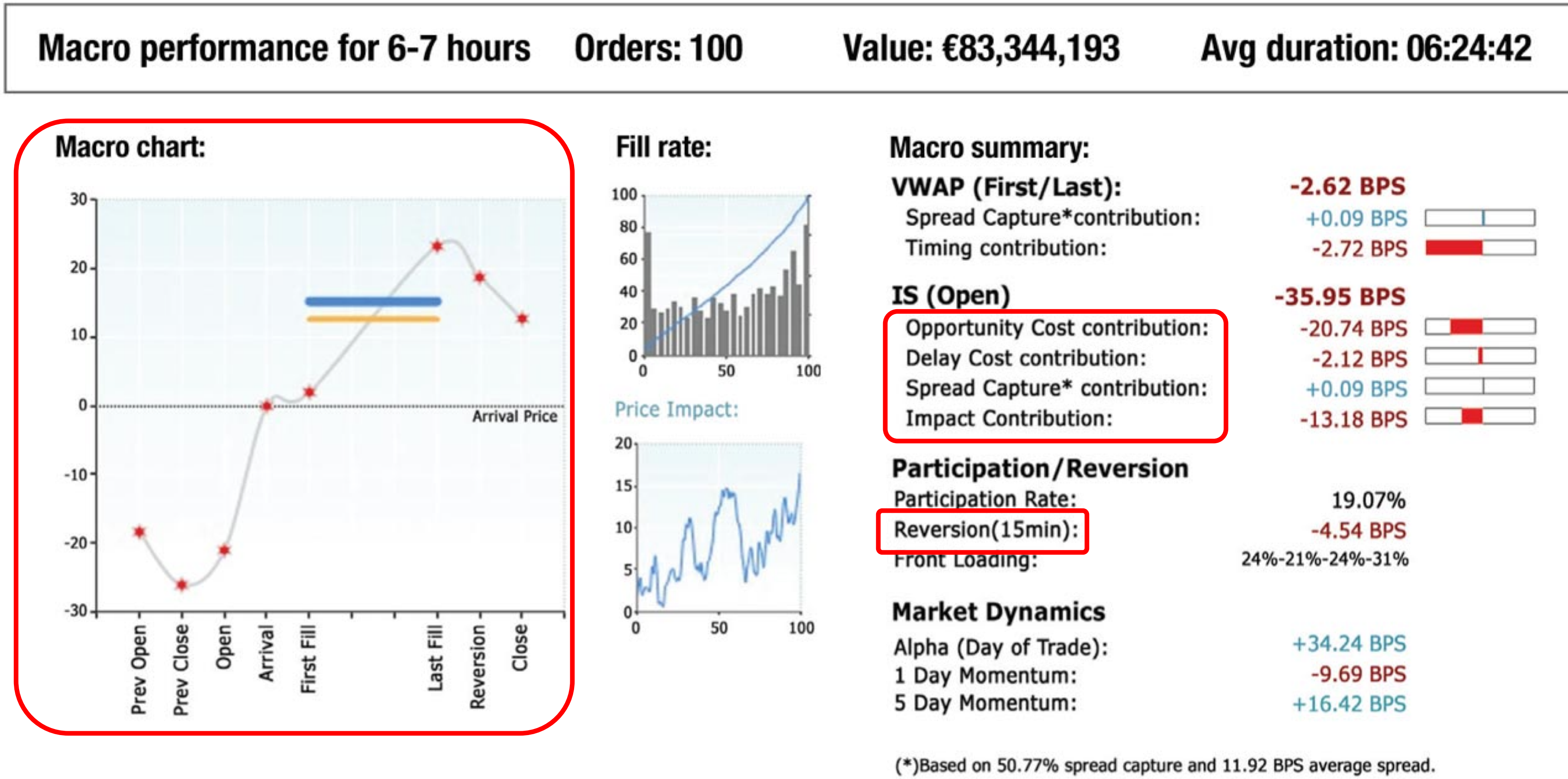
Activity	Total	Buys	Sells	Execution Quality		BPS	Performance		BPS	Timing					
Count	4,710	2,654	2,056	IS	10,390,311	50	vs Interval VWAP	-6,125,763	-30	s 3 Months Before	-21,608,390	-104	vs Next Day	-6,765,176	-32
Ordered Shares	981,927,686	664,780,333	317,147,353	Expected Cost	11,774,811	56	vs Open	-12,374,361	-59	s 1 Month Before	5,755,420	27	vs 2 Days After	-5,514,169	-26
Ordered Amount	2,093,944,861	1,134,473,507	959,471,354	Performance	1,384,501	7	vs Close	-9,006,591	-43	s 1 Week Before	-9,407,428	-45	vs 1 Week After	-8,195,449	-39
Executed Shares	981,927,686	664,780,333	317,147,353	Opportunity Cost	0	0	vs Price at End	-9,216,447	-44	s 2 Days Before	-20,112,450	-95	vs 1 Month After	930,808	4
Executed Amount	2,130,533,949	1,153,471,350	977,062,599	Opportunity Cost 'End'	0	0	vs VWAP	-10,188,174	-49	s 1 Day Before	-14,486,893	-69	vs 3 Months After	8,375,598	39
Order Life Average	388			Opportunity Cost 'Client'	0	0	vs Interval High	739,719	4						



