Performance Metrics for Algorithmic Traders

Dale W.R. Rosenthal[1]

University of Illinois at Chicago, Department of Finance

26 January 2012

[1]daler@uic.edu; tigger.uic.edu/∼daler

Introduction

- Trading changed rapidly over past decade. Now see:
  - More matching of buyers and sellers by computer; and,
  - Automation of many tactical trading decisions.
- Microstructure research guides optimal trading decisions.
- One particular change: how large orders are handled.
  - Block trades rarer, now often sliced into smaller orders.
- US 2010: \$13.4 bn spent on trading infrastructure.[2]
- Key question: What is the value of trading infrastructure?
- Implementation shortfall (IS) is too blunt to measure this.

[2] This includes \$1 bn on smart order routing to best prices.

Results Preview

- Proposed metrics more informative than *IS* for order slicers.
- Relate to price impact model (and allow parameter recovery).
- Can evaluate people or automated trading processes.
  - Moved prices too much at end of trading?
  - Execution, timing, or scheduling skill. . . or luck/noise?
- Find tests for front-running/information dissipation.
- Estimated savings: up to 4 bp/trade = 15% lower expenses.
  - $\Rightarrow$ \$7.3 bn/year savings for US equity mutual funds.

UIC Liautaud

Splitting Orders

- Kyle (1985): split orders to hide private information (alpha)
- Bertsimas and Lo (1998): split to reduce trading costs.
- Almgren and Chriss (2001): minimize mean-variance cost.
- Engle and Ferstenberg (2007): portfolio choice affected.
  - Optimize portfolio and order splitting together.
- 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* use: performance auditor sees full info.[3]
  - 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.

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

UIC Liautaud

Measuring Performance

- Common metric: Perold's (1988) *implementation shortfall*.
  - Parent order traded value − order starting value.[4]
- Instead, ask multiple counterfactual questions.
  - Some relate to parts of the trading process (*e.g.* software).
- Keep in mind: are metrics gameable?
  - Some are gameable, suitable only for internal use;
  - Others resist gaming, also suitable for external use.

[4] Assuming trading completed.

UIC Liautaud

Types of Decompositions

Answer questions with two types of metrics:

1. *Parent Order Metrics*, measuring:
   - information leakage, adverse selection, price impact.
2. *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.

But we first need a fair price for a time period.

UIC Liautaud

VWAP is Fair

- 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 Liautaud

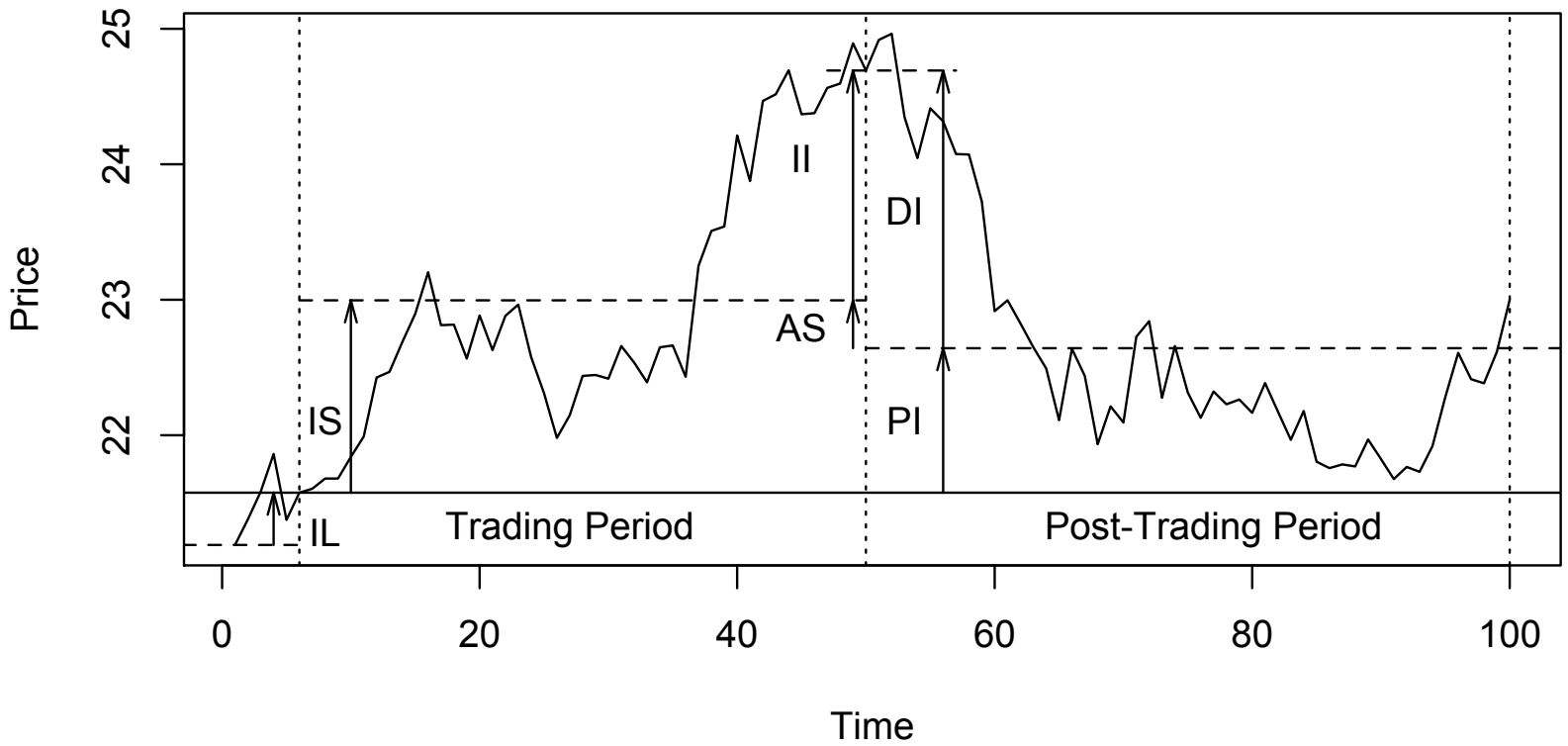
Parent Order Questions

Metrics derived from asking about counterfactuals:

- What if trading began when somebody external saw our order?
- What was the marginal (incremental) cost of our last trade?
- What is the profit of providing liquidity to that last trade?
- What is the lasting effect our trading had on prices?
- How much worse did we do than that lasting effect on prices?

UIC Liautaud

Parent Order Metrics: Diagram

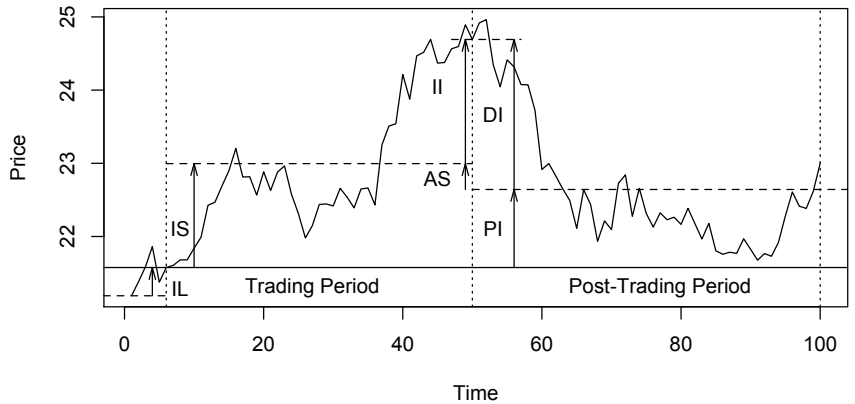


Parent Order Metrics. Dashed lines represent average fill price (trading period) and next-period VWAP.

Note relationships: $IS = PI + AS = PI + DI - II$.

UIC Liautaud

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 Liautaud

$q, \tilde{q}$ = shares ordered, filled $\quad p_t, \bar{p}_t$ = price, average fill price at t

Intertemporal Metrics

- Can decompose benchmark-relative shortfall (*e.g. IS*).
- Break time into contiguous bins (j).
- Then ask about counterfactuals related to decisions/software.
- What if:
  - child orders filled at fair prices (bin VWAPs)? (MS SORT)
  - child orders filled when scheduled? (urgency settings)
  - fills followed average volume distribution? (MS BXS)
  - fills followed realized volume distribution? (noise)
- Since this is a decomposition, metrics are not orthogonal.

Intertemporal Metrics: Decompositions

Decompose Realized Implementation Shortfall *RIS*:

$$RIS = \underbrace{\sum_j \tilde{q}_j(\bar{p}_j - \check{p}_j)}_{\text{Trading Shortfall}} + \underbrace{\sum_j (\tilde{q}_j - \tilde{q}\frac{q_j}{q})\check{p}_j}_{\text{Fill Time Shortfall}} + \underbrace{\sum_j \tilde{q}(\frac{q_j}{q} - \bar{D}_j)\check{p}_j}_{\text{Order Timing Shortfall}}$$
$$+ \underbrace{\sum_j \tilde{q}(\bar{D}_j - D_j)\check{p}_j}_{\text{Volume Shortfall}} + \underbrace{\sum_j \tilde{q}D_j\check{p}_j - \tilde{q}p_0}_{\text{Perfect VWAP Shortfall}} \quad (1)$$

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

UIC Liautaud

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 (2011) model.
  - Replenishing order book $\Rightarrow$ some impact decays to 0.
- Impact has 3 (or 4) components:

$$E(\bar{p}_j) = p_0 + \underbrace{\sum_{k=1}^{j} \pi \tilde{q}_k}_{\text{permanent}} + \underbrace{\sum_{k=1}^{j} \delta^{j+1-k} \tilde{q}_k}_{\text{decaying}} + \underbrace{\tau \frac{\tilde{q}_j}{t_j} + \phi[\![\tilde{q}_j]\!]}_{\substack{\text{temporary} \\ \text{(only trader pays)}}} . \quad (2)$$

- Careful action can reduce decaying, temporary effects.

UIC Liautaud

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

If we choose t_j s.t. $\tilde{q}_j = \frac{\tilde{q}}{n}$, then get:[5]

- Implementation Shortfall: combination of all impact forms.

$$E(IS) = \pi\tilde{q}\frac{n+1}{2n} + \frac{\tilde{q}\delta}{n(1-\delta)} + \tau\frac{\tilde{q}}{n^2}\sum_{j=1}^{n}\frac{1}{t_j} + \phi[\![q]\!] + o(\tfrac{1}{n}) \quad (3)$$

- Incremental Impact: combines permanent, temporary impact.

$$E(II) = \pi\tilde{q}\frac{n-1}{2n} - \tau\frac{\tilde{q}}{n^2}\sum_{j=1}^{n}\frac{1}{t_j} - \phi[\![q]\!] + o(\tfrac{1}{n}) \quad (4)$$

[5] *N.B.* Variances in paper; no distributional assumptions.

UIC Liautaud

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

If next-period volume distribution not degenerate:

- Adverse Selection: combines all impact forms.

$$E(AS) = \frac{\tilde{q}\delta}{n(1-\delta)} - \pi\tilde{q}\frac{n+1}{2n} + \tau\frac{\tilde{q}}{n^2}\sum_{j=1}^{n}\frac{1}{t_j} + \phi[\![q]\!] + o(\tfrac{1}{n}) \quad (5)$$

- Decaying Impact: eponymous, related to decaying impact.

$$E(DI) = \frac{\tilde{q}\delta}{n(1-\delta)} + o(\tfrac{1}{n}) \quad (6)$$

- Permanent Impact: eponymous, related to permanent impact.

$$E(PI) = \pi\tilde{q} + o(\tfrac{1}{n}) \quad (7)$$

- Note: IS, IS, AS are amalgams; DI, PI are very clean.

UIC Liautaud

Analysis: Intertemporal Metrics (TS)

- Trading Shortfall: related only to temporary impact.

$$E(TS) = \sum_{j=1}^{n} \tilde{q}_j (\tau \frac{|\tilde{q}_j|}{t_j} + \phi[\![q]\!])(1 - \frac{|\tilde{q}_j|}{V_j}) \tag{8}$$

- 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)

$$E(FTS) = \tilde{q}\,\mathrm{Cov}(\frac{\tilde{q}_{\cdot}}{\tilde{q}} - \frac{q_{\cdot}}{q}, \check{p}_{\cdot}) \tag{9}$$

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

$$E(OTS) = \tilde{q}\,\mathrm{Cov}(\frac{q_{\cdot}}{q} - \bar{D}_{\cdot}, \check{p}_{\cdot}) \tag{10}$$

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

$$E(VS) = \tilde{q}\,\mathrm{Cov}(\bar{D}_{\cdot} - D_{\cdot}, \check{p}_{\cdot}) \tag{11}$$

- Can also look across instruments to study each bin.

UIC Liautaud

Recovering Impact Model Parameters

- Can use clean forms for DI, PI, and TS for inference.
  - *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 ($O(\frac{1}{n^2})$, etc.).

$$PI = \beta_{0,PI} + \pi\tilde{q} + \epsilon_{PI} \tag{12}$$

$$DI = \beta_{0,DI} + \frac{\delta}{1-\delta}\tilde{q} + \epsilon_{DI} \tag{13}$$

$$TS_j = \beta_{0,TS} + \tau\tilde{q}_j\frac{|\tilde{q}_j|}{t_j}\left(1 - \frac{|\tilde{q}_j|}{V_j}\right) + \phi\tilde{q}_j\left(1 - \frac{|\tilde{q}_j|}{V_j}\right) + \epsilon_{TS} \tag{14}$$

UIC Liautaud

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.

Interpretation of *IL*, *IS*

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

$$t = \frac{[\![q]\!](p_0 - p_-)}{\sigma_p \sqrt{t_0 - t_-}} \tag{15}$$

  where σ_p is *price* volatility $(= p\sigma_r)$.
- Implementation Shortfall *IS*: unclean for performance tuning.
  - However, *IS* is applicable to all orders.
  - Pricing of unfilled quantity may be slightly gamed.

UIC Liautaud

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.

UIC Liautaud

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 correct AS with PI:

$$E(AS + PI\frac{n+1}{2n}) = \frac{\tilde{q}\delta}{n(1-\delta)} + \tau\frac{\tilde{q}}{n^2}\sum_{j=1}^{n}\frac{1}{t_j} + \phi[\![q]\!] + o(\tfrac{1}{n}) \tag{16}$$

- Intuition: correct AS for including some permanent impact.

UIC Liautaud

Interpreting Trading Shortfall

- Trading Shortfall TS: clean measure of execution 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:

$$P(b \text{ of } n/2 \text{ worst } TS_j\text{'s in first half}) = \binom{n/2}{b} \frac{1}{2^{n/2}}. \qquad (17)$$

- 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 Liautaud

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.)

UIC Liautaud

Conclusion

- Proposed more informative metrics for algorithmic traders.
- Metrics relate to counterfactuals, trading decisions/software
- May express metrics in terms of realistic price impact model.
  - Can even use a few metrics to recover model parameters.
- Helps managers assess where traders/software excels (or not).
  - Execution, timing, or scheduling skill. . . or luck/noise.
  - Useful for evaluating people or automated trading processes.
  - Is person/process overly timid, volatile, or consistent?
  - Helps reward superior service and punish subpar service.
- Found tests for possible front-running, information dissipation.
- Extension: relate performance variation to other schedules.
  - *e.g.* surprises in volatility, spread, depth, volume.

Merci beaucoup de votre attention!

UIC Liautaud