Performance Metrics for Algorithmic Traders

Dale W.R. Rosenthal

University of Illinois at Chicago, Department of Finance

CQA Spring 2010 Conference
14 April 2010

Introduction

- Trading changed rapidly over past decade. Now see:
  - More matching of buyers and sellers by computer; and,
  - Automation of many tactical trading decisions.
- This is a great triumph for microstructure:
  - Microstructure research suggests optimal trading decisions.
- One particular change: how large orders are handled.
  - Less block trading, more slicing into smaller orders.
- Key question: how well did we do at trading such orders?

UIC

Splitting Orders

- Kyle (1984): split orders to hide private information (alpha)
- Bertsimas and Lo (1998): split to reduce trading costs.
- Almgren and Chriss (2001):
  - Split orders to minimize mean-variance trading costs.
- Engle and Ferstenberg (2006): also considered time risk.
- No surprise to you: many orders now split.
- Berke (2010) estimates 30% of volume from split orders.

Terminology

Need to be clear about terminology.

- E & F: slice *parent orders* into schedule of *child orders*.
- Call a collection of parent orders a *portfolio order*.
- *Algorithmic trading*: automated order creation, management.
- *Internal*[1] use: performance auditor sees full info.
    - Knows about unsent orders; can see gaming attempts.
    - *e.g.* Fund strategist optimizing in-house trading engine.
- *External* use: performance auditor lacks full info.
    - External metrics must be resistant to gaming.
    - *e.g.* Fund manager examining external execution providers.

[1] *Internal* vs *external* is as in Lehmann (2003).

Implementation Shortfall

- Common metric: Perold's (1988) *implementation shortfall*.
  - Portfolio order metric; often computed for parent orders.
  - Parent order traded* value − order starting value.
- Problem 1: * unfilled quantity priced at ending midpoint.
- Problem 2: For large orders, we expect price impact.
- (Weak) gaming: claim end time which lowers unfilled cost.

UIC

Measuring Performance

- Instead, examine multiple dimensions of performance:
  - Was a portfolio/parent order treated with discretion?
  - Was a desired child order schedule *ex post* optimal?
  - Did we benefit or suffer by deviating from this schedule?
  - Did we do well at trading the given child orders?
  - Or, were we just lucky?
- Keep in mind: are metrics gameable?
  - Some are gameable, suitable only for internal use;
  - Others resist gaming, also suitable for external use.

UIC

Decompositions

Types of Decompositions

Answer questions with three types of metrics:

1. *Full Implementation Shortfall*:
    - conservative re-definition of implementation shortfall.
2. *Parent Order Metrics*, measuring:
    - information leakage, adverse selection, price impact.
3. *Intertemporal Metrics*, using child orders to measure:
    - different types of skill versus luck/noise.

I assume no alpha to ease math; could adjust for alpha.
Also assume market impact precludes (dynamic) arbitrage.

Full Implementation Shortfall

- Conservative: satisfy unfilled quantity from order book.
- Thus the Bounded Full Implementation Shortfall $BFIS_t$:

$$BFIS_t = \overbrace{\underbrace{\tilde{q}_t(\bar{p}_t - p_0)}_{\text{Realized IS}} + \underbrace{(q - \tilde{q}_t)(p_{\text{mid},t} - p_0)}_{\text{Simple Opportunity Cost}}}^{\text{Implementation Shortfall } IS_t} + \underbrace{(q - \tilde{q}_t)(\bar{p}_{\text{walk},t} - p_{\text{mid},t})}_{\text{Bounded Immediacy Cost } BIC_t} \quad (1)$$

where $q, \tilde{q}_t$ = signed order shares, shares filled by time t

$p_t, \bar{p}_t$ = price at, average fill price up to time t

$\bar{p}_{\text{walk},t}$ = average price to fill $q - \tilde{q}_t$ from order book.

UIC

VWAP

- VWAP: volume-weighted average price; common benchmark.
- Experience, Opiela (2006): cannot beat VWAP without alpha.
- For doubters: Only need this to hold over short timespans.