Source: Charles River

Post-Trade TCA – Examples

Figure 2:

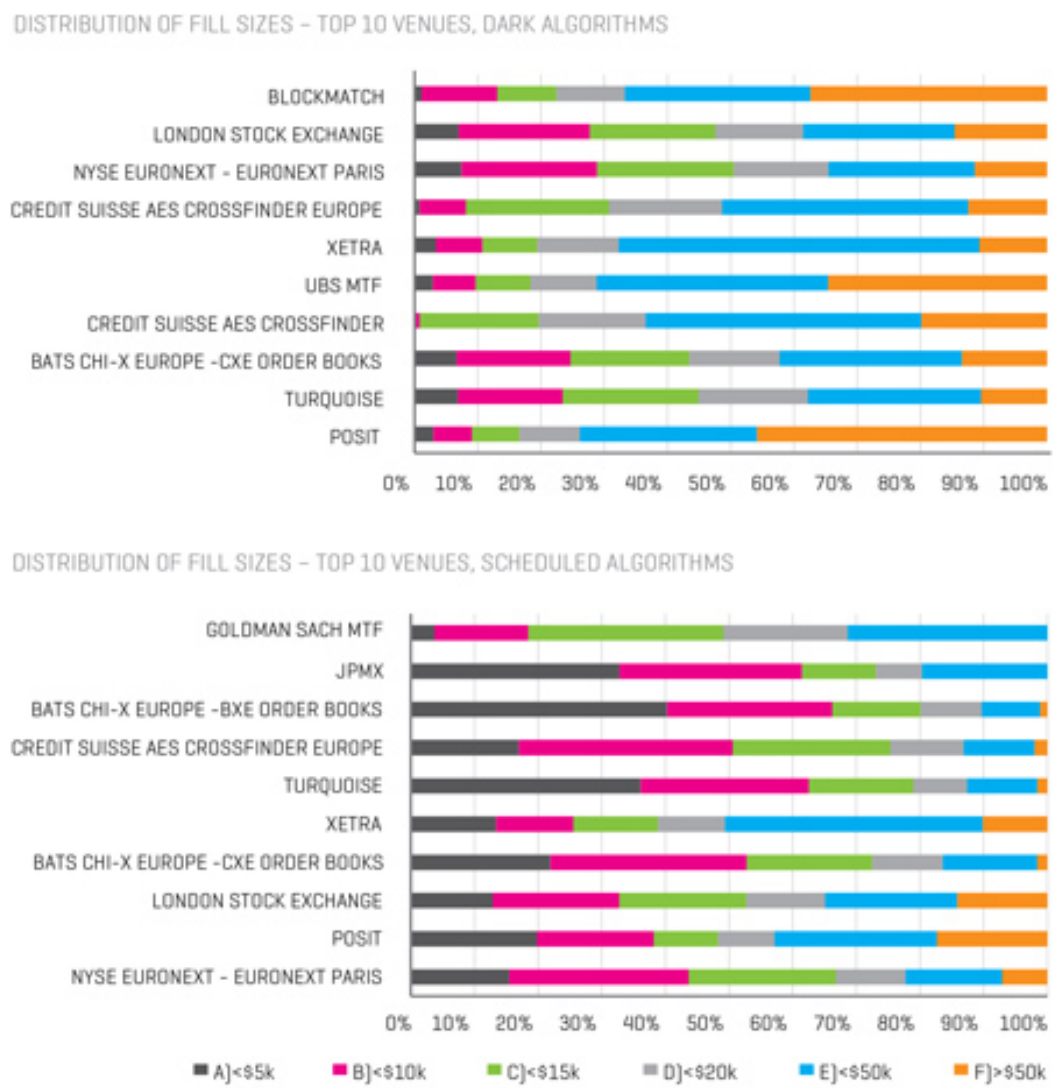


Source: LiquidMetrix

Execution Venue Analysis

- Distribution of Executions between Venues
- Execution Size (bigger is better)
- Slippage (from what?)
- <u>Adverse selection / Reversion</u>

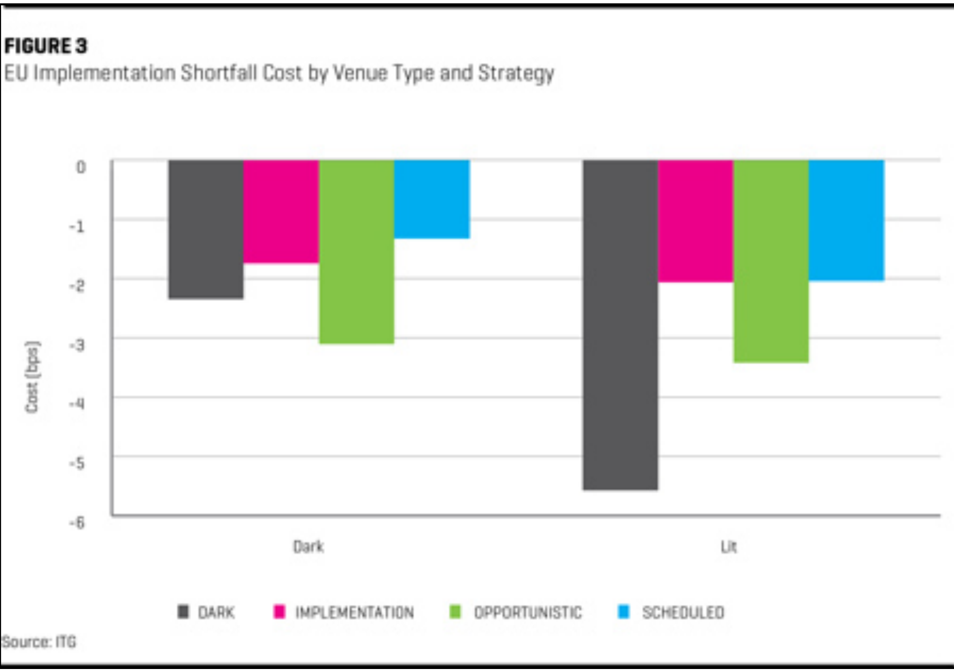
FIGURE 2
EU Examples of Average Fill Size by Venue Type and Strategy

