Execution strategies
Why is it so hard to optimize?

Tianwu (Michael) Cai & George Sofianos
Goldman Sachs Equity Execution Strategies

Q Group Seminar
October 2007

Overview

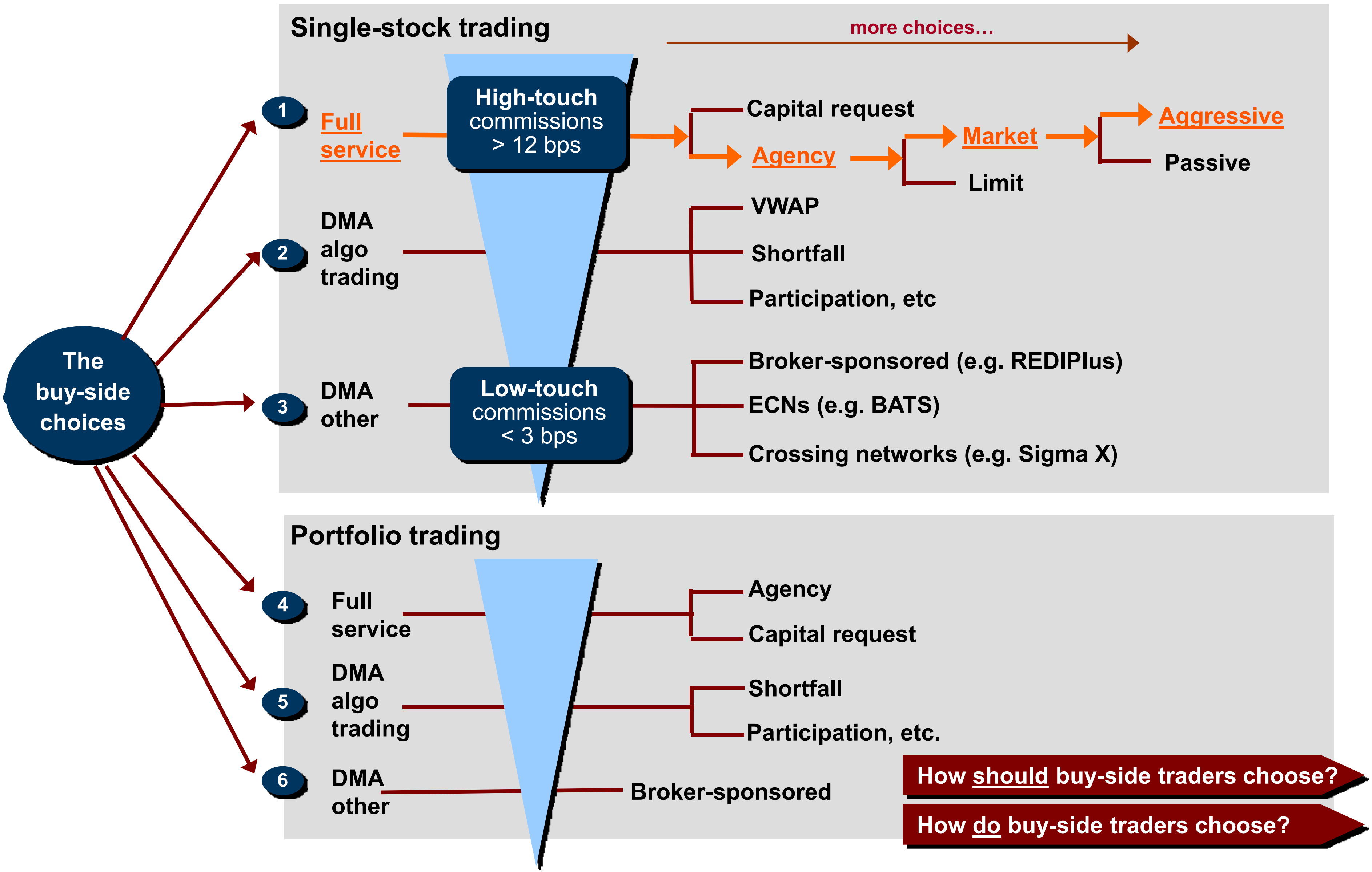
- Big change over past decade: buy-side traders now choose from a wide spectrum of execution choices
  - This was not always the case
  - Ten years ago few execution choices and buy-side traders would delegate most of these choices to high-touch executing brokers
- Expected short-term alpha is perhaps the most important determinant of the appropriate execution strategy[1]
- We have a large data set of buy-side trader executions across the whole spectrum of execution choices
- What is the empirical evidence: do traders optimize across execution choices?
- We find no evidence traders optimize executions based on expected ST-alpha
- Why?
  - ST-alpha is difficult to forecast?
  - Sub-optimal flow of information between PMs and traders
  - Transition period: traders are still learning
  - The offering is too complicated, traders cannot easily choose
- What are the implications?
  - Cannot properly evaluate execution strategies & algorithms if traders do not optimize their usage of strategies & algorithms
  - Should not design execution strategies & algorithms that require unrealistic information, e.g. dynamic algorithms
  - Traders should be modest in their execution choices; should not fine-tune too much

The choice of execution strategy is shifting away from center of order flow

1. See Abrokwah & Sofianos, "Shortfall Surprises," Journal of Trading, Summer 2007

Wide spectrum of execution choices

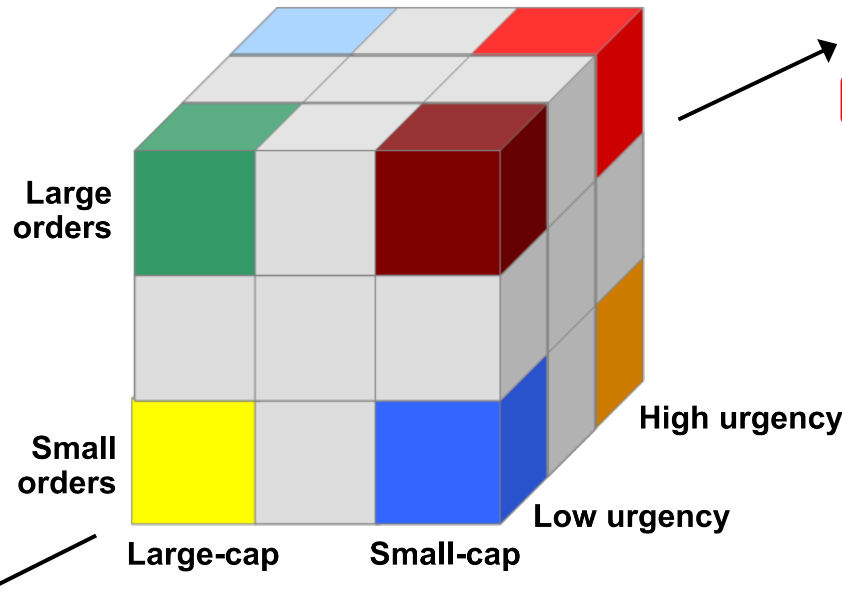
Increasingly, buy-side traders make most of these execution choices

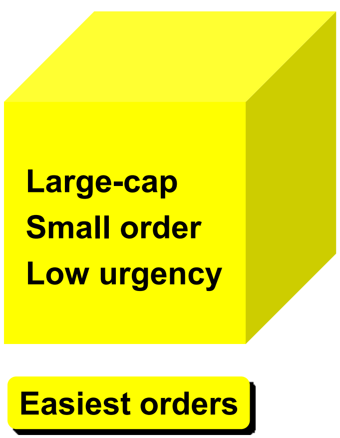


How **should** buy-side traders choose: the Cube framework[1]

- Choice of execution strategy depends on order difficulty
- Order difficulty depends on
  - Order size
  - Stock liquidity
  - Trade urgency

Hypothesis: for difficult orders buy-side traders should choose high-touch; for easy orders should choose low-touch