Proposition (VWAP is Fair)

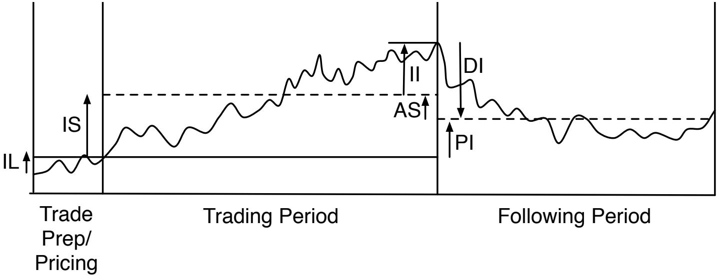
Assuming no alpha and arbitrage-free market impact, VWAP is a fair metric, i.e. it cannot be beaten in expectation.

Proof (Sketch).

For 1, 2 traders: implied by arbitrage-free market impact, no info on others' orders, VWAP being average of fair (arb-free) prices.
For 3+ traders: Result follows by induction. □

UIC

Parent Order Metrics: Diagram

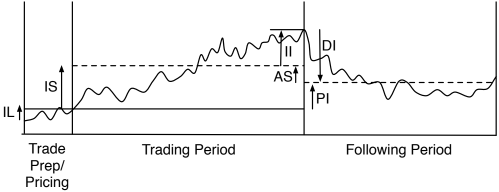


Parent Order Metrics. Dashed lines represent average fill price (trading period) and next-period VWAP.

Note decomposition of IS: $IS_T = PI + AS = PI + DI - II$.

UIC

Parent Order Metrics



- *Information Leakage*: $IL = q(p_0 - p_-)$
  - increased cost from idea/revelation to trading start.
- *Incremental Impact*: $II = \tilde{q}(p_T - \bar{p}_T)$
  - average fill price to end-of-trading value change.
- *Adverse Selection*: $AS = \tilde{q}(\bar{p}_T - \check{p}_+)$
  - average fill price to next-period VWAP value change.
- *Decaying Impact*: $DI = \tilde{q}(p_T - \check{p}_+)$
  - end-of-trade to next-period VWAP value change.
- *Permanent Impact*: $PI = \tilde{q}(\check{p}_+ - p_0)$
  - trading start to next-period VWAP value change.

UIC

Intertemporal Metrics

- Many orders traded vs. benchmark (*e.g.* VWAP, current mid).
- Can look at benchmark-relative shortfall, similar to IS.
- Break time into bins (j); consider scenarios. What if:
  - each child order had been filled at bin VWAP?
  - fills had followed planned execution schedule?
  - fills had followed realized volume schedule?

Intertemporal Metrics: Decompositions

$$FBS = \underbrace{(q-\tilde{q})(\bar{p}_{\text{walk},T} - p_{\text{mid},T})}_{\text{Immediacy Cost}} + \underbrace{(q-\tilde{q})(\bar{p}_{\text{mid},T} - p_{\text{bmk}})}_{\text{Opportunity Cost}} + \underbrace{\tilde{q}\bar{p}_T - \tilde{q}p_{\text{bmk}}}_{\text{Realized Benchmark Shortfall } RBS} \quad (2)$$

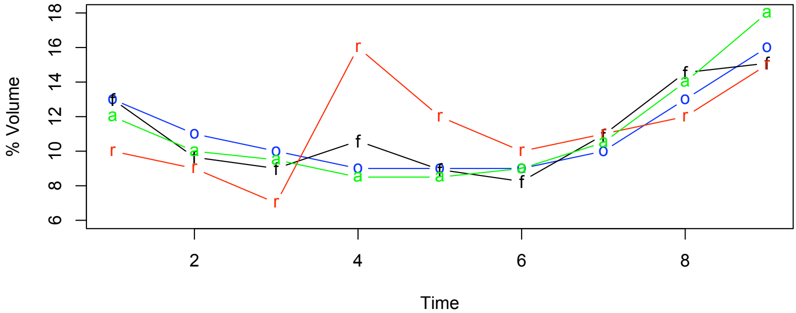
- Full Benchmark Shortfall *FBS* like Full *IS*, *RBS* like *RIS*.
- Decompose Realized Benchmark Shortfall *RBS*, like *IS*.