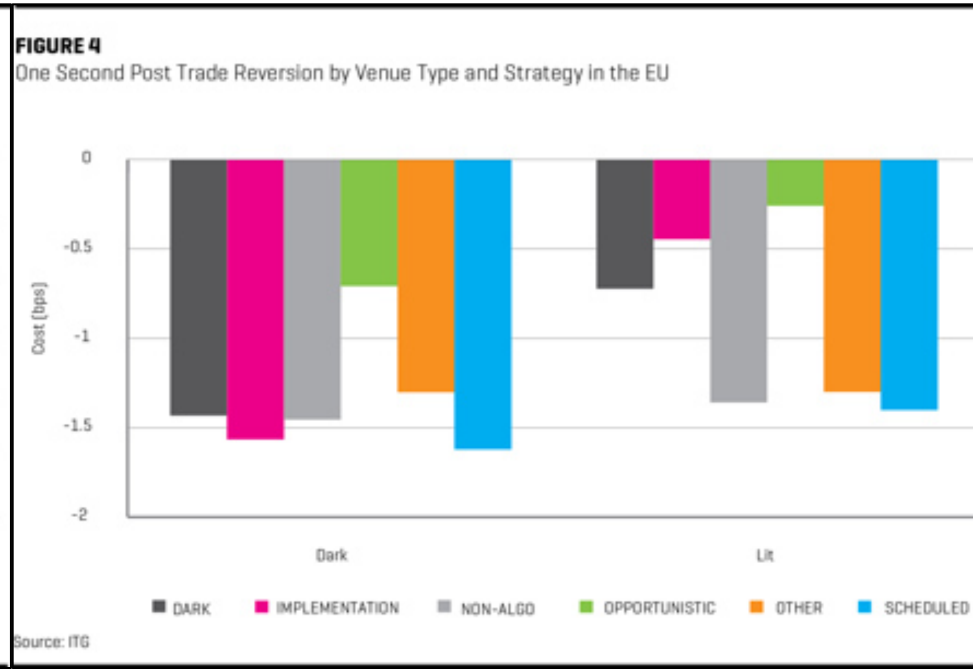


Source: ITG

Source: ITG

Execution Venue Analysis cont'd.