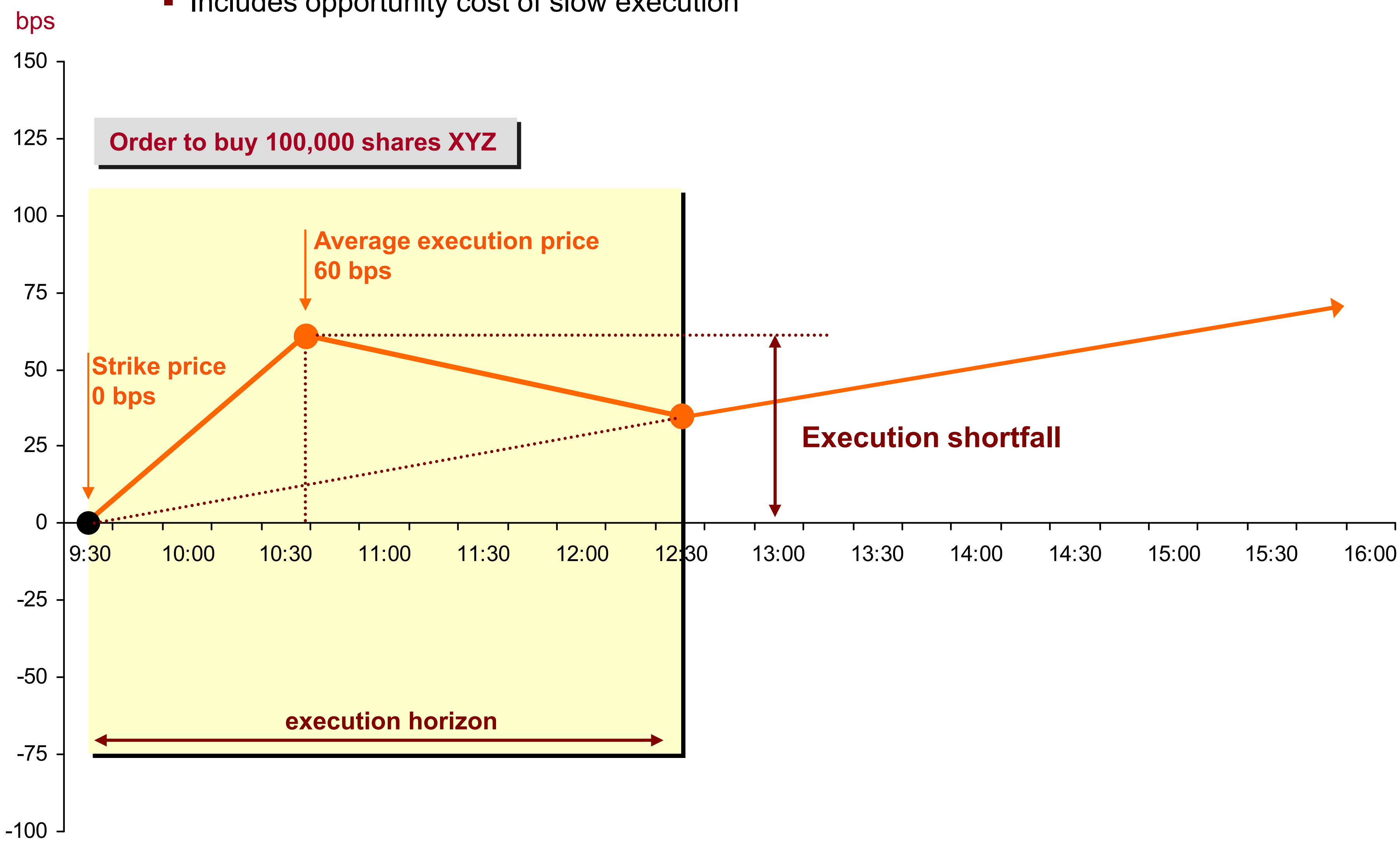


- Trade urgency has 2 components:
  - Short-term (ST) alpha
  - Execution risk

1. For details see Goldman Sachs, Street Smart report, Issue 22

The trading cost measure: execution shortfall

- Definition of execution shortfall (buy orders)
  - Execution price minus price when trader received the order (strike price)
  - Includes opportunity cost of slow execution



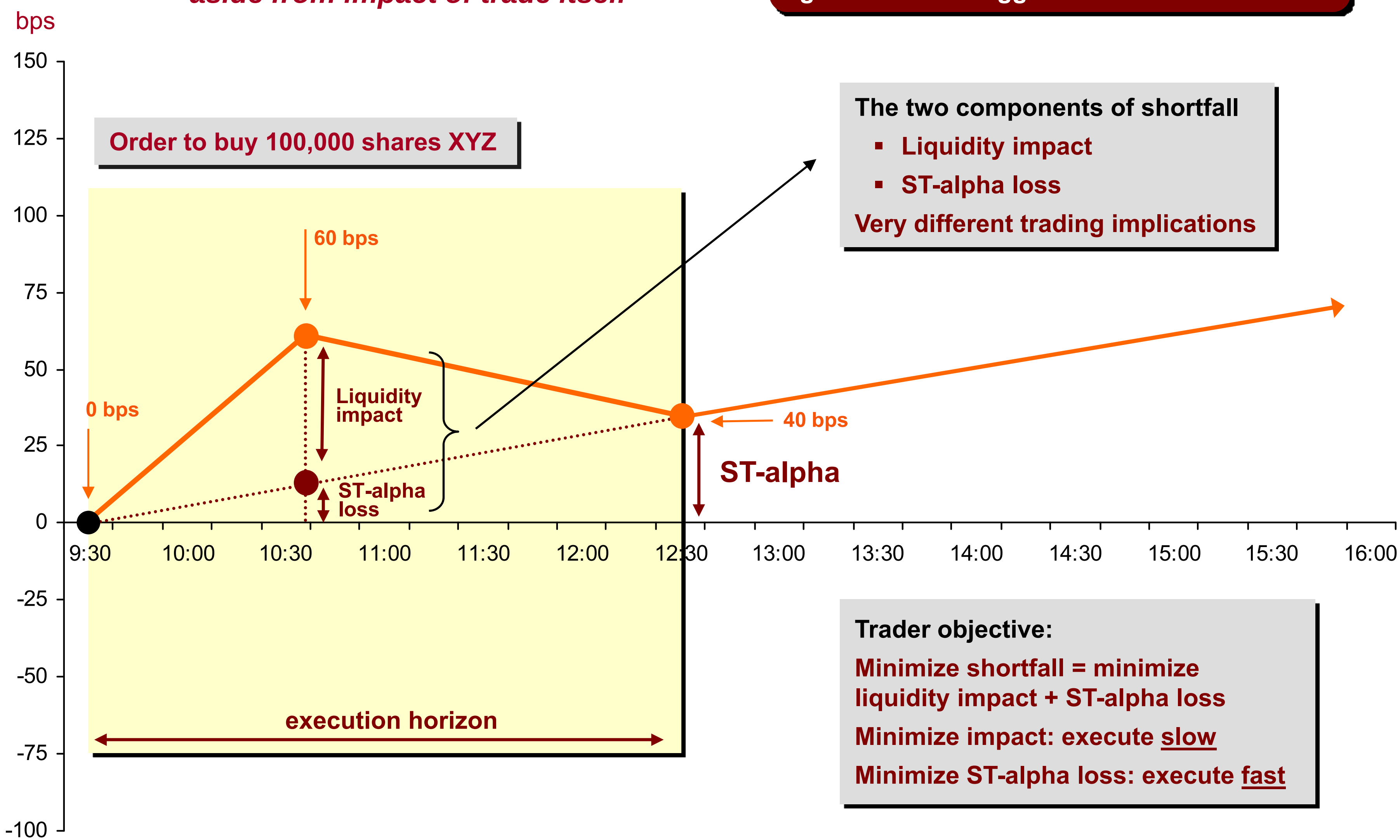
Hypothetical example

Trade urgency: ST-alpha

1

- Definition of ST-alpha (buy orders)
  - Price increase over execution horizon, ***aside from impact of trade itself***

Hypothesis: higher expected ST-alpha means higher trading urgency; traders should choose higher execution aggressiveness



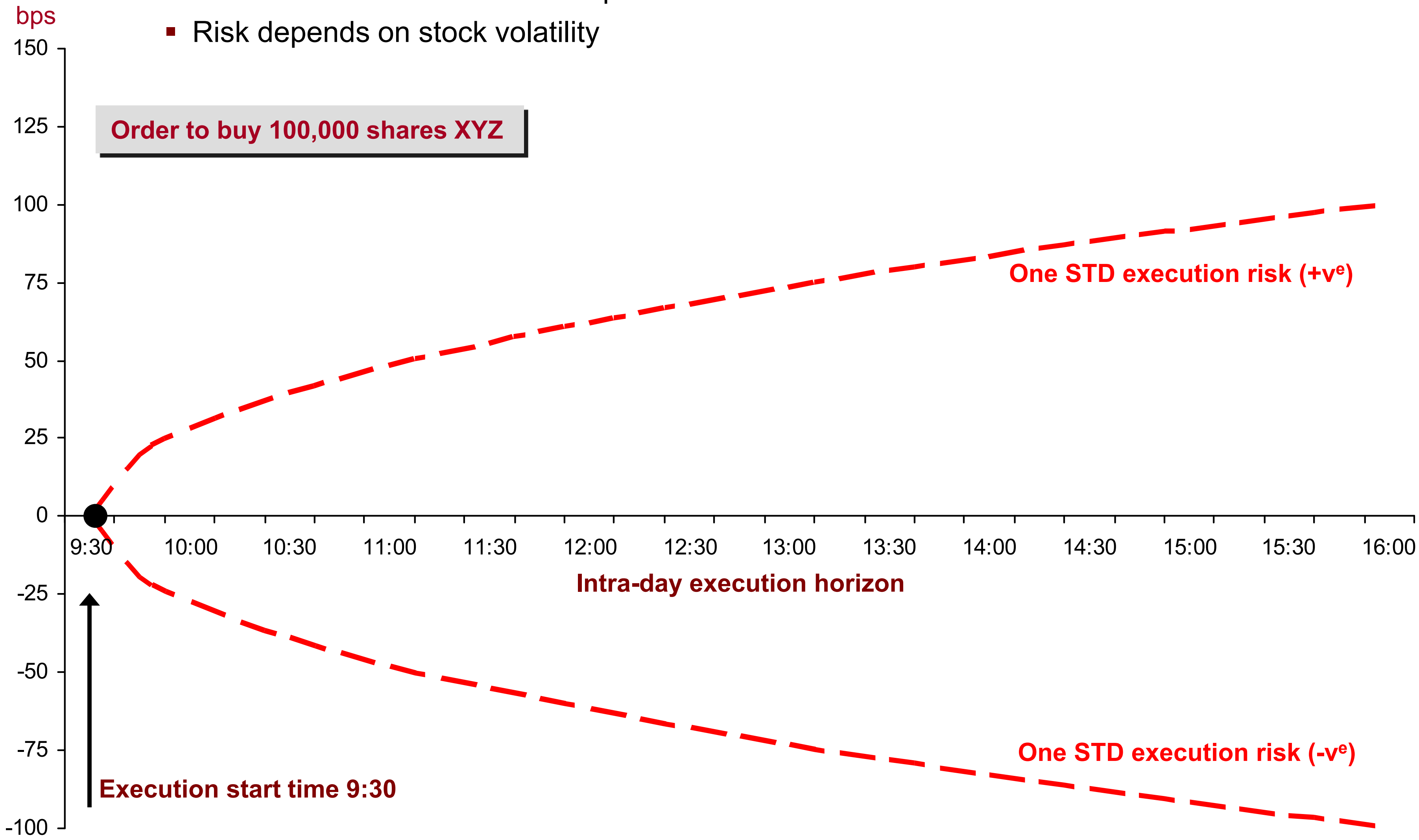
Hypothetical example

Trade urgency: volatility and execution risk

2

- Pure execution risk
  - Price equally likely to go up or down
  - Risk increases with time to completion
  - Risk depends on stock volatility