$$RBS = \underbrace{\sum_j \tilde{q}_j(\bar{p}_j - \check{p}_j)}_{\text{Trading Shortfall}} + \underbrace{\sum_j (\tilde{q}_j - \tilde{q}\frac{q_j}{q})\check{p}_j}_{\text{Fill Time Shortfall}} + \underbrace{\sum_j \tilde{q}(\frac{q_j}{q} - D_j)\check{p}_j}_{\text{Order Timing Shortfall}} + \underbrace{\sum_j \tilde{q}(D_j - \bar{D}_j)\check{p}_j}_{\text{Volume Shortfall}} + \underbrace{\sum_j \tilde{q}(\bar{D}_j\check{p}_j - p_{\text{bmk}})}_{\text{Simple VWAP Shortfall}} \quad (3)$$

UIC

Comparing Schedules

Decompositions compare trading using different schedules.



Order and Volume Schedules.

Blue o's $= q_j/q$, orders planned;
Black f's $= \tilde{q}_j/\tilde{q}$, actual fills;
Red r's $= D_j = V_j/V$, realized volume;
Green a's $= \bar{D}_j = E(V_j/V)$, average volume.

Intertemporal Metrics: Explanations

$$RBS = \underbrace{\sum_j \tilde{q}_j(\bar{p}_j - \check{p}_j)}_{\text{Trading Shortfall}} + \underbrace{\sum_j (\tilde{q}_j - \tilde{q}\frac{q_j}{q})\check{p}_j}_{\text{Fill Time Shortfall}} + \underbrace{\sum_j \tilde{q}(\frac{q_j}{q} - D_j)\check{p}_j}_{\text{Order Timing Shortfall}} + \underbrace{\sum_j \tilde{q}(D_j - \bar{D}_j)\check{p}_j}_{\text{Volume Shortfall}} + \underbrace{\sum_j \tilde{q}(\bar{D}_j\check{p}_j - p_{\text{bmk}})}_{\text{Simple VWAP Shortfall}} \tag{4}$$

- *Trading Shortfall*: due to fills worse than bin VWAPs.
- *Fill Time Shortfall*: due to fill times other than planned.
    - Note the counterfactual: fills at bin VWAPs.
- *Order Timing Shortfall*: order plan vs. actual volume dist.
- *Volume Shortfall*: due to variation of volume distribution.

UIC

Intertemporal Metrics: Comments

- If $p_{\text{bmk}} = p_0$: $FBS = FIS$, $BS = IS$.
- *Fill Time Shortfall*: measures cost of deviating from plan.
- *Order Timing Shortfall*: measures order scheduling skill.
  - *Caveat*: We affect actual volume; separate our volume.

$$OTS = \underbrace{\sum_j \tilde{q} \frac{\tilde{q}_j}{V_j} (\frac{q_j}{q} - D_j) \check{p}_j}_{\text{Endogenous}} + \underbrace{\sum_j \tilde{q} \frac{V_j - \tilde{q}_j}{V_j} (\frac{q_j}{q} - D_j) \check{p}_j}_{\text{Exogenous}} . \quad (5)$$

UIC

Analysis

Price Impact of Trading

- Analyze these metrics in light of a price impact model.
- Want arbitrage-free model, *cf* Huberman and Stanzl (2004).
- Recall: We split orders to allow liquidity to replenish.
- Use Obizhaeva and Wang (2010) model.
  - Replenishing order book ⇒ some impact decays to 0.
- Impact has 3 (or 4) components:

$$E(\bar{p}_j) = p_0 + \underbrace{\sum_{k=1}^{j} \pi \tilde{q}_k}_{\text{permanent}} + \underbrace{\sum_{k=1}^{j} \delta^{j+1-k} \tilde{q}_k}_{\text{decaying}} + \underbrace{\tau \frac{\tilde{q}_j}{t_j} + \phi \,\text{sgn}(\tilde{q}_j)}_{\substack{\text{temporary} \\ \text{(only trader pays)}}} . \quad (6)$$

- Careful action can reduce decaying, temporary effects.