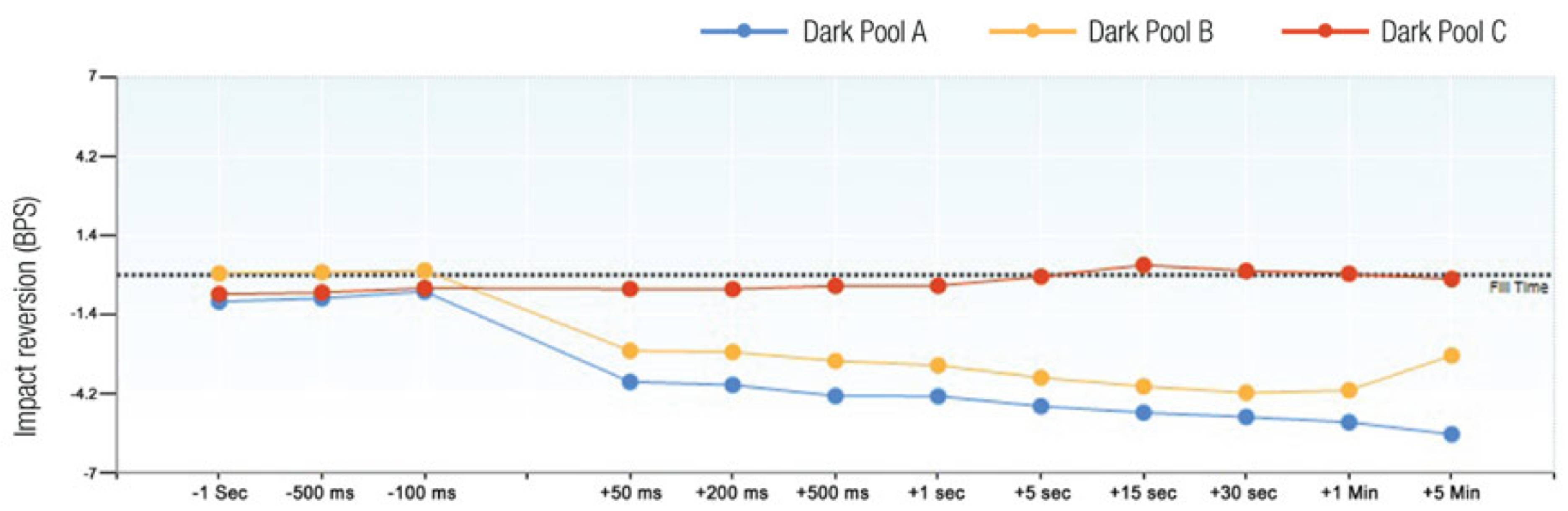
Source: ITG

- Slippage from arrival – causality?
  - Multiple venues
  - Overlapping orders
  - Decoupling timing from order placement is impossible

- Adverse selection / Reversion
  - What happened after the trade?
  - Excess Reversion is bad (could have traded better later)