Hypothesis: assuming risk aversion, higher execution risk means higher trading urgency; traders should choose higher execution aggressiveness



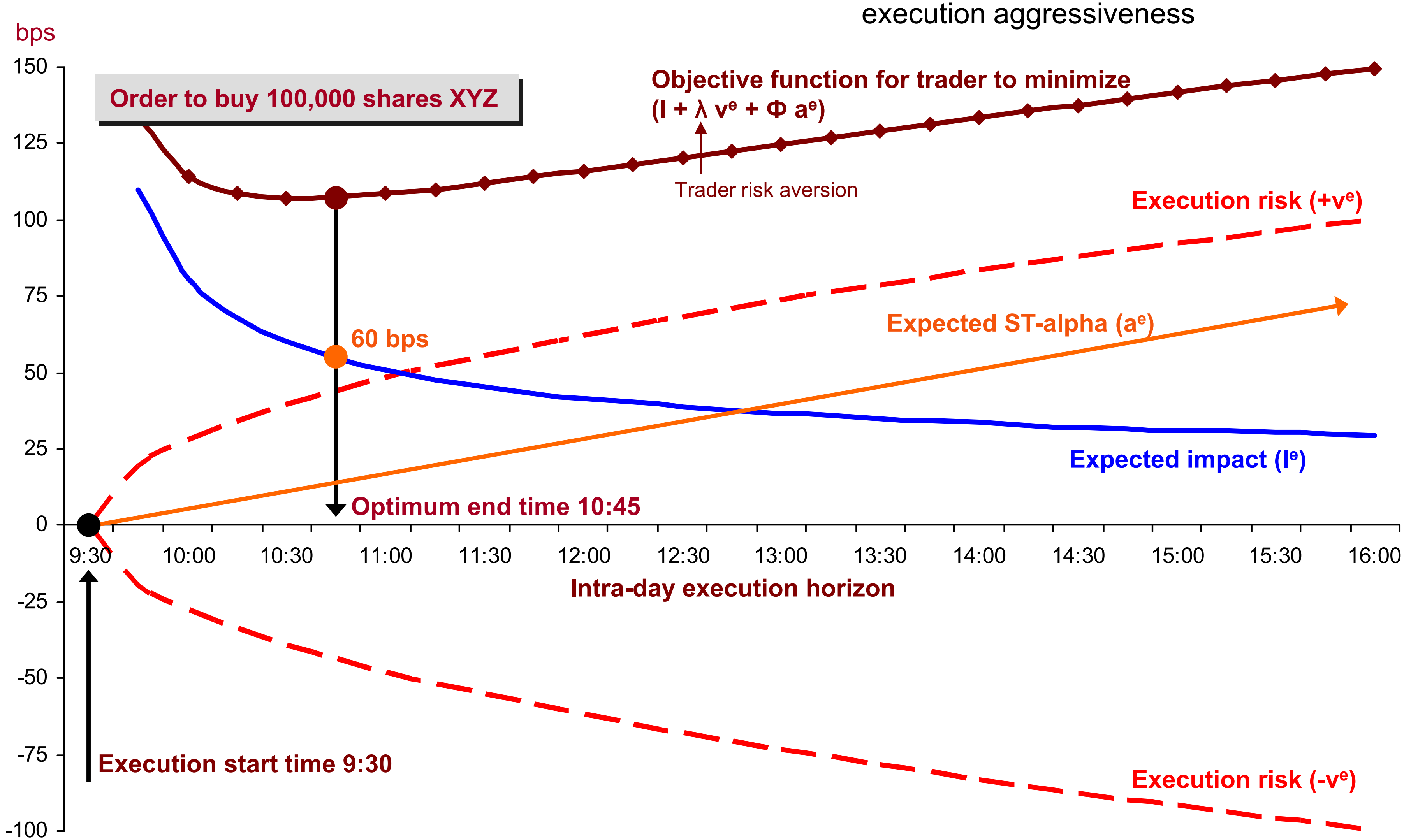
Hypothetical example

The fundamental execution optimization

- Choose execution aggressiveness to minimize impact plus execution risk plus expected ST-alpha

Implications

- Higher volatility, higher execution risk, higher execution aggressiveness
- Higher expected ST-alpha, higher execution aggressiveness



Hypothetical example

How do buy-side traders choose? Overview

Some intriguing evidence on suboptimal execution optimization

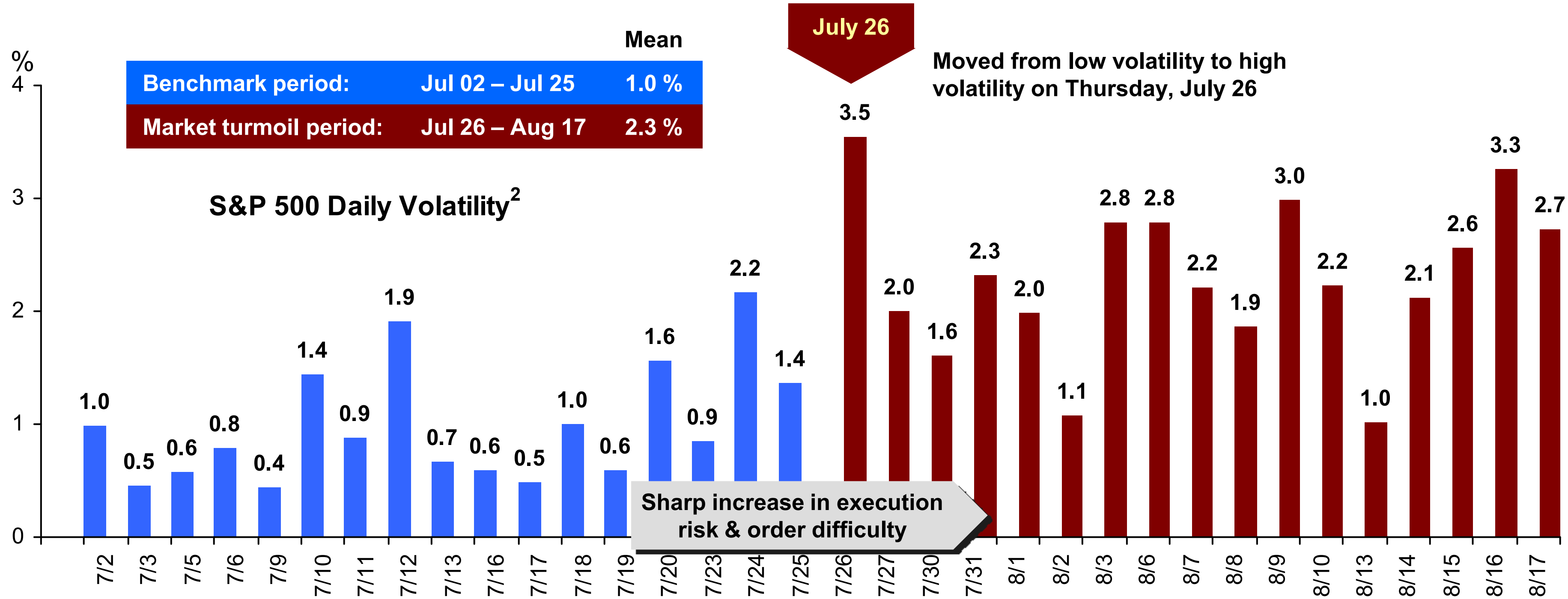
- Effect of recent market turmoil on execution choices
  - No shift from low-touch to high touch
  - No attempt to reduce execution risk: are traders risk neutral?
- Choice between high-touch and low-touch
  - Some evidence of differentiation
  - Comparisons of short-term price dynamics
- Choice between passive and aggressive algorithms
  - “VWAP or Shortfall Algorithms” paper[1]
  - Just as likely to get it wrong as right
- Within an algorithm choice of aggressiveness level
  - Shortfall (4Cast) algorithm risk preference settings
  - Just as likely to get it wrong as right
- Choice of aggressiveness level across all execution strategies
  - Regressions of participation rate on possible determinants
  - Again, just as likely to get it wrong as right

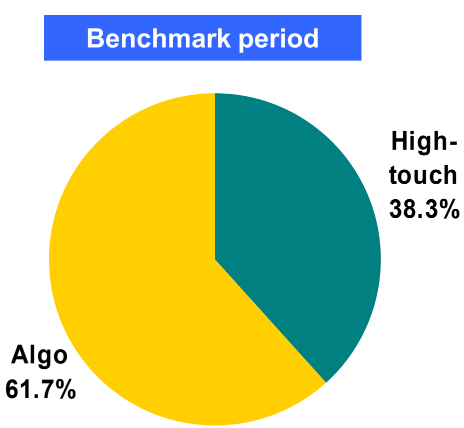
Our main focus

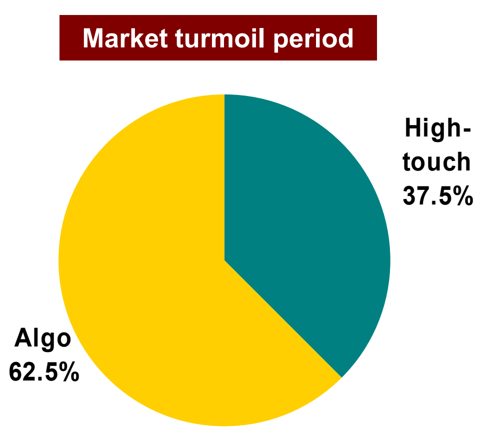
1. Rakhlin & Sofianos, “The Choice of Execution Algorithm: VWAP or Shortfall,” Journal of Trading, Winter 2006

Effect of recent market turmoil on execution choices[1]

		Mean
Benchmark period:	Jul 02 – Jul 25	1.0 %
Market turmoil period:	Jul 26 – Aug 17	2.3 %







Hypothesis: market turmoil increased order difficulty so we should see a shift to high-touch

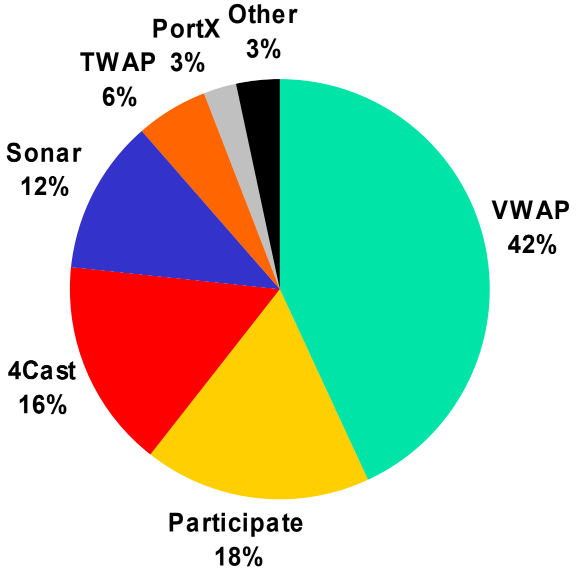
We see no such shift in our data!