UIC

Analysis: Parent Order Metrics (*RIS* and *II*)

If we choose t_j s.t. $\tilde{q}_j = \frac{\tilde{q}}{n}$, then get:

- Implementation Shortfall: combination of all impact forms.

$$E(RIS) = \pi\tilde{q}\frac{n+1}{2n} + \frac{\tilde{q}\delta}{n(1-\delta)} + \tau\frac{\tilde{q}}{n^2}\sum_{j=1}^{n}\frac{1}{t_j} + \phi\,\mathrm{sgn}(q) + o(1/n) \quad (7)$$

- Incremental Impact: combines permanent, temporary impact.

$$E(II) = \pi\tilde{q}\frac{n-1}{2n} - \tau\frac{\tilde{q}}{n^2}\sum_{j=1}^{n}\frac{1}{t_j} - \phi\,\mathrm{sgn}(q) + o(1/n) \quad (8)$$

- *N.B.* Variances in paper; no distributional assumptions.

UIC

Analysis: Parent Order Metrics (AS, DI, PI)

Assuming next-period volume distribution is non-degenerate:

- Adverse Selection: combines all impact forms.

$$E(AS) = \frac{\tilde{q}\delta}{n(1-\delta)} - \pi\tilde{q}\frac{n+1}{2n} + \tau\frac{\tilde{q}}{n^2}\sum_{j=1}^{n}\frac{1}{t_j} + \phi\,\text{sgn}(q) + o(1/n) \tag{9}$$

- Decaying Impact: related only to decaying impact.

$$E(DI) = \frac{\tilde{q}\delta}{n(1-\delta)} + o(1/n) \tag{10}$$

- Permanent Impact: related mostly to permanent impact.

$$E(PI) = \pi\tilde{q} + o(1/n) \tag{11}$$

Analysis: Intertemporal Metrics (TS)

- Trading Shortfall: related only to temporary impact.

$$E(TS) = \sum_{j=1}^{n} \tilde{q}_j(\tau \frac{|\tilde{q}_j|}{t_j} + \phi \, \mathrm{sgn}(q))(1 - \frac{|\tilde{q}_j|}{V_j}) \qquad (12)$$

- Other metrics not so cleanly related to impact.
- However, other metrics may be stated as covariances.

Analysis: Intertemporal Metrics as Covariances

- Fill Time Shortfall: Cov(overfills, worse prices)

$$E(FTS) = \tilde{q}\,\mathrm{Cov}(\frac{\tilde{q}_\cdot}{\tilde{q}} - \frac{q_\cdot}{q}, \check{p}_\cdot) \tag{13}$$

- Order Timing Shortfall: Cov(larger orders, worse prices)

$$E(OTS) = \tilde{q}\,\mathrm{Cov}(\frac{q_\cdot}{q} - D_\cdot, \check{p}_\cdot) \tag{14}$$

- Volume Shortfall: Cov(volume surprises, worse prices)

$$E(VS) = \tilde{q}\,\mathrm{Cov}(D_\cdot - \bar{D}_\cdot, \check{p}_\cdot) \tag{15}$$

- Can also look across instruments to study each bin.

UIC

Recovering Impact Model Parameters

- Note that DI, PI, and TS are very clean in form.
  - *Caveat*: DI is not robust to gaming. (More later.)
- Recover impact model parameters via regression, rewriting.
- The β_0's are nuisance parameters.
- Could also add bias terms ($O(1/n^2)$, etc.).

$$PI = \beta_{0,PI} + \pi\tilde{q} + \epsilon_{PI} \tag{16}$$

$$DI = \beta_{0,DI} + \frac{\delta}{1-\delta}\tilde{q} + \epsilon_{DI} \tag{17}$$

$$TS_j = \beta_{0,TS} + \tau\tilde{q}_j\frac{|\tilde{q}_j|}{t_j}\left(1 - \frac{|\tilde{q}_j|}{V_j}\right) + \phi\tilde{q}_j\left(1 - \frac{|\tilde{q}_j|}{V_j}\right) + \epsilon_{TS} \tag{18}$$

UIC

Interpretation

A Note on Gaming

- Noted earlier: some measures only suitable for internal use.
  - These are metrics which may be gamed in subtle ways.
  - Extra care should be taken if they are used externally.
- Other metrics are gaming-resistant, better for external use.
  - That may still be gamed, but. . .
  - The effect is either obvious or small at most.