  - Momentum = information leakage?

Execution Venue Analysis cont’d.

Figure 5: High frequency price movements just before and after fills



Source: LiquidMetrix

Spread Capture

Figure 3: Dark pools – spread capture histogram

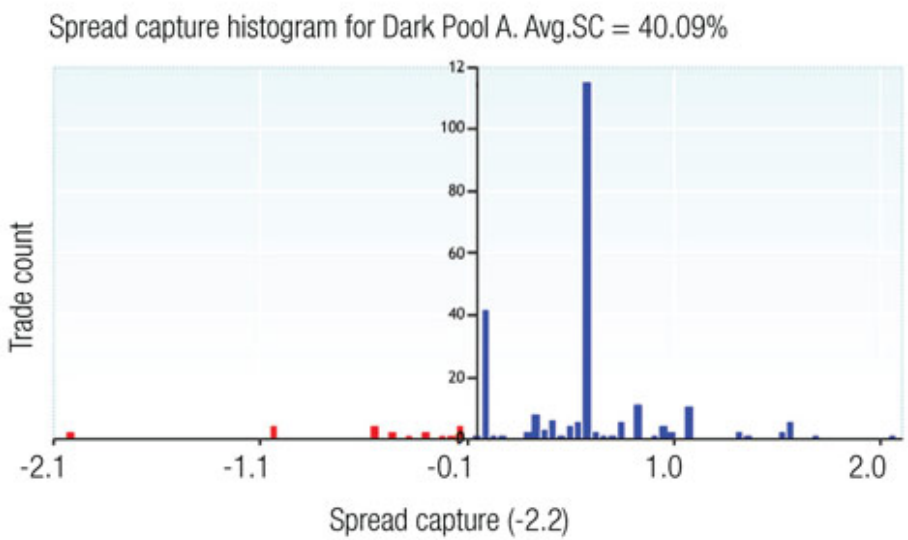
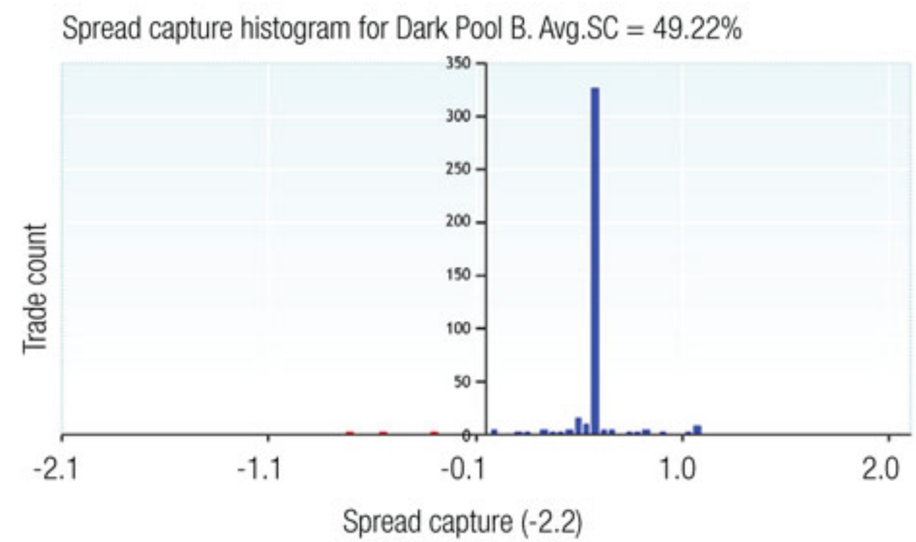