Also, within high-touch we see little shift towards higher aggressiveness to offset the increased execution risk

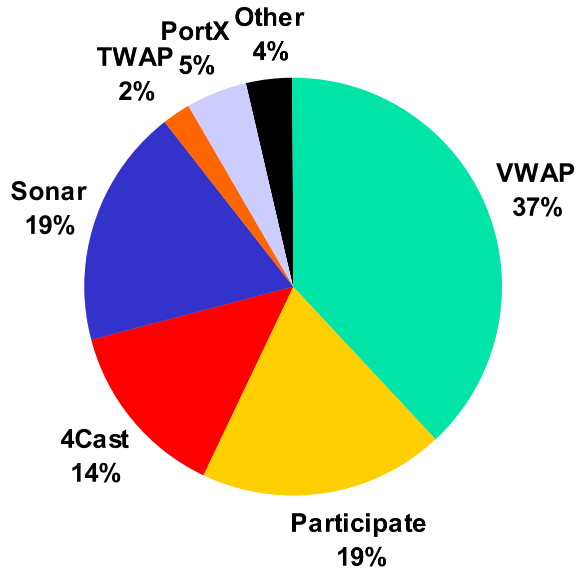
1. Abrokwah & Sofianos, Street Smart, Issue 27, Aug 27, 2007.
2. Daily high minus daily low as % of average high plus low.

More on the recent market turmoil[1]: choice of algorithms

Benchmark period



Market turmoil period



$ value executed

Hypothesis: market turmoil increased execution risk so we should see a shift to more aggressive algorithms

We see no such shift in our data!

Little change in average order characteristics across the various GSAT algorithms

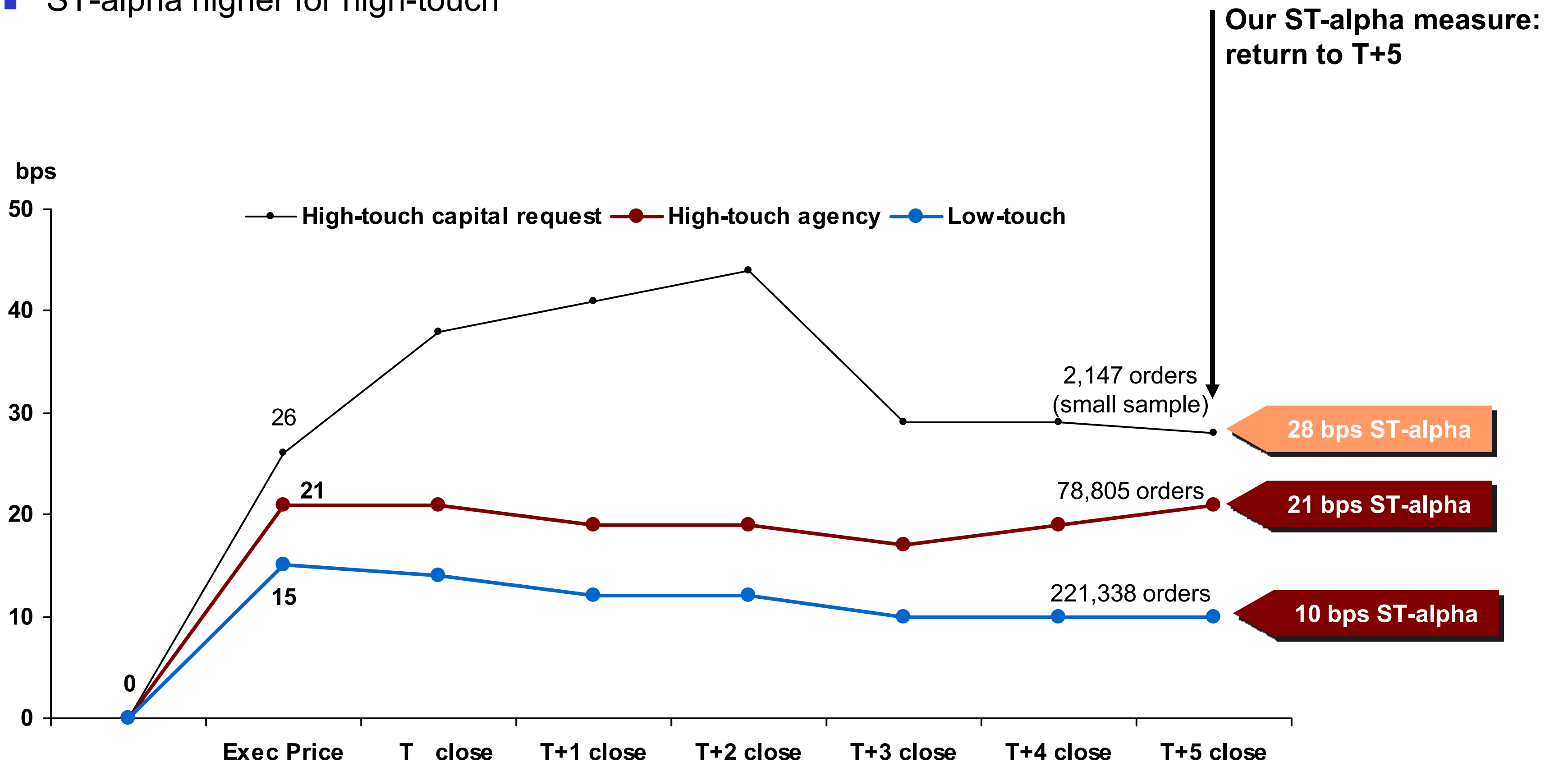
The participation rate actually decreased slightly!

	Order size # Orders	% ADV	# Shares	Arrival strike time	Part. rate	**Expected shortfall (bps)**	**Actual shortfall (bps)**	**Alpha-to-close (bps)**	Exec. half life
Algorithmic trading (US GSAT)									
Benchmark	119,105	4.8%	9,913	11:13	10%	**12**	**14**	**15**	105 min
Market turmoil	171,025	3.6%	10,875	11:30	9%	**14**	**25**	**23**	90 min

1. Cai & Sofianos, Street Smart, Issue 28, Sep 20, 2007.

Choice between high-touch and low-touch

- Evidence of some order flow segmentation between high-touch and low-touch
- ST-alpha higher for high-touch



In constructing the short-term price dynamic we exclude orders with clustering (same client, symbol, side) T to T+5

Choice between passive and aggressive algorithms[1]

- VWAP is a passive algorithm, Shortfall (4Cast) is a more aggressive algorithm
- What is the evidence on how buy-side traders choose?

Sub-optimal choice between VWAP and 4Cast[a]

	# Orders	Expected Shortfall[b] Two Hours	Expected Shortfall[b] Actual Horizon	Part. Rate[c]	Actual[d] Shortfall	ST-alpha[e]	Execution Half-life[f]
VWAP algorithm	28,022	12 bps	12 bps	10%	**12 bps** *(1.0)*	**14 bps** *(1.7)*	201 min
4Cast algorithm	5,033	11 bps	14 bps	22%	**15 bps** *(1.2)*	**12 bps** *(3.6)*	59 min

a. Standard errors in brackets below the estimates.
b. From Goldman Sachs expected shortfall model assuming the execution starts at order arrival and ends two hours later or to the close, whichever comes first. $ value-weighted.
c. $ value filled as % of total $ value executed in the market over the execution horizon. If numerator exceeds denominator we add numerator to denominator.
d. For buys: volume-weighted execution price minus strike price as percent of strike price. For sells: Strike price minus volume-weighted execution price as percent of strike price.
e. Alpha-to-same-day-close. For buys: same-day official close minus strike price as % of strike price. For sells: strike price minus same-day official close as % of strike price.
f. $ value-weighted (in case of multiple executions) time to execution in minutes.

ST-alpha actually lower for 4Cast!

Similar pre-trade order characteristics

Higher impact of more aggressive executions: but why incur the extra cost?

Hypothesis: Orders with higher ST-alpha should go to the more aggressive 4Cast algorithm

Our data shows this is not the case!