UIC

Interpretation of *IL*, *IS*

- Information Leakage *IL*: good for external use.
  - Yields a t-test for possible front-running:

$$t = \frac{\text{sgn}(q)(p_0 - p_-)}{\sigma_p \sqrt{t_0 - t_-}} \tag{19}$$

  where σ_p is *price* volatility $(= p\sigma_r)$.
- Implementation Shortfall *IS*: unclean for performance tuning.
  - However, *IS* and *FIS* are applicable to all orders.
  - Pricing of unfilled quantity may be slightly gamed.
  - *IS*: briefly narrow spread; nudge claimed end time.
  - *FIS*: briefly fill far side of book; nudge claimed end time.

UIC

Interpretation of *II*, *DI*

- Incremental Impact *II*: where we leave the market.
  - High *II*: may have attracted liquidity providers. (Bad.)
  - Maybe should have traded over longer period; or,
  - Last orders were too aggressive. (Why "get done"?)
  - Very gameable: "end" time affects whole metric.
- Decaying Impact *DI*: more direct eponymous measure.
  - High *DI* suggests trading over longer period; or,
  - Chose poor times to send child orders.
  - Very gameable: "end" time affects whole metric.
- Ease of gaming *II*, *DI* suggests only using them internally.

Interpretation of PI, AS

- Permanent Impact PI: measures inescapable impact.
  - Should expect PI to be consistent over time.
  - May be useful to measure effects of market changes.
- Adverse Selection AS: depends on all impact forms.
  - Like different ways models impound such fears?
  - High AS should suggest high adverse selection cost.
  - However, not so clear with this impact model.
- PI, AS: gaming-resistant (use of average prices).
  - Liquidity provision skews next period? Look farther ahead.
  - Thus may be suitable for external use.

Single Parent Order Metric?

- Is there a *portmanteau* parent order metric for tuning?
- Maybe. Can combine PI and AS:

$$E(AS + PI\frac{n+1}{2n}) = \frac{\tilde{q}\delta}{n(1-\delta)} + \tau\frac{\tilde{q}}{n^2}\sum_{j=1}^{n}\frac{1}{t_j} + \phi\,\mathrm{sgn}(q) + o(1/n) \tag{20}$$

- Intuition behind this?

UIC

Interpreting Trading Shortfall

- Trading Shortfall TS: clean measure of trading skill.
  - Good traders have consistently small TS.
  - Disciplined but bad: consistently large TS.
  - Sloppy: noisy/inconsistent TS.
- TS may even indicate front-running.
  - Front-runner position accumulation, disposal biases TS.
  - Would see high TS earlier, low TS later; can test this:

$$P(b \text{ of } n/2 \text{ worst } TS_j\text{'s in first half}) = \binom{n/2}{b} \frac{1}{2^{n/2}}. \quad (21)$$

- Might also use TS if alpha traded without Kyle model.
- Gaming (obvious): If all volume in bin j, $TS_j = 0$.
- Gaming (subtle): letting external provider define bin times.

UIC

Interpreting Fill Time, Order Timing Shortfalls

- Fill Time Shortfall *FTS*: skill of gauging aggressiveness.
  - Good ("cool hand"): consistently low *FTS*.
  - Too passive: low *FTS* earlier, high *FTS* later.
- Order Timing Shortfall *OTS*: order scheduling skill.
  - Good: consistent and/or low *OTS*.
  - (Some benchmarks schedule orders to lower variance.)
- Volume Shortfall *VS*: tough to interpret; noise.
  - Unless one has skill at predicting volume surprises. (!)
- Gaming *FTS*, *OTS* is tough if bin times pre-defined.
- Gaming *VS* pointless. (*VS* is noise; ignore it anyway.)

Conclusion

- Introduced new metrics for algorithmic traders.
  - More informative than *IS* for order slicers.
- Some metrics relate to realistic impact model parameters.
- A few allow recovery of model parameters.
- More detailed metrics allow us to measure different skills:
  - trading vs. patience vs. scheduling vs. luck/noise.
  - Can separate trading, scheduling skill (people or software).
  - Can see where outside execution providers excel (or not).
- Found tests for possible front-running, information dissipation.
- Extension: relate performance variation to other schedules.
  - *e.g.* surprises in volatility, spread, depth, volume.
- Demand your fills, timestamps: gold in them thar fills.

Comments: daler@uic.edu More info: www.uic.edu/∼daler