PRINCETON UNIVERSITY

ORF 474: High Frequency Trading
Spring 2020
Robert Almgren

Lecture 11b

April 22, 2022

Macro transaction cost model

- You are a broker, or execution desk
- Client, or portfolio manager, wants to execute a trade
- A large trade will push the market against you
  - large buy order pushes price up as you are buying
  - large sell order pushes price down as you are selling
- How much will the market move?
- Answer based on results for similar trades in past
  - do not need theoretical explanation

The broker's problem

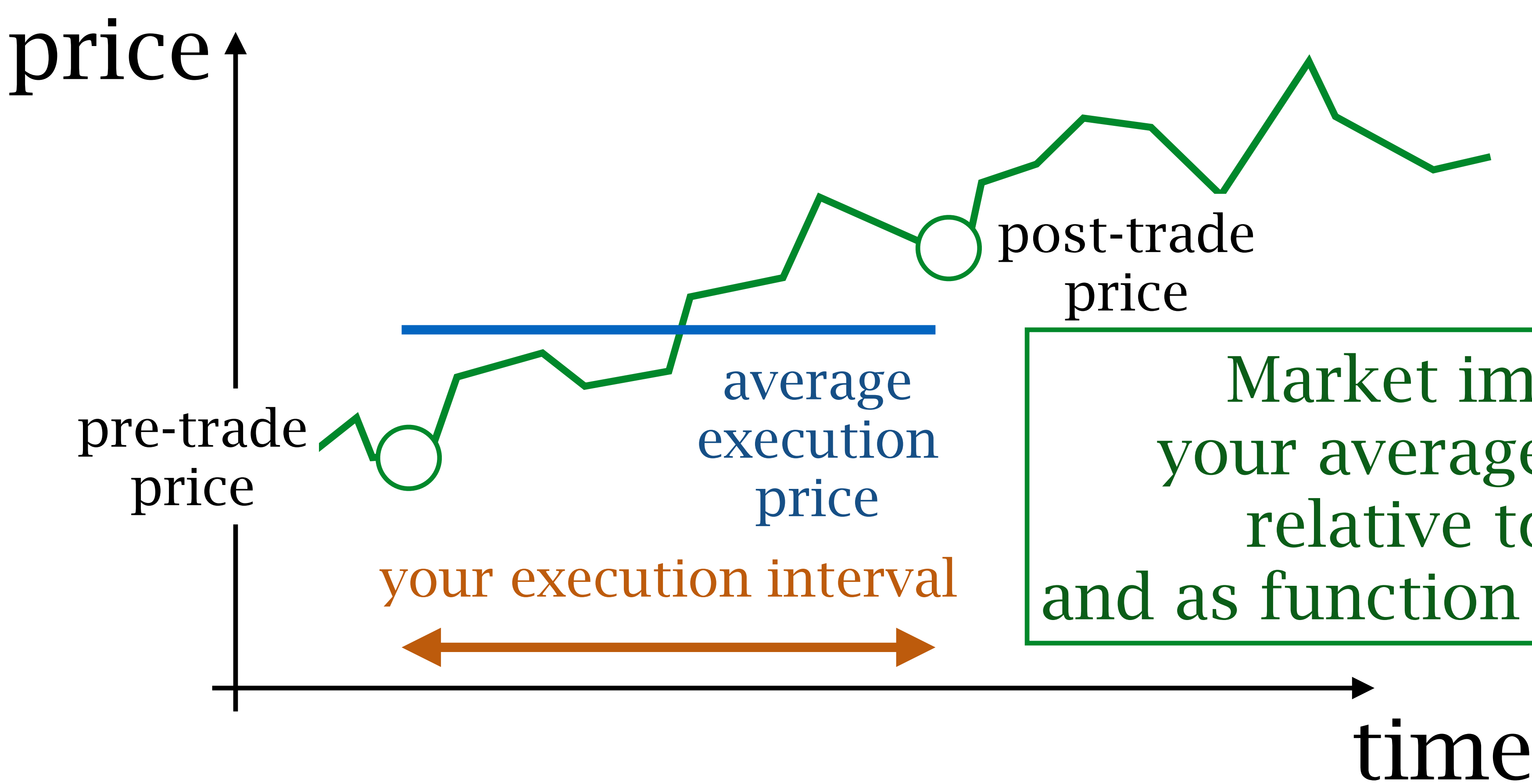
- Execute order as agent for client
  separation of execution from investment decision
- Goal: best final average execution price
- What is a good price?
- Evaluate relative to benchmark
  benchmark defines an "ideal" trade
  different benchmarks give different strategies

Slippage vs impact

- Micro impact: effect on market of individual trade
- Macro impact -- two things to consider:

how does your order change the market?

what is the expected cost of your order?



Market impact: focus on your average execution price relative to benchmarks and as function of order parameters

Slippage is important

quantitativebrokers
BEST EXECUTION FOR FIXED INCOME AND FUTURES

February 12, 2016

Page 1

Algorithm Performance Comparison

Market	Bid/Ask	Bulge Bracket Banks AP	Bulge Bracket Banks I-VWAP	Quantitative Brokers AP	Quantitative Brokers I-VWAP	QB Improvement AP	QB Improvement I-VWAP
US 5-yr Note	7.81	3.14	2.22	1.53	−0.76	−1.61	−2.98
US 10-yr Note	15.63	6.84	4.81	2.44	−0.94	−4.40	−5.75
US 30-yr Bond	31.25	13.6	9.16	4.18	−1.59	−9.42	−10.75
Eurex Bobl	11.39	3.65	1.61	2.05	−0.59	−1.60	−2.20
Eurex Bund	11.39	5.89	1.51	4.2	−1.17	−1.69	−2.68
LIFFE Long Gilt	14.1	6.69	1.46	6.25	−2.14	−0.44	−3.60
E-mini S&P	12.5	5.8	3.54	3.56	−1.49	−2.24	−5.03
E-mini NASDAQ	5	4.2	1.45	2.82	−1.11	−1.38	−2.56
NYMEX Heating Oil	4.2	13.95	4.12	9.66	−2.49	−4.29	−6.61
NYMEX Crude Oil	10	13.7	2.55	9.32	−3.52	−4.38	−6.07
NYMEX Nat. Gas	10	12.76	3.15	10.64	−3.44	−2.12	−6.59
Weighted Average:		6.94	3.17	4.06	−1.44	−2.88	−4.61

Table 2: Relative performance of bulge bracket and QB algorithms. Positive numbers (black) indicate positive costs or underperformance to benchmark, negative numbers (green) indicate negative slippage or outperformance to benchmark. AP denotes Arrival price and I-VWAP denotes interval VWAP. All figures are in US dollars per lot, using recent EUR and GBP exchange rates for Eurex and LIFFE products. The bid/ask spread in dollars is given for comparison. QB consistently achieved lower slippage than the bulge bracket banks.

the Bolt algorithm again outperformed the bank algos on a weighted basis by an average of \$4.61 (\$3.17 + \$1.44) per lot. Assuming the entire set of 3.3MM contracts traded were sent either to one venue or the other, utilization of Quantitative Brokers' Bolt algorithm offered a potential transaction cost savings relative to arrival price of over \$9.5MM in 34 months. On a time-weighted basis, this represents approximately 77 basis points of annual performance across the Mosaic Institutional and Emerald Futures Programs.