1. Rakhlin & Sofianos, "The Choice of Execution Algorithm: VWAP or Shortfall," Journal of Trading, Winter 2006

A simple empirical model of optimal execution aggressiveness

- The model
  - $\pi^* = F(\underline{X}^e, v^e, a^e)$
- Where
  - π^* = optimally chosen execution aggressiveness
  - v^e = expected volatility
  - a^e = expected ST-alpha
  - $\underline{X}^e$ = other factors determining execution aggressiveness
- Using a linear approximation for illustration purposes:
  - $\pi^* = \underline{b}_1 \underline{X}^e + b_2 v^e + b_3 a^e + u$
- Our main hypothesis:
  - $b_2 > 0, b_3 > 0$

Examples of execution aggressiveness

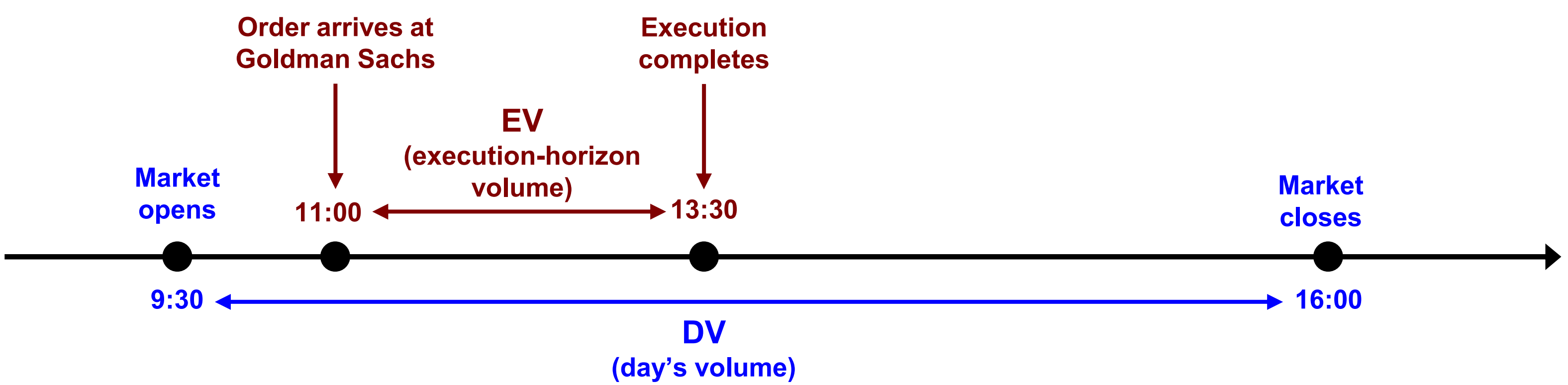
- Choose short execution horizon
- Choose high participation rate
- Request capital
- Choose must-fill market orders

All else equal, optimizing traders should execute more aggressively for higher expected volatility and ST-alpha

- To empirically test the model, we measure:
  - π^* by the <u>actual</u> intra-day participation rate P^a
  - v^e by the previous 21 days median volatility
  - a^e by the <u>actual</u> T+5 price return (perfect foresight)
  - $\underline{X}^e$ other factors include: stock capitalization, quoted spreads, listing venue, limit order
- Unobservable factors that could explain chosen aggressiveness
  - Forced liquidations
  - Holding constraints, etc.
  - By definition, we cannot control for these unobservable factors
- **In our analysis we focus on intra-day executions**

The actual intra-day participation rate

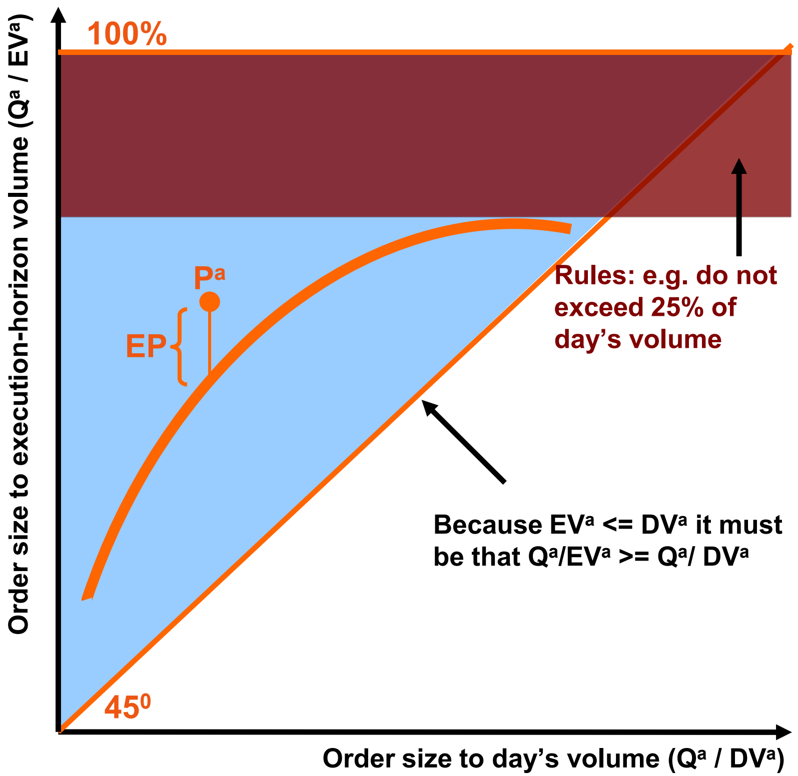
- Definition of actual intra-day participation (part) rate
  - $P^a = Q^a / EV^a$
- Where
  - Q^a = actual quantity executed
  - EV^a = actual consolidated market volume in stock over execution horizon
- The part rate is a natural measure of order aggressiveness
- But: must deal with several measurement issues
  - Correlation between part rate and order size as % of day's volume (DV)
  - Ceilings on part rate
  - Ideally we need the <u>intended</u> part rate but we only have the <u>actual</u> part rate
  - Bi-modal distribution of part rate: small-order fast executions
  - In certain pathological cases, part rate can exceed 100% (we exclude them)
- We examine each of these issues in turn



By definition: EV <= DV

Correlation between part rate & order size as % of day's volume

- The intra-day part rate can never be less than order size as % of day's volume
  - $P^a = Q^a / EV^a >= Q^a / DV^a$
- Because
  - $EV^a <= DV^a$
- Also ceilings on part rate
  - Part rate cannot exceed 100%[1]
  - Trading rules: e.g. do not exceed 25% of day's volume



Solution 1: A two-step approach

- **Step one**
  - **Regress part rate on order size as % of day's volume: Q^a / EV^a on Q^a / DV^a**
  - **Non-linear to take account of ceilings**
  - **Call the residual <u>the excess part rate</u> (EP)**
- **Step two**
  - **Regress EP on the factors that may explain the choice of high aggressiveness**

Solution 2

- **One-step, but add order size as % of day's volume as an additional explanatory variable**

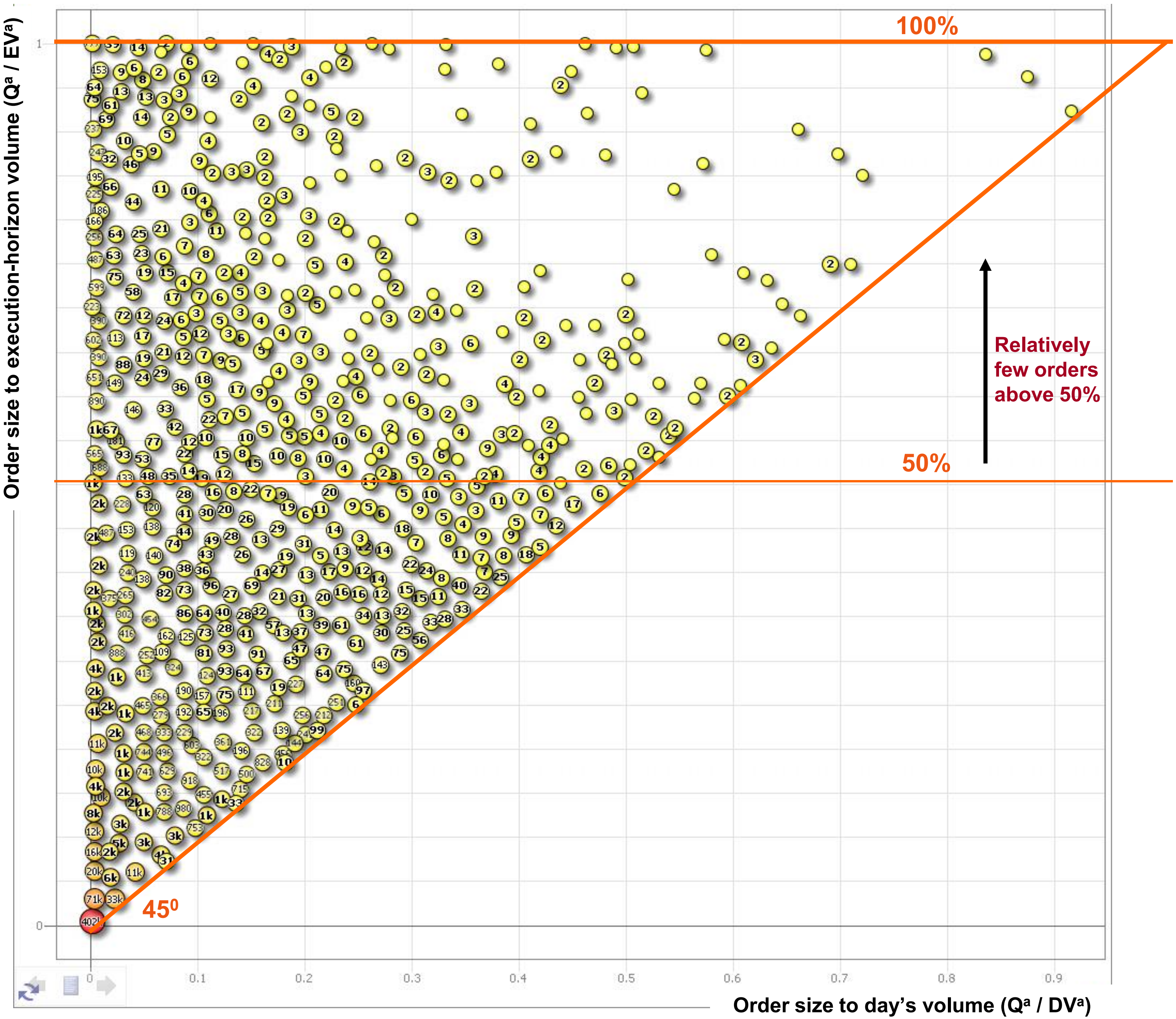
Solution 3

- **One-step, but use the difference between part rate and order size as % of day's volume as the dependent variable**

It does not make much difference which solution we use!