Figure 4: Dark pools – spread capture histogram



Source: LiquidMetrix

- For passive flow (market making, passive strategies)
- What is the impact of order placement?
- Peg Order Types:
  - Passive,
  - Mid,
  - Market / Aggressive

Rebate Capture – Concepts and Issues

- Maker - Taker

  Maker-taker fee model is a pricing structure in which a market generally pays its members a per share rebate to provide (i.e., “make”) liquidity in securities and assesses on them a fee to remove (i.e., “take”) liquidity.2 For example, a maker-taker market may charge $0.003 per share to take liquidity (i.e., 30 cents per 100 shares) and pay a rebate of $0.002 per share to post liquidity (i.e., 20 cents per 100 shares). In this example, the market would earn as its revenue the difference between the two of $0.001 (i.e., 10 cents per 100 shares)

- Inverted markets (Taker - Maker)
- How to choose between venues
  - Regulation NMS Requirements
  - Liquidity
  - Cost

Fee Schedule Examples

	BZX	EDGX	BYX	EDGA
Key Differentiators	Hidden Price Improvement	Retail, Attribution	Retail Price Improvement	Routing, Low Cost
Pricing Model	Traditional "Maker-Taker" model	Traditional "Maker-Taker" model	"Taker-Maker" with low fee and rebates	"Taker-Maker" with low fee and rebates
Exchange Code	Z	K	Y	J
Fee Schedule	View	View	View	View

Source: Bats.com

Fee Schedules: EDGX, EDGA

EDGX

Category	Adding Liquidity	Removing Liquidity	Routing and Removing Liquidity
Securities at or above $1.00	$(0.0020)[1]	$0.0029	$0.0029
Securities below $1.00	$(0.00003)	0.30% of Dollar Value	0.30% of Dollar Value
Standard Fee Codes	B, V, Y, 3, 4	N, W, 6, BB, ZR	X

EDGA

Category	Adding Liquidity	Removing Liquidity	Routing and Removing Liquidity
Securities at or above $1.00	$0.0005[4]	$(0.0002)[1]	$0.0029
Securities below $1.00	FREE	FREE[1]	0.30% of Dollar Value
Standard Fee Codes	B, V, Y, 3, 4	N, W, 6, BB, CR, PR, XR	X

Source: Bats.com

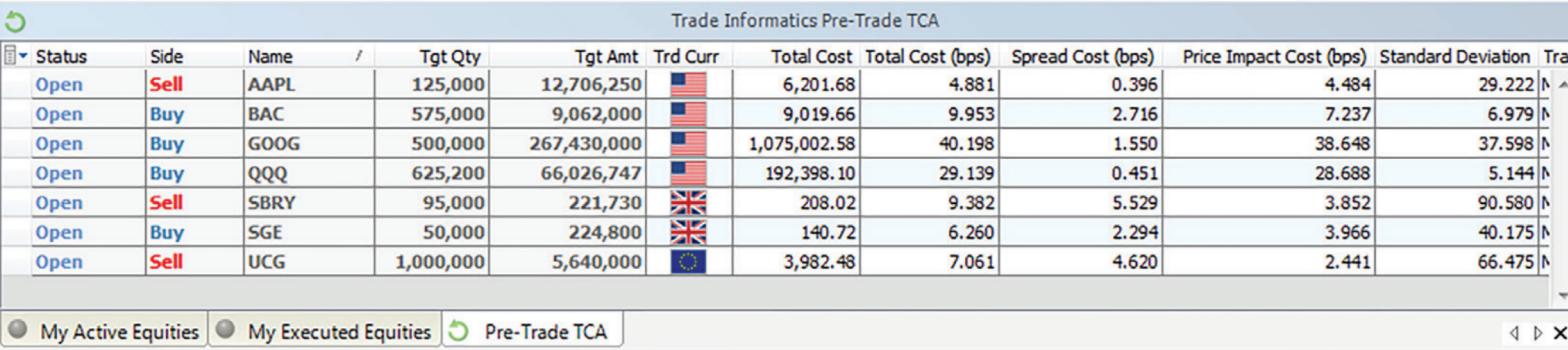
Fee Schedules – EDGX (most of it…)