Slippage

- Difference of average execution price and benchmark

  execution - benchmark for buys

  benchmark - execution for sells
- Positive slippage is bad, negative is good
- For agency execution, minimize mean slippage

Pedigree of "arrival price" or "implementation shortfall"

The implementation shortfall: Paper versus reality

André F. Perold

Journal of Portfolio Management; Spring 1988; pg. 4

Reality involves the cost of trading and *the cost of not trading.*

After selecting which stocks to buy and which to sell, "all" you have to do is implement your decisions. If you had the luxury of transacting on paper, your job would already be done. On paper, transactions occur by mere stroke of the pen. You can transact at all times in unlimited quantities with no price impact and free of all commissions. There are no doubts as to whether and at what price your order will be filled. If you could transact on paper, you would always be invested in your ideal portfolio.

Execution cost also measures price impact. For the purposes of this discussion, let us define price impact to be the difference between the price you could have transacted at on paper (the average of the bid and ask at the time of the decision to trade) and the price you actually transacted at, whether immediately following the decision to trade or later. For example, if you buy at the ask (or sell at the bid) prevailing at the time of the decision to trade, your price impact will be half the bid–ask spread.

Price impact may occur because you have to move the market temporarily away from its current price in order to induce someone to supply the liquidity you are seeking. From time to time, there may be negative price impact, because you are able to take advantage of someone on the other side who needs the liquidity more than you do. When the price impact is purely a liquidity effect, the price of the stock will usually return to the level it was at before you traded.

Price impact may occur also because the market suspects you know something. Think of the block trader who has to find the other side of the trade for you. If you often show up with "soiled merchandise," he is going to go out of business if he always accommodates you at current prices and bags his clients on your behalf. More likely, he will adjust the price somewhat. The smarter he thinks you are, the bigger the adjustment. Once you have traded, the price may not return to its previous level because the cat is now out of the bag. In that case, part of the price impact will be permanent.

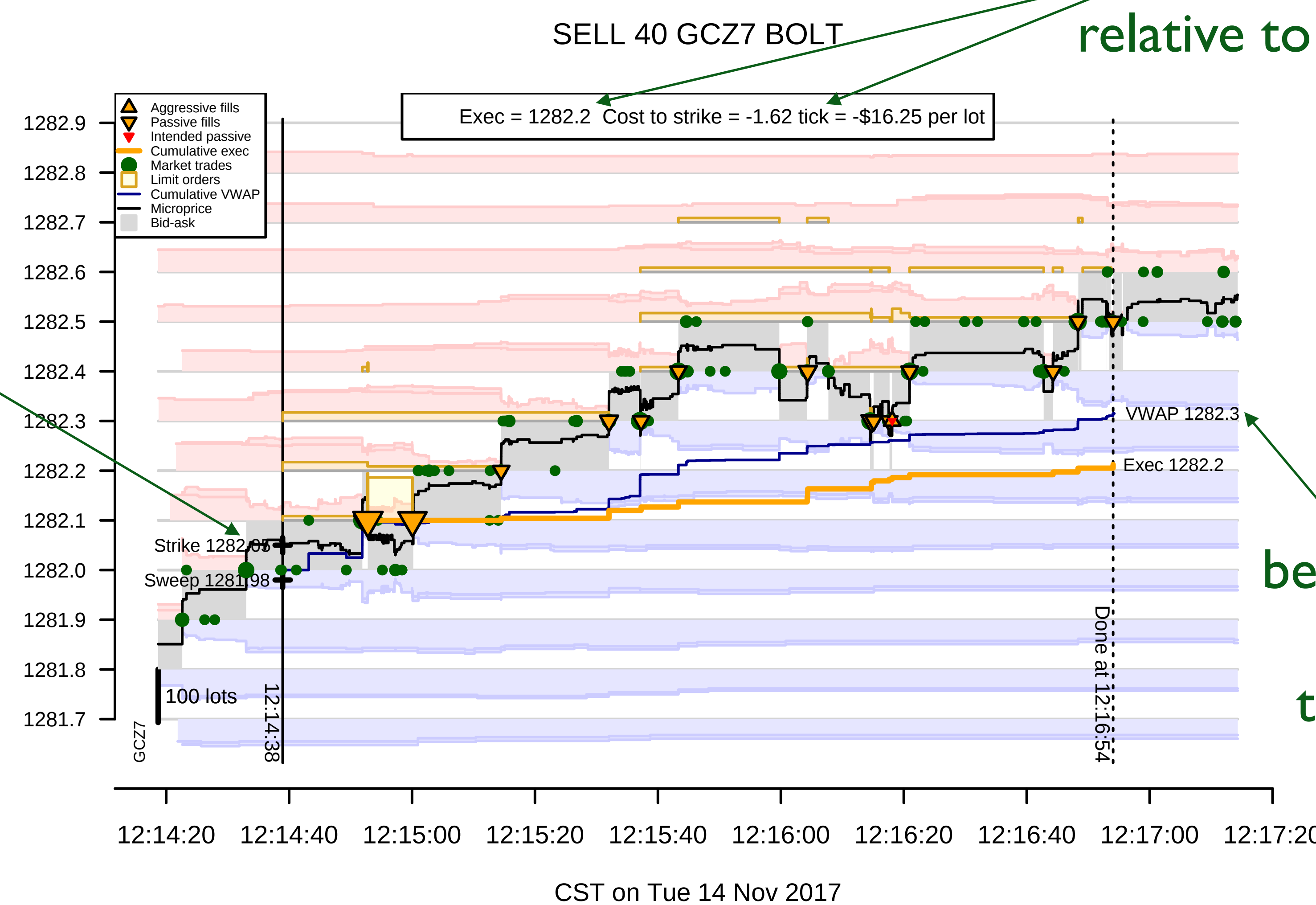
Different benchmarks and algorithms (QB)

- Arrival price
- Average price on interval (TWAP or VWAP)
- Settlement price
- Multi-leg target price
- etc (many more possibilities)