We report results for Solution 2.

The evidence in our data on the part rate



Intended and actual part rates

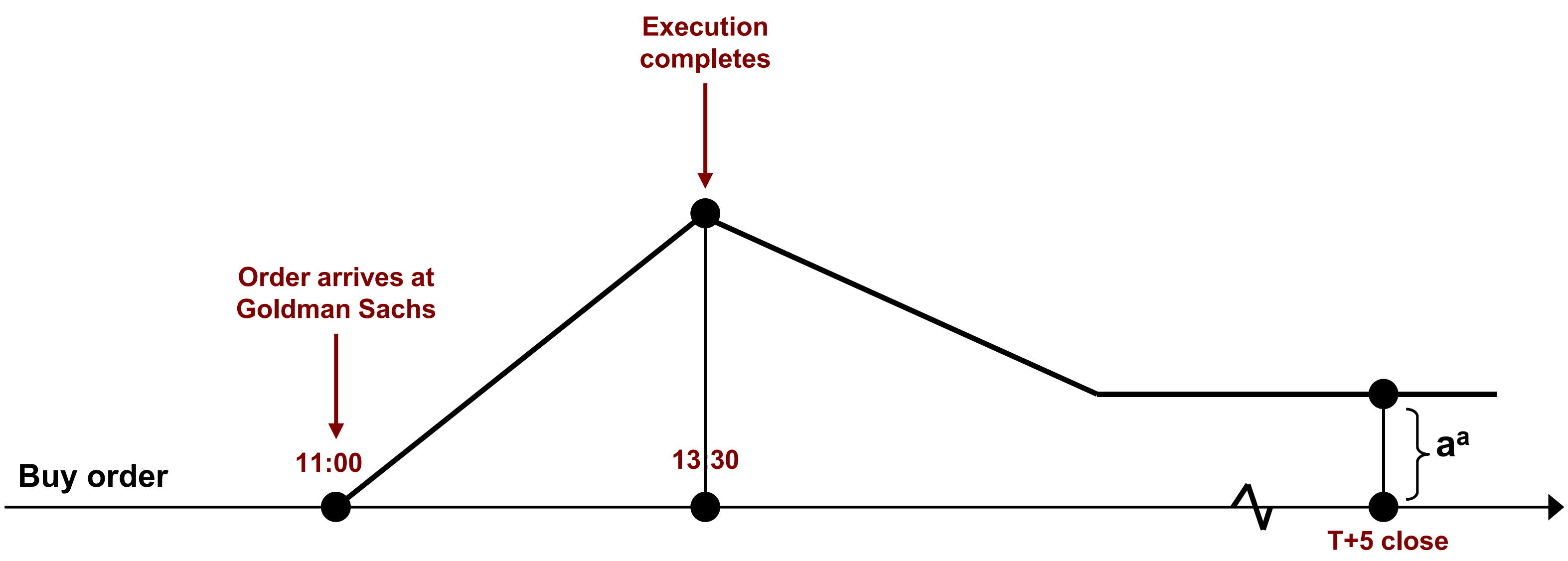
- Traders optimize their <u>intended</u> part rate (P^e), based on information available pre-trade
  - $P^e = G(\underline{X}^e, v^e, a^e)$
- But what we observe is the actual part rate P^a
- Reasons why actual part rate may differ from intended part rate:
  - Actual execution-horizon volume (EV^a) different than expected (EV^e)
  - Actual executed order size (Q^a) different than intended (Q^e)
    - Cancelled market orders
    - Limit orders that do not fully execute
- We control for these factors by including in our regressions:
  - Volume surprises (EV^a/EV^e)
    - We measure expected volume as the prior 21 trading days median
  - Fill rates (Q^a/Q^e)
    - Where we measure intended order size by the order size submitted
  - Limit order dummy

Hypothesis: positive volume surprise (EV^a > EV^e) lowers actual part rate

Hypothesis: low fill rate lowers actual part rate

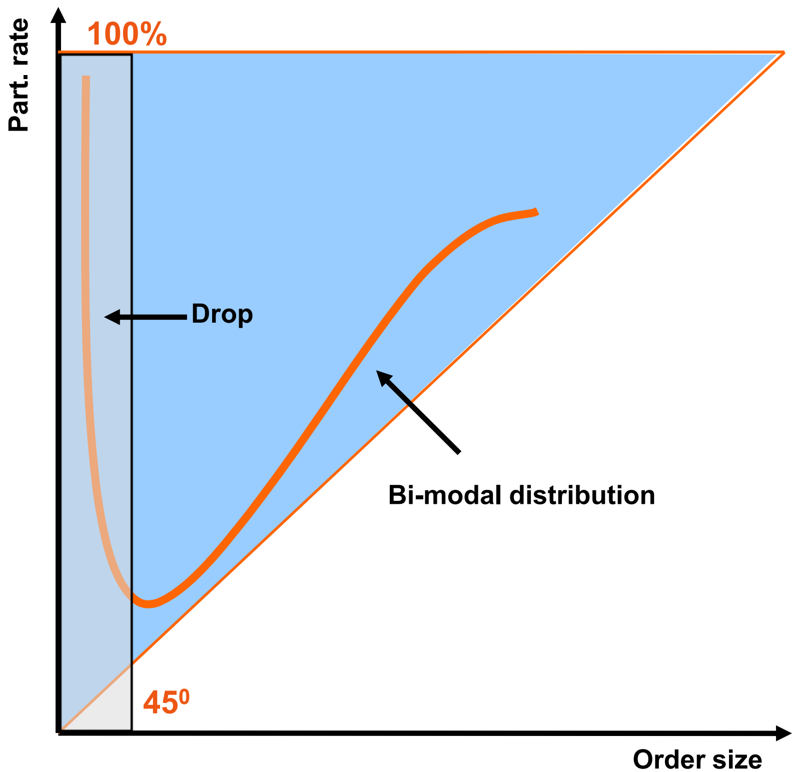
ST-alpha again

- Back to our model….
  - $P^e = G(\underline{X}^e, v^e, a^e)$
- We cannot observe a^e
- We use <u>actual</u> ST-alpha (a^a) through T+5 close (see diagram below)
  - T+5 to allow for reversal to play through
- Implications of using actual (perfect foresight) ST-alpha
  - Cannot distinguish between inability to optimize or inability to forecast ST-alpha
- Complication: multi-day executions - impact may stretch over several days
  - We drop orders that "cluster" T through T+5
    - Same-trader, same-symbol, same-side within the 5-day window



Small orders with fast executions

- Another issue with the part rate:
  - It has a bi-modal distribution as a function of order size
- Small orders that execute fast may have a very high part rate but it is not aggressiveness
  - The liquidity is there just take it
  - 100 shares of MSFT executing instantaneously may have a 100% part rate
- Our model applies to relatively large orders where there is a credible time-impact trade-off
  - What length execution time provides a credible time-impact trade-off?



Solution: Drop orders with fast executions

- **Sensitivity analysis using different cut-offs**

Summary of main estimated aggressiveness model

- Dependent variable
  - $Log(P^a / (1-P^a))$
  - Where P^a is the actual part. rate (Q^a / EV^a)

The 9 explanatory variables

		Relation to part. rate	Detail	Notation	Functional form
1	Size as % of DV	**POSITIVE (+)**	Executed order size as % of same day's volume	Q^a / DV^a	Log
2	Volume surprise	**NEGATIVE (-)**	Expected day's volume divided by actual day's volume	DV^e / DV^a	Log
3	Fill rate	**NEGATIVE (-)**	Actual quantity executed divided by intended quantity	Q^a / Q^e	Log
4	Market cap		Stock market capitalization		Log
5	Spread		21-day median daily average bid-ask spread (in bps)		
6	Listing venue		NYSE=1, NASDAQ=0		
7	Limit order	**NEGATIVE (-)**	Limit=1, Market=0		
8	**Volatility**	**POSITIVE (+)**	**21-day median % day's high price minus low price**	v^e	
9	**ST-alpha**	**POSITIVE (+)**	**Percent price return from order arrival to T+5 close (in bps)**	a^a	

- Main results are not sensitive to the regression functional form

Sample description

- US equity trading, Goldman Sachs client orders
- June 06 to June 07 (13 months)
  - High-touch single-stock (US Shares) orders
  - Low-touch GSAT Direct orders (GSAT=Goldman Sachs Algorithmic Trading)
  - Low-touch GSAT REDIPlus orders (Feb to Jun 07)
- We aggregate intra-day same orders
  - Same-trader, same-symbol, same-side[1]
- We do extensive filtering, e.g.
  - Only stocks with US country of origin
  - Only common stocks
  - Regular settlement
  - Intra-day executions (09:30 – 15:55)
  - Execution half-life > 1 min
  - Drop odd lots
  - Drop if exec. quantity > exec. horizon volume
  - Drop outliers, e.g. 5-day return > 20%, shortfall > 1000 bps
  - Drop orders with missing information
- We also exclude multi-day T to T+5 orders
  - Same-trader, same-symbol, same-side[1]

- Final sample: ~300,000 orders

~Number of orders in sample			
	High-touch single-stock	Low-touch algo (GSAT)	Total
Initial	1,400,000	1,600,000	3,000,000
After aggregation	300,000	700,000	1,000,000
After filtering	170,000	530,000	700,000
Excl. multi-day	70,000	230,000	**~300,000**