Fee Code	Description	Fee/(Rebate) Securities at or above $1.00	Fee/(Rebate) Securities below $1.00
10	Routed to NYSE Arca, adds liquidity (Tape B)	(0.00220)	FREE
2	Routed to NASDAQ using INET routing strategy (Tape B)	0.00300	0.30% of Dollar Value
3[1]	Adds liquidity to EDGX, pre and post market (Tapes A or C)	(0.00200)	(0.00003)
4[1,2]	Adds liquidity to EDGX, pre and post market (Tape B)	(0.00200)	(0.00003)
6	Removes liquidity from EDGX, pre and post market (All Tapes)	0.00290	0.30% of Dollar Value
7	Routed, pre and post market	0.00300	0.30% of Dollar Value
8	Routed to NYSE MKT LLC, adds liquidity	(0.00150)	FREE
9	Routed to NYSE Arca, adds liquidity (Tapes A or C)	(0.00210)	FREE
A	Routed to NASDAQ, adds liquidity	(0.00150)	FREE
AA	Routed to EDGA using ALLB routing strategy	(0.00020)	FREE
AY	Routed to BYX using ALLB routing strategy	(0.00150)	0.10% of Dollar Value
AZ	Routed to BZX using ALLB routing strategy	0.00300	0.30% of Dollar Value
B[1,2]	Adds liquidity to EDGX (Tape B)	(0.00200)	(0.00003)
BB	Removes liquidity from EDGX (Tape B)	0.00290	0.30% of Dollar Value
BY	Routed to BYX using Destination Specific ("DIRC"), ROUC or ROUE routing strategy	(0.00150)	0.10% of Dollar Value
C	Routed to BX	(0.00110)	0.10% of Dollar Value
CL[9]	Routed to listing market closing process, except for NYSE Arca & BZX	0.00100	0.30% of Dollar Value
D	Routed to NYSE or routed using RDOT routing strategy	0.00275	0.30% of Dollar Value
EA[7]	Internalization, adds liquidity	0.00045	0.15% of Dollar Value
ER[7]	Internalization, removes liquidity	0.00045	0.15% of Dollar Value
F	Routed to NYSE, adds liquidity	(0.00150)	FREE
G	Routed to NYSE Arca (Tapes A or C)	0.00300	0.30% of Dollar Value