Arrival price

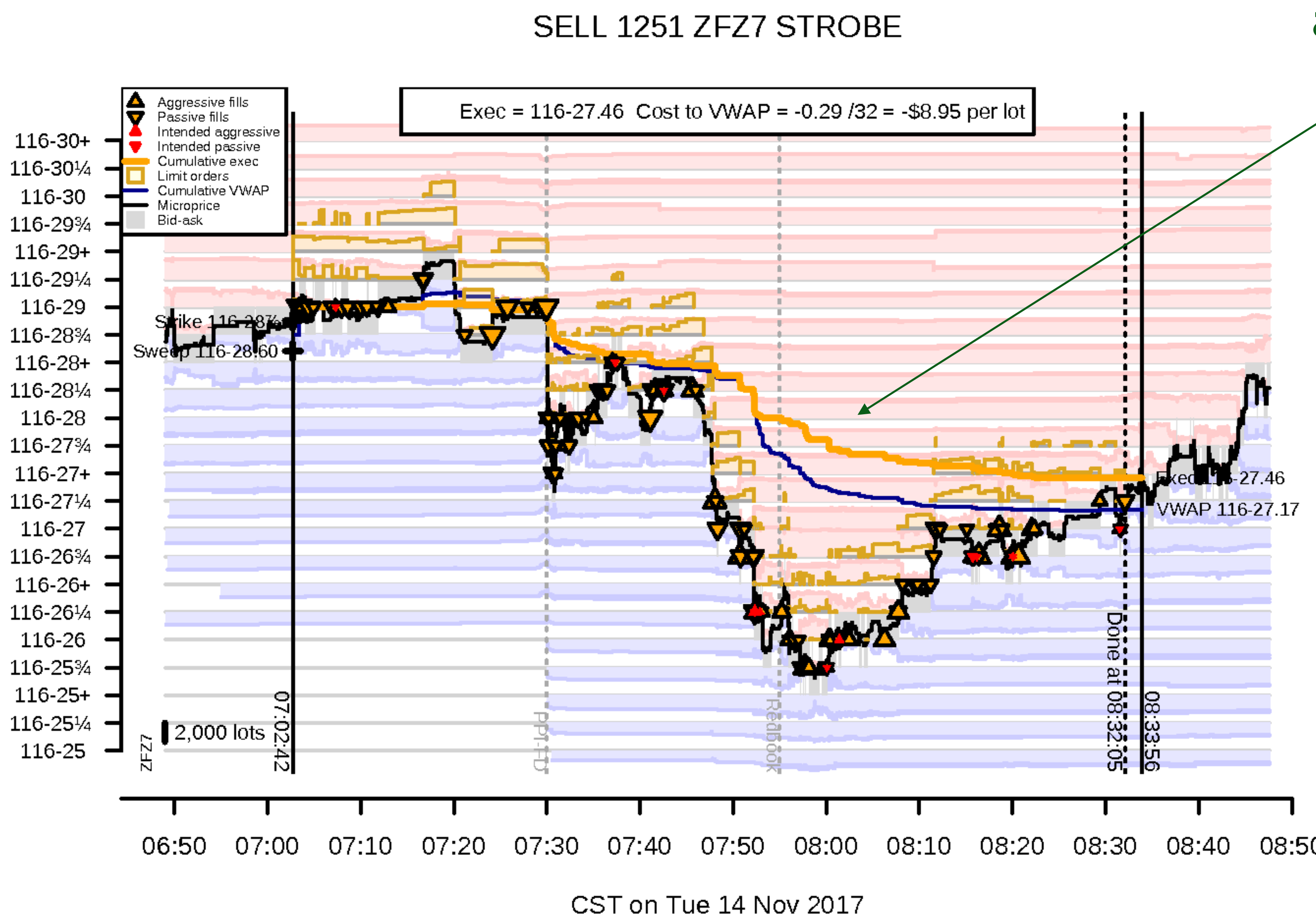
Arrival price benchmark ("strike")

Report execution price and slippage relative to benchmark



Also report other benchmarks for interest (but these are not targeted by this algo)

Average price on interval

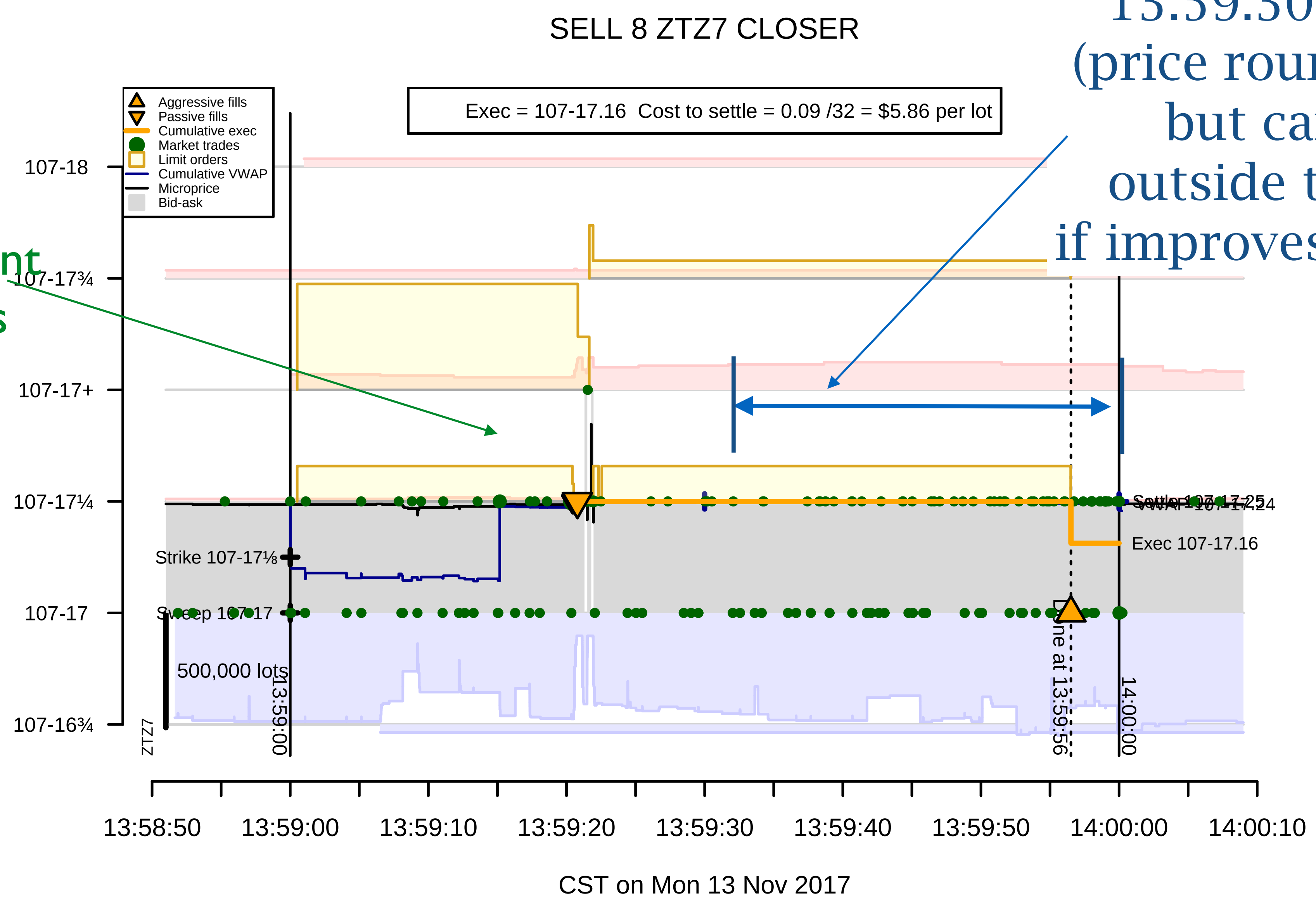


Execution approximately follows volume curve, but opportunistic when can improve performance

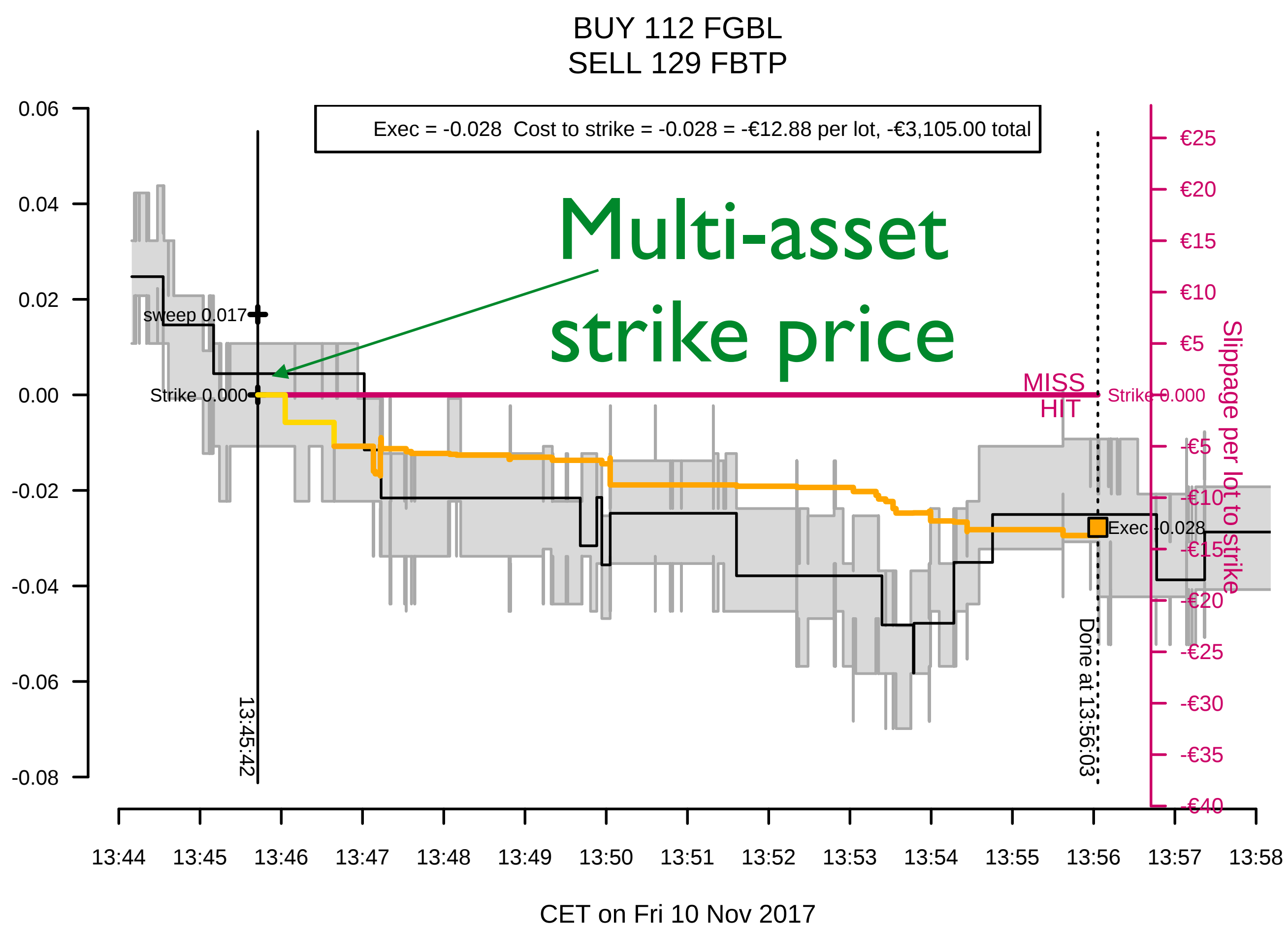
Closer: settlement window

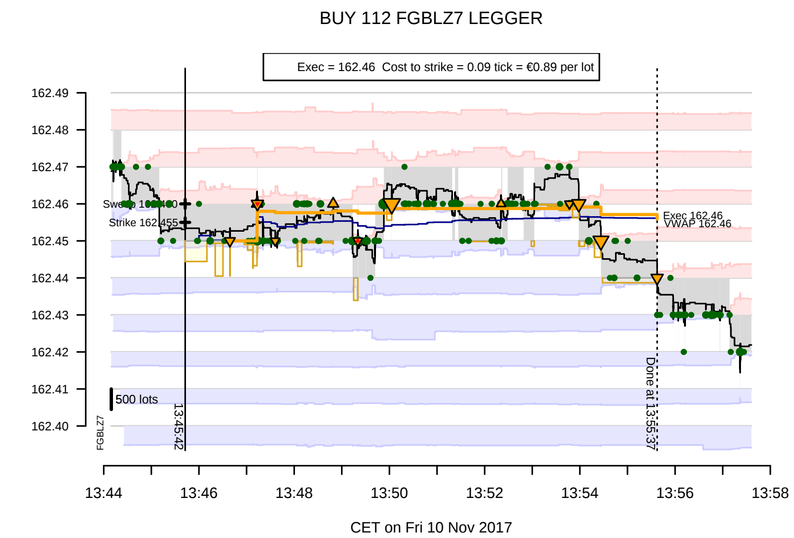
Passive fill before settlement window opens

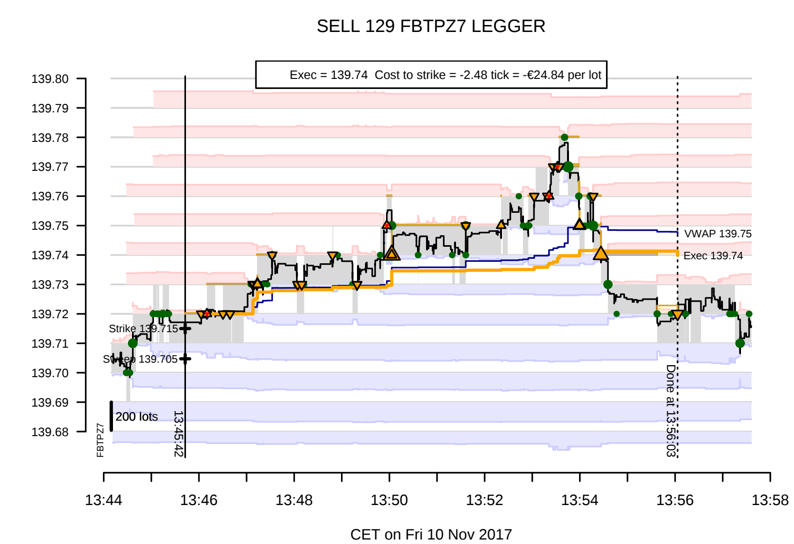
Settlement price =VWAP 13:59:30 to 14:00:00 (price rounded to grid)) but can execute outside that window if improves performance



Legger: multi-asset strike price







We focus on arrival price

- Arrival price is most fundamental
  - represents trade completed immediately at decision price
- Arrival price is cleanest benchmark
  - reference point is in past, not affected by trading
- Arrival price is most challenging to model
  - market direction is biggest contributor
  - lots of statistical noise
  - market impact and alpha are inextricable

Goal of market impact modeling

- Predict

  Slippage for a particular contemplated order

  buy 500 10-year Treasury over the next 2 hours

  Dependence on variables that can be adjusted

  what happens if we trade 200 or 1000, or take 3 hours?

- Uses
  - Trade decision-making

    how should we choose execution parameters?

  - Post-trade analysis

    what products / brokers / traders were good or bad relative to model?

Data resources

- Unfortunately, available only to industry participants

  brokers executing for clients

  asset managers executing their own orders

- Large variety of orders executed in past

  thousands per day

  many different products and market conditions

- 'What happened the last time that we did something "like" this?'

Caveat: Endogeneity

- Decision to trade depends on market
  send buy order when positive alpha signal
- Size of trade may depend on strength of signal
  would look like more impact for larger orders
- Other traders may follow same signal
  trade flow is correlated
- Execution may be adjusted depending on market

Paradox of slippage analysis

- Micro models do not predict macro models
- Micro models:

  impact is roughly linear in trade size

  impact is additive

  impact depends on time
- Macro models (empirical):

  impact is nonlinear in trade size (square root)

  impact is not additive

  execution time seems not to matter much.

Price impact trajectory

234 *The Impact of Metaorders*

TRADES, QUOTES AND PRICES

Financial Markets Under the Microscope

JEAN-PHILIPPE BOUCHAUD

Capital Fund Management, Paris

JULIUS BONART

University College London

JONATHAN DONIER

Spotify

MARTIN GOULD

Spotify

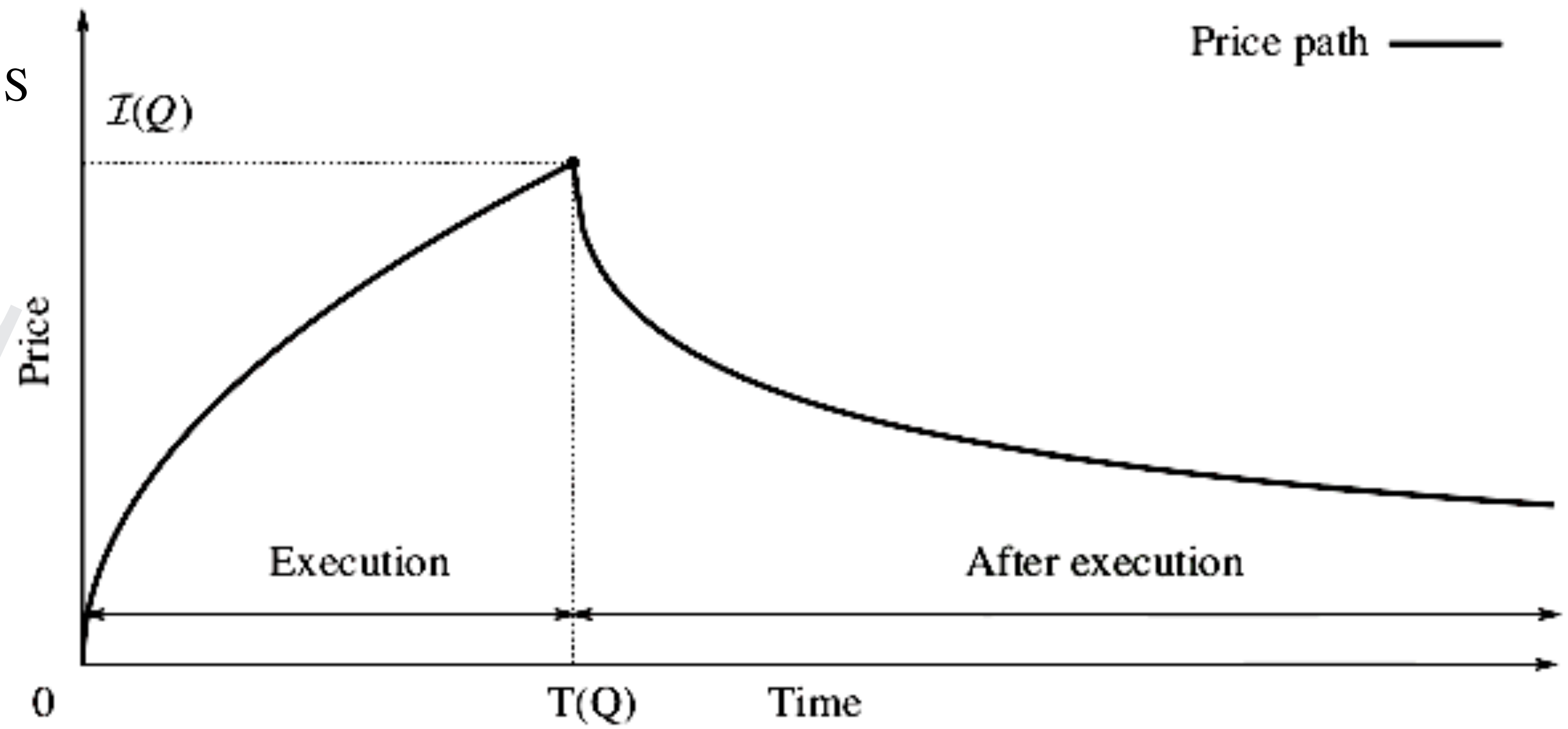


Figure 12.1. Average shape of the impact path. Over the course of its execution, a buy metaorder pushes the price up, until it reaches a peak impact. Upon completion, the buying pressure stops and the price reverts abruptly. Some impact is however still observable long after the metaorder execution is completed, and sometimes persists permanently. For real data, one should expect a large dispersion around the average impact path.

2005 model

Quantitative trading | Cutting edge

Equity market impact

The impact of large trades on prices is very important and widely discussed, but rarely measured. Using a large data set from a major bank and a simple but realistic theoretical model, Robert Almgren, Chee Thum, Emmanuel Hauptmann and Hong Li propose that impact is a 3/5 power law of block size, with specific dependence on trade duration, daily volume, volatility and shares outstanding. The results can be directly incorporated into an optimal trade scheduling algorithm and pre- and post-trade cost estimation

A. Distinguishing features of our model

Advantages
- Calibrated from real data
- Includes time component
- Incorporates intra-day profiles
- Uses non-linear impact functions
- Confidence levels for coefficients

Disadvantages
- Based only on Citigroup data
- Little data for small-cap stocks
- Little data for very large trades

WWW.RISK.NET • JULY 2005 RISK

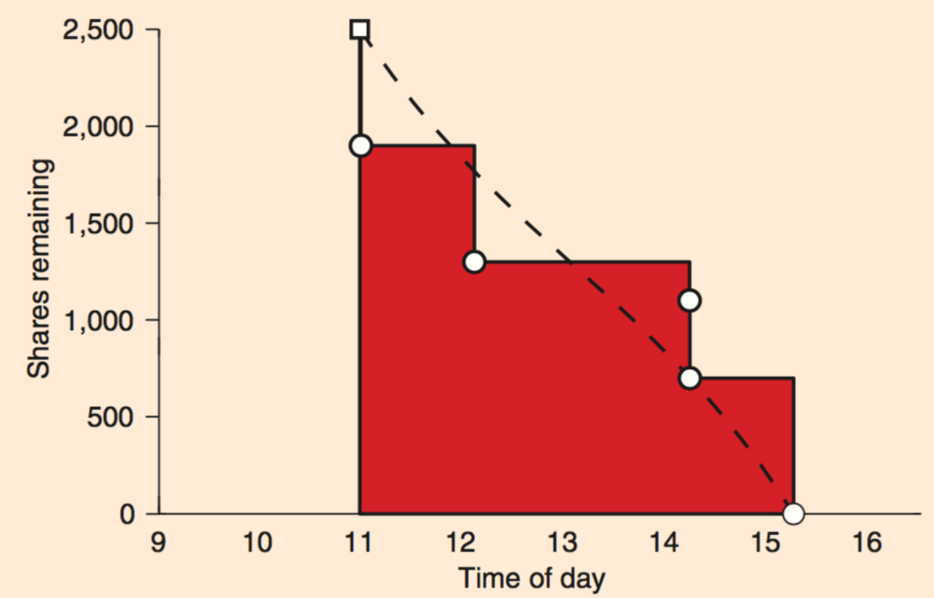
Available data is typical of institutional trade records

- The stock symbol, requested order size (number of shares) and sign (buy or sell) of the entire order. Client identification is removed.
- The times and methods by which transactions were submitted by the Citigroup trader to the market. We take the time t_0 of the first transaction to be the start of the order. Some of these transactions are sent as market orders, some are sent as limit orders, and some are submitted to Citigroup's automated VWAP server. Except for the starting time t_0, and except to exclude VWAP orders, we make no use of this transaction information.
- The times, sizes, and prices of execution corresponding to each transaction. Some transactions are cancelled or only partially executed; we use only the completed price and size. We denote execution times by $t_1, \ldots, t_n$, sizes by $x_1, \ldots, x_n$, and prices by $S_1, \ldots, S_n$.

December 2001 through June 2003.

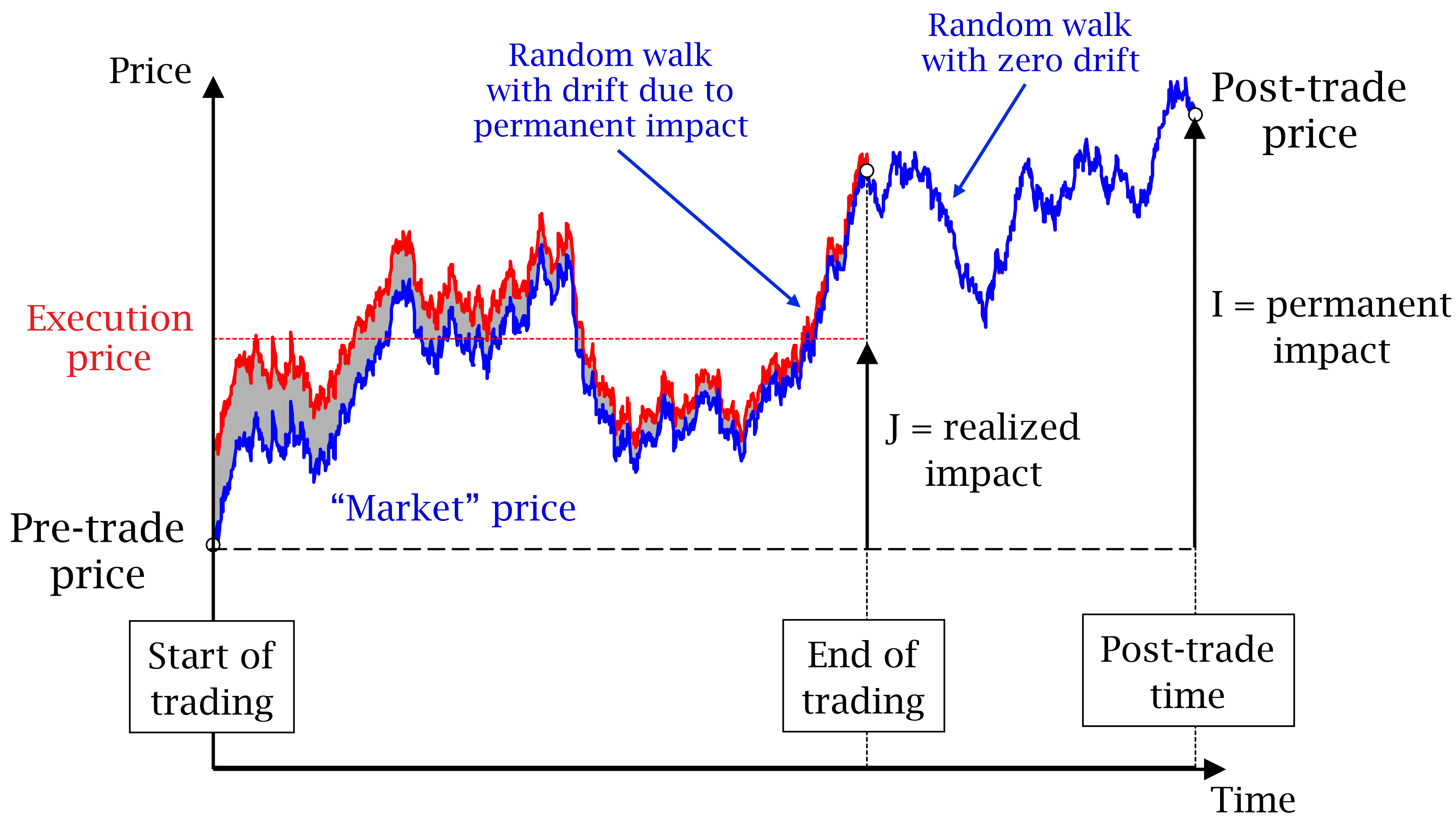
29,509 orders in our data set.

1. A typical trading trajectory



The vertical axis is shares remaining and each step downwards is one execution. The trajectory starts at the first transaction recorded in the system; the program ends when the last execution has been completed. The dashed line is our continuous-time approximation

Permanent and temporary impact



$$
\begin{aligned}
S_0 &= \text{market price before this order begins executing} \\
S_{\text{post}} &= \text{market price after this order is completed} \\
\bar{S} &= \text{average realized price on the order}
\end{aligned}
$$

$$
\begin{aligned}
\text{Permanent impact:} \quad I &= \frac{S_{\text{post}} - S_0}{S_0} \\
\text{Realized impact:} \quad J &= \frac{\bar{S} - S_0}{S_0}.
\end{aligned}
$$

Conclusion of 2005 study

Micro model:
can use this for trajectory optimization

micro model:

$$dS = S_0\, g(v)\, d\tau + S_0\, \sigma\, dB,$$

Integrate:

$$\frac{I}{\sigma} = \gamma\, T\, \mathrm{sgn}(X) \left|\frac{X}{VT}\right|^{\alpha} \left(\frac{\Theta}{V}\right)^{\delta} + \langle\text{noise}\rangle \tag{7}$$

$$\frac{1}{\sigma}\left(J - \frac{I}{2}\right) = \eta\, \mathrm{sgn}(X) \left|\frac{X}{VT}\right|^{\beta} + \langle\text{noise}\rangle \tag{8}$$

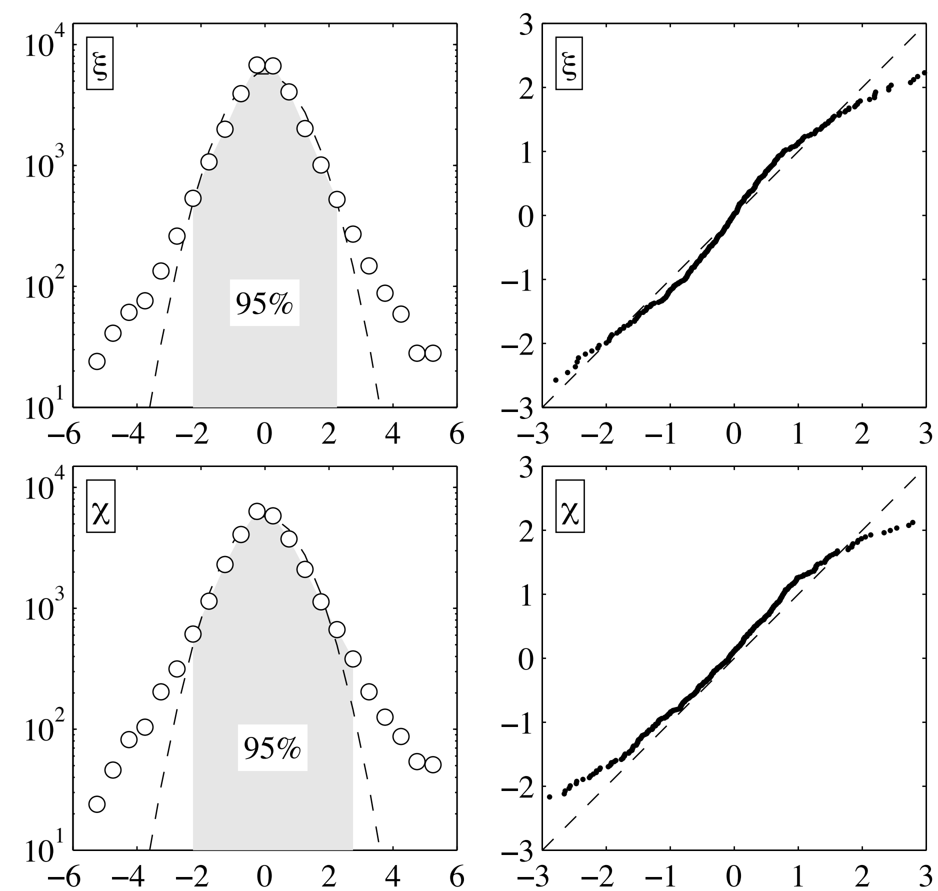
Calibrate to actual data

$$\frac{I}{\sigma} = \gamma\, T\, \mathrm{sgn}(X) \left|\frac{X}{VT}\right|^{\alpha} \left(\frac{\Theta}{V}\right)^{\delta} + \langle\text{noise}\rangle \tag{7}$$

$$\frac{1}{\sigma}\left(J - \frac{I}{2}\right) = \eta\, \mathrm{sgn}(X) \left|\frac{X}{VT}\right|^{\beta} + \langle\text{noise}\rangle \tag{8}$$

$$\alpha = 0.891 \pm 0.10$$
$$\delta = 0.267 \pm 0.22$$
$$\beta = 0.600 \pm 0.038.$$

Residuals are non-normal:
very fat tails
(like all market data)



Liquidity impact is concave
as function of trade size (β<1)
Close to square root (β=0.5) but different

General problem of macro cost modeling

- Given lots of historical data
  orders executed by algorithm
  with all child order information
- Formulate some kind of practical model
  to meet trading needs

Classic problem in statistics / machine learning

- One output variable: slippage
- Many input variables:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \longrightarrow y$$

Order parameters:

symbol, side, size, start time, duration, etc

Market parameters known before trading:

forecast volume, volatility, spread, quote size

real-time volume, volatility, spread, quote size

Market parameters discovered during trading:

price direction (most important)

evolution of volume, volatility, etc

Analytical techniques

$$y = f(x_1, \dots, x_n)$$

$$y = \alpha + \beta_1 x_1 + \cdots + \beta_n x_n$$

- Regression
  - specific function depending on parameters
  - easy interpretation, not always accurate
- Supervised learning
  - many powerful modern techniques
    - neural nets, trees, support vector machines, etc
  - not always easy to interpret

Challenge in market impact modeling

- Very large noise relative to signal
- Criterion of minimum discrepancy is hard to apply
- Main criterion: residuals, etc, should not depend other variables

Strategy

1. Single asset fitting
2. Multi-asset fitting across universe of futures products

Example:

- ES (SP500 futures) trades,
  Sep 2018 through Dec 2019

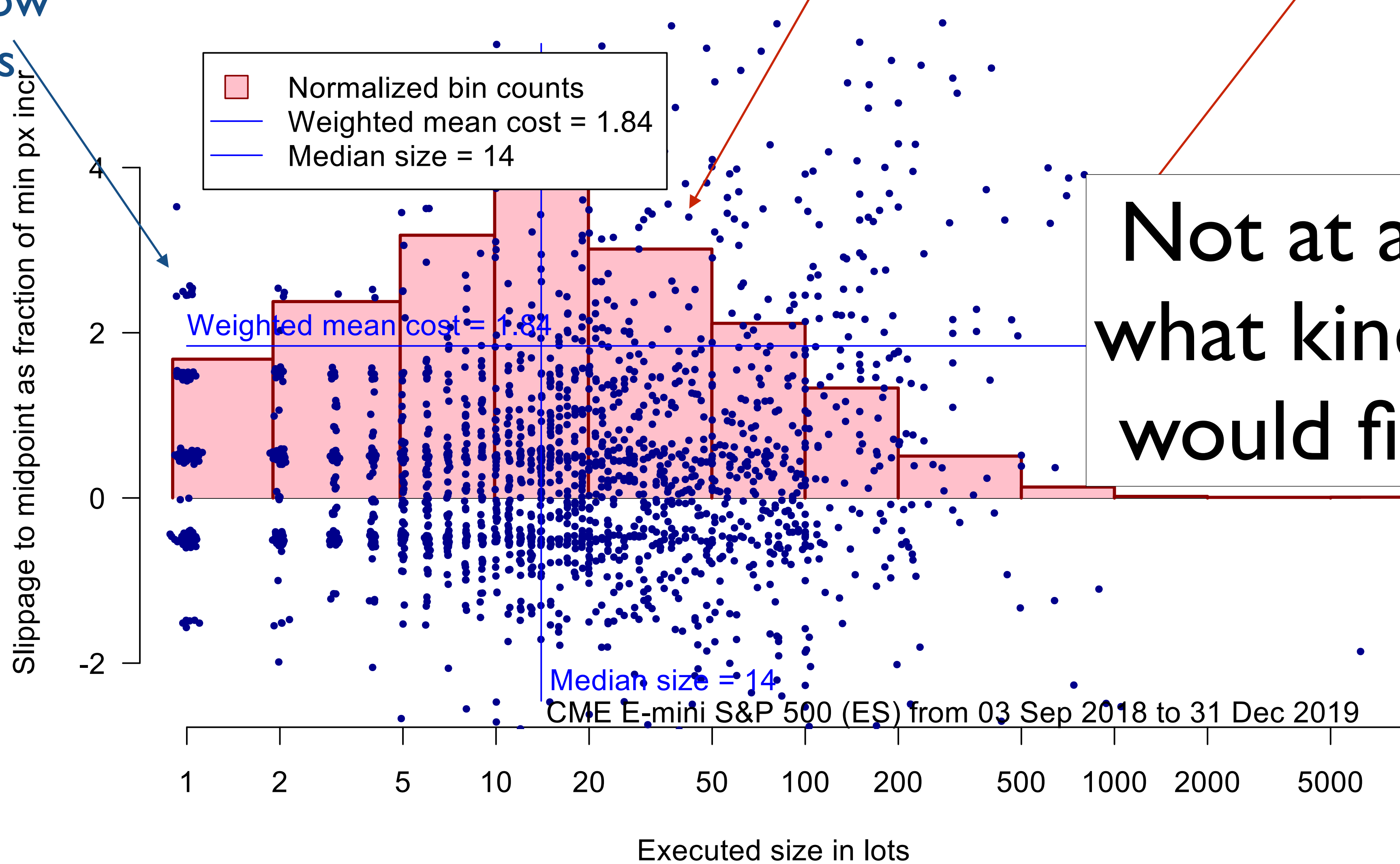
  all clients merged together (except private data)
- Outright contracts only
- Mostly weighted toward front month
- Price units are in exchange's minimum price increment
- For "market impact", most important variable is size

  hypothesis: large trades have higher slippage

Majority of orders are small

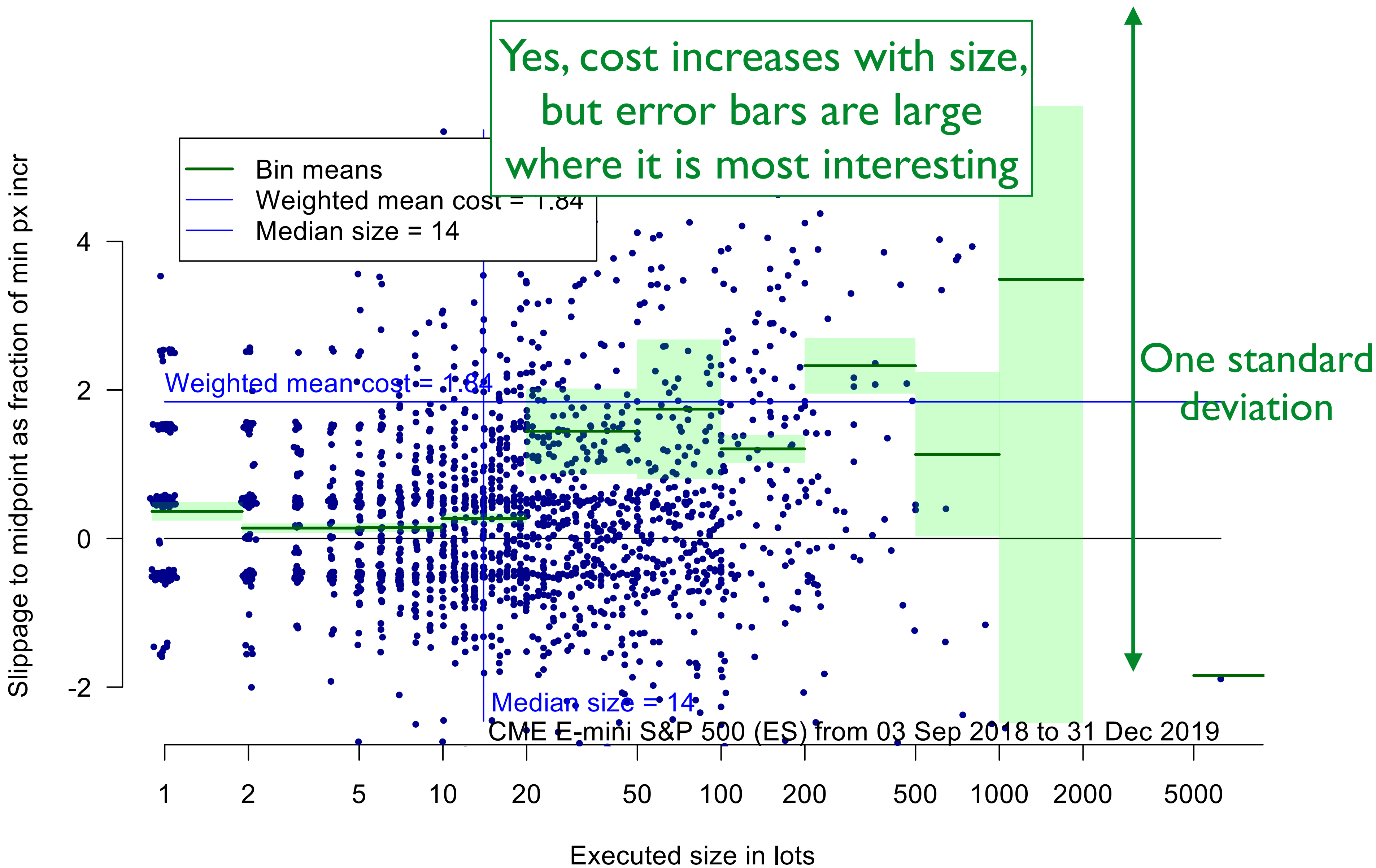
But cost is experienced on these few large orders

Slight "jitter" to show overlapping points

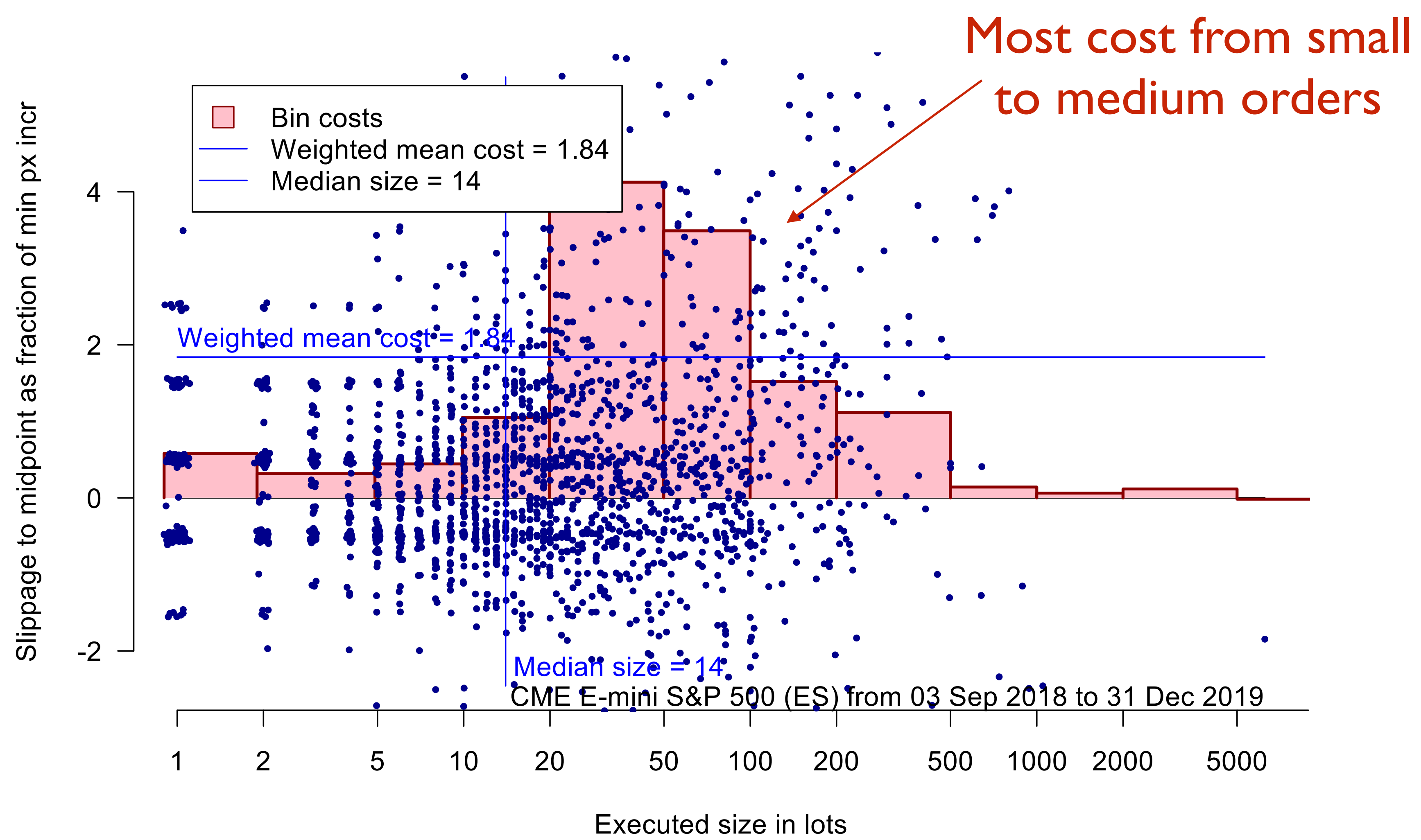