1. The order after aggregation assumes strike time at first order's arrival and end time at last order's last execution

Part rate regressions: results and sensitivity analysis

Summary statistics: the t-values

	Hypo-thesis	Execution half-life:				Half-life >15 & order size (%DV):			
		>1 min	>5 min	**>15 min**	>30 min	>0.1%	>0.5%	>1.0%	>2.0%
		1	2	3	4	5	6	7	8
Intercept		-21.5	18.3	**36.6**	9.9	22.9	4.8	2.0	4.8
Size as % of DV	(+)	**426.1**	**649.8**	**817.9**	**996.6**	**547.4**	**327.8**	**241.6**	**170.3**
Volume surprise	(-)	1.9	-7.6	**-10.0**	-14.2	-12.6	-6.1	-2.1	-0.4
Fill rate	(-)	-35.3	-43.2	**-40.9**	-36.7	-34.8	-23.4	-15.9	-9.3
Market cap		39.8	23.3	**11.3**	8.9	21.8	28.9	23.5	12.8
Spread		23.8	9.1	**6.5**	5.2	8.7	5.6	2.5	0.0
Listing venue		16.5	9.1	**0.3**	-6.4	-4.9	-4.4	-2.6	-1.7
Limit order	(-)	-4.7	-11.2	**-8.1**	-1.1	-7.1	-5.9	-4.6	-0.2
Volatility	(+)	10.5	9.6	**0.0**	-2.0	5.8	12.4	11.2	6.7
ST-alpha	(+)	**-0.8**	**-1.8**	**-0.8**	**0.5**	**-0.4**	**-0.7**	**-1.2**	**-1.5**
Sample size		302,290	245,410	**200,766**	163,972	156,369	107,259	81,942	55,407
R-square		41%	67%	**80%**	88%	68%	54%	48%	44%

Dominates

Never positive & significant!

Preferred

- No evidence of ST-alpha optimization
  - In all specifications!
  - ST-alpha is <u>never</u> positive & significant
  - A robust result
- Volume surprise, fill rates, limit orders: significant and with predicted sign
  - Most of the time!
- Volatility: mixed evidence
- Tried many other specifications (market orders only, algo only, high-touch only, etc.), similar results!

The two-step approach

The t-values

	Hypo-thesis	Half-life >15 min
Step one: part rate on size as % DV		
Intercept		**140.5**
Size as % of DV	**(+)**	**900.5**
R-square		**80%**
Step two: excess part rate		
Intercept		**-10.3**
Volume surprise	**(-)**	**-9.2**
Fill rate	**(-)**	**-40.5**
Market cap		**9.9**
Spread		**7.1**
Listing venue		**0.8**
Limit order	**(-)**	**-6.8**
Volatility	**(+)**	**-0.4**
ST-alpha	**(+)**	**-0.6**
R-square		**1%**

Most explanatory power in first step

Little explanatory power in explaining the excess part rate

- Step one
  - Regress part rate on order size as % of day's volume (Q^a / EV^a on Q^a / DV^a)
  - Call the residual the excess part rate (EP)
- Step two
  - Regress EP on the factors that may explain choice of high aggressiveness

- Consistent with our one-step regression results:
  - Most explanatory power is in the first step regressing the part rate on order size as % of day's volume
  - In the second step, little explanatory power in explaining the excess part rate

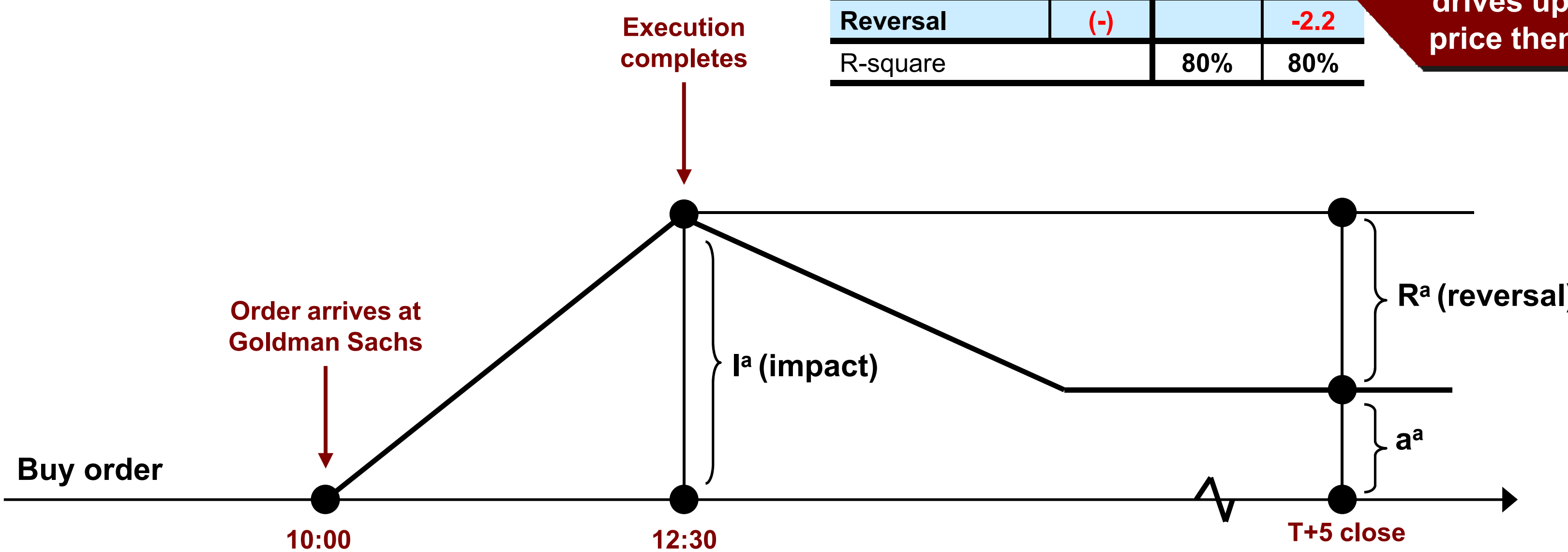
Liquidity impact and reversal

- Am interesting decomposition of ST-alpha
  - Liquidity impact (I^a)
  - Reversal (R^a)
- Where
  - $a^a = I^a + R^a$
- We also ran regressions on the impact and reversal decomposition
  - $P^a = \underline{b}_1 \underline{X}^e + b_2 v^e + b_4 I^a + b_5 R^a$
  - Hypothesis: $b_4 > 0$, $b_5 < 0$

The t-values

	Hypo-thesis	Half-life >15 min	
		A	**B**
Intercept		36.6	36.5
Size as % of DV	(+)	817.9	813.4
Volume surprise	(-)	-10.0	-9.8
Fill rate	(-)	-40.9	-40.9
Market cap		11.3	11.1
Spread		6.5	6.5
Listing venue		0.3	0.3
Limit order	(-)	-8.1	-7.6
Volatility	(+)	0.0	-0.2
ST-alpha	(+)	**-0.8**	
Liquidity impact	(+)		**5.6**
Reversal	(-)		**-2.2**
R-square		**80%**	**80%**

High part rate drives up price but price then reverses



Evidence on algo choice: choosing aggressiveness settings

- A different dependent variable in our regressions
  - No longer part rate
  - Our results are not sensitive to the use of part rate as our aggressiveness indicator

The 4Cast probit z-values

	All 4Cast orders
Intercept	-18.9
Size as % of DV	14.2
Volume surprise	4.8
Fill rate	-3.9
Market cap	25.6
Spread	3.2
Listing venue	11.1
Limit order	59.7
Volatility	**-0.6**
ST-alpha	**-2.5**
Sample size	25,977

Again no evidence traders optimize

- We already discussed the evidence on the choice between passive and aggressive algorithms
  - No evidence traders optimize
- How about within an algorithm?
  - Do traders optimize in the choice of aggressiveness settings?
- Focus on the Goldman Sachs 4Cast algorithm
- Ten aggressiveness settings
  - 1 (least aggressive) to 10 (most aggressive)
  - Divided them into two groups:
  - 1 – 4 Aggressive dummy variable=0
  - 5 – 10 Aggressive dummy variable=1
- We estimated a multifactor probit model to explain the [0,1] aggressiveness choice
- Main findings:
  - Volatility does not explain the choice
  - ST-alpha does not explain the choice
- No evidence traders optimize the 4Cast algorithm's aggressiveness settings
  - In terms of volatility
  - In terms of ST-alpha

Why is it so hard to optimize? Focus on evidence on ST-alpha

- Shortfall Surprises paper[1]: ST-alpha main reason why actual shortfall differs from expected shortfall
  - Expected ST-alpha by far the most important factor in predicting shortfall
  - Expected ST-alpha by far the most important factor in choosing the right execution strategy and managing shortfall
- But our data suggests no correlation between actual ST-alpha and choice of execution strategy
- Why?
- Possibility one: traders do not optimize
  - A transition period, they are still learning
  - The offering is too complicated, traders cannot easily choose
- Possibility two: traders cannot easily predict ST-alpha

The sell-side should simplify its offering, e.g. streamline algorithms ✓

- We prefer the 2nd possibility
- The question then becomes: why is it so hard to predict ST-alpha?
- At least four reasons
  - Organizational failure: suboptimal communication between portfolio managers and traders
  - ST-alpha is intrinsically hard to predict: exactly how alpha signal gets incorporated in price
  - ST-alpha is intrinsically hard to predict: short-term order flow considerations
  - If can predict ST-alpha should be trading on that!

1. See Abrokwah & Sofianos, "Shortfall Surprises," Journal of Trading, Summer 2007

The investment & execution process

Organizational failure

The investment process

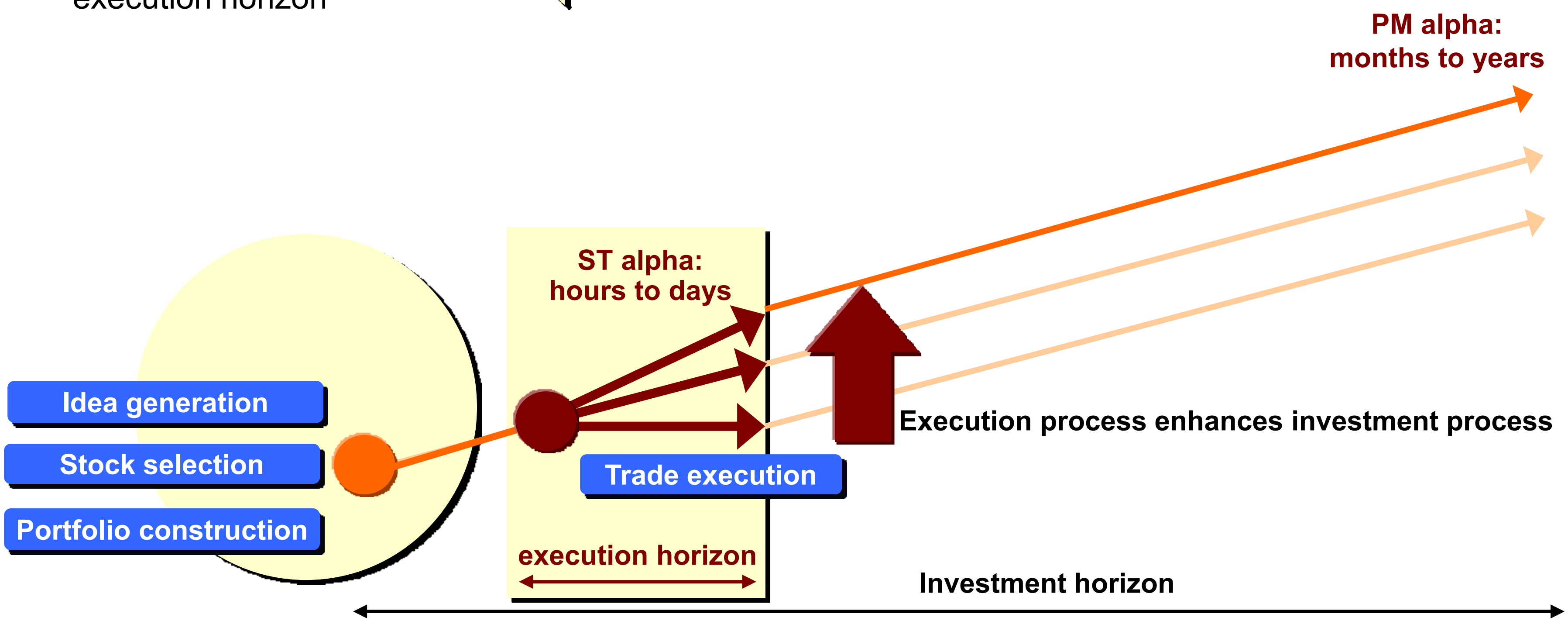
- Idea generation
- Stock selection
- Portfolio construction
- Maximize PM alpha

Strategic

The execution process

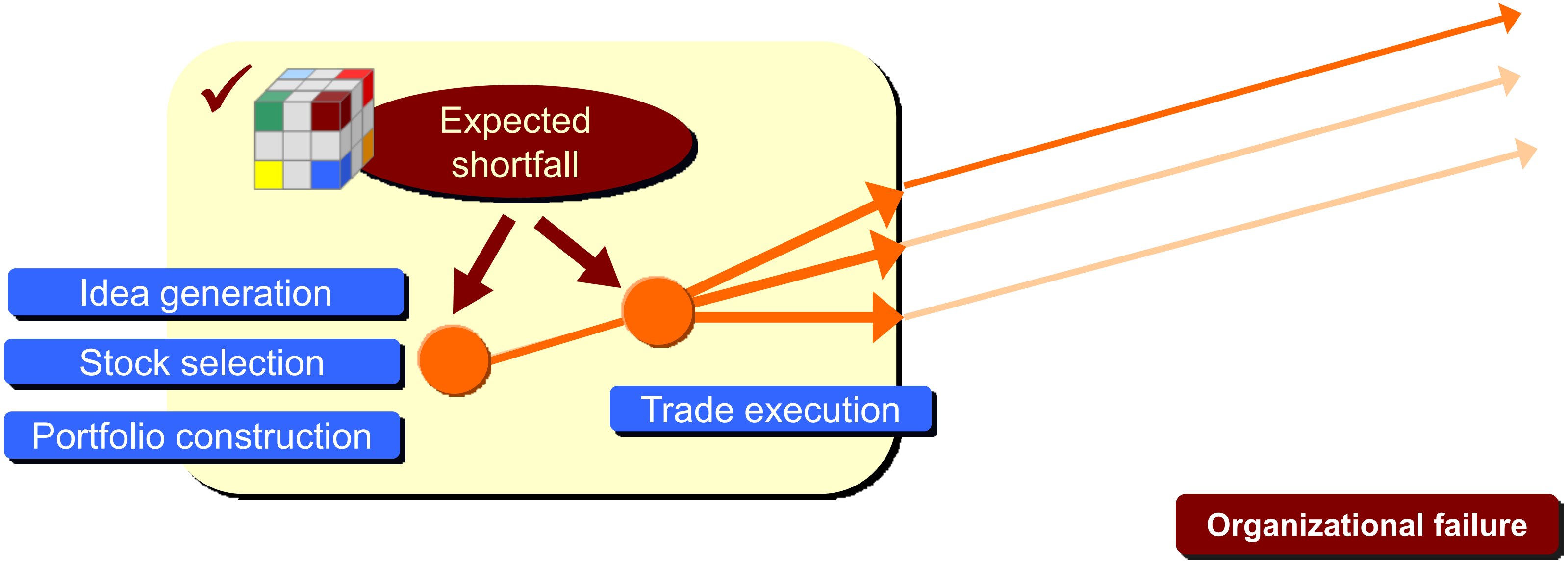
- Minimize alpha loss over the execution horizon

Tactical



The ideal organizational structure

- Complete integration of idea generation, stock selection, portfolio construction & execution
- Ideas are balanced with expected trading costs
- Dialog exists among idea generation, portfolio construction & execution



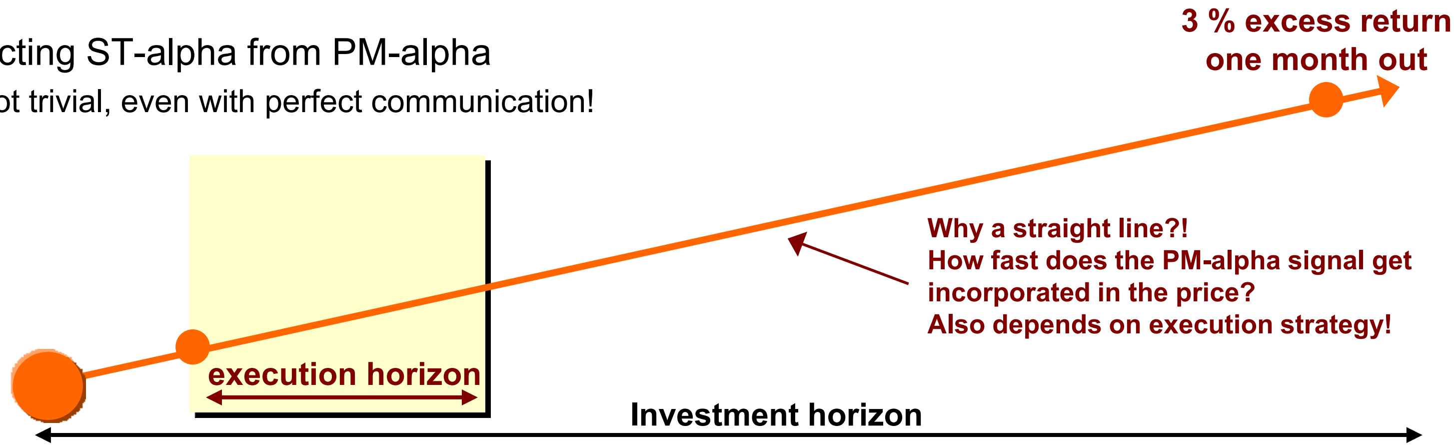
In reality, sub-optimal information flow across three stages in the process

Estimating ST alpha

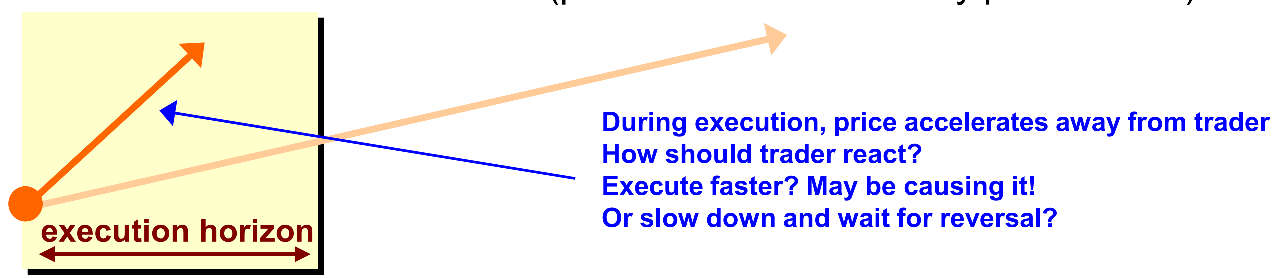
- **PM input: investment strategy**
- **Trader input: "feel for the market"**
- **Data analysis: the "science" part**

Challenges in estimating ST-alpha

- Extracting ST-alpha from PM-alpha
  - Not trivial, even with perfect communication!



- Distinguishing between momentum and reversal (permanent and transitory price moves)



- What determines ST-alpha?
  - Short-term order-flow dynamics more than long-term PM-alpha signals
- Traders "feel for the market" is important
  - But for traders to have a good "feel for the market" must have central view of order flow
  - NYSE specialists & floor brokers used to, but <u>nobody</u> has any more!
- Traders that are good in predicting ST-alpha will quickly move to other opportunities!