HA[11]	Non-displayed order, adds liquidity	(0.00150)	(0.00003)
HI[11]	Non-displayed order that receives price improvement, adds liquidity	FREE	FREE
I	Routed to EDGA	0.00290	0.30% of Dollar Value
J	Routed to NASDAQ	0.00300	0.30% of Dollar Value
K	Routed to PSX using ROUC or ROUE routing strategy	0.00290	0.30% of Dollar Value
L	Routed to NASDAQ using INET routing strategy (Tapes A or C)	0.00300	0.30% of Dollar Value
MM	Non-displayed order, adds liquidity using Mid-Point Peg	(0.00150)	(0.00003)
N	Removes liquidity from EDGX (Tape C)	0.00290	0.30% of Dollar Value
NA	Routed to BZX, NYSE, NYSE Arca, NYSE MKT or Nasdaq; adds non-displayed liquidity	FREE	FREE
NB	Routed to any exchange not covered by Fee Code NA, adds non-displayed liquidity	0.00300	0.30% of Dollar Value
O[5]	Routed to listing market opening or re-opening cross	0.00100	0.30% of Dollar Value
OO	EDGX Opening or Re-Opening	0.00100	FREE
PR	Removes liquidity from EDGX using ROUQ routing strategy	0.00290	0.30% of Dollar Value
Q	Routed to a non-exchange destination using ROUC routing strategy	0.00200	0.30% of Dollar Value
R	Re-routed by NYSE	0.00300	0.30% of Dollar Value
RA	Routed to EDGA, adds liquidity	0.00050	FREE
RB	Routed to BX, adds liquidity	0.00200	FREE
RN	Routed to NASDAQ using ROOC routing strategy, adds liquidity	(0.00150)	FREE
RP	Non-displayed order, adds liquidity using Supplemental Peg	(0.00150)	(0.00003)
RQ	Routed using ROUQ routing strategy	0.00290	0.30% of Dollar Value
RR	Routed to EDGA using DIRC routing strategy	(0.00020)	0.30% of Dollar Value
RT	Routed using ROUT routing strategy	0.00300	0.30% of Dollar Value
RX	Routed using ROUX routing strategy	0.00300	0.30% of Dollar Value
RY	Routed to BYX, adds liquidity	0.00180	FREE

Source: Bats.com

Pre-Trade TCA



Trade Informatics Pre-Trade TCA

Status	Side	Name	Tgt Qty	Tgt Amt	Trd Curr	Total Cost	Total Cost (bps)	Spread Cost (bps)	Price Impact Cost (bps)	Standard Deviation	Trad
Open	Sell	AAPL	125,000	12,706,250		6,201.68	4.881	0.396	4.484	29.222	N
Open	Buy	BAC	575,000	9,062,000		9,019.66	9.953	2.716	7.237	6.979	N
Open	Buy	GOOG	500,000	267,430,000		1,075,002.58	40.198	1.550	38.648	37.598	N
Open	Buy	QQQ	625,200	66,026,747		192,398.10	29.139	0.451	28.688	5.144	N
Open	Sell	SBRY	95,000	221,730		208.02	9.382	5.529	3.852	90.580	N
Open	Buy	SGE	50,000	224,800		140.72	6.260	2.294	3.966	40.175	N
Open	Sell	UCG	1,000,000	5,640,000		3,982.48	7.061	4.620	2.441	66.475	N

Source: Charles River

Pre-Trade TCA provides portfolio managers with estimates of trade costs and market impact based on extensive historical trade information for a particular name. It evaluates all relevant trade execution strategies and recommends the strategy that is most consistent with a manager's risk preferences.

- Pre-Trade TCA is popular in program trading
- Some pre-trade analyses include correlation analysis
- Coupled with program trading services / algos to help construct optimal and/or dynamic trade profiles

TCA Considerations and Best Practices

- Normalizing results across brokers, algorithms and market conditions
- Broker specific report formats (we have seen several just today)
- Arrival price: what is it?
  - Last trade? Midpoint at arrival? Weighted?
  - What if the stock hasn't traded yet at arrival? (pre-open)
  - What if stock is illiquid? Is last trade or current midpoint better?
- VWAP
  - What prints are included / excluded?
  - Primary (common in Europe)
  - Consolidated (better in US?)
  - Is Auction volume included?
- Measurements of reversion / adverse selection
  - Time resolution of price movement – seconds, milliseconds?
- Venue identifiers (last Market – includes routed orders?)

TCA Considerations – Data Analytics

Key challenges in data management and analytics

- Large volume of data
  - Billions of records on stock quotes and trades
  - Large volume of client orders and executions with millions on records
- Trade-Quote Matching
  - Which trade refers to which quote? Not a simple match!
  - Lee-Ready algorithm a popular choice – Add 5 second delay to quote to match a fill
- Real-time analytics
  - Live market data feed to perform benchmark analysis
  - Latency in both market data feed and client order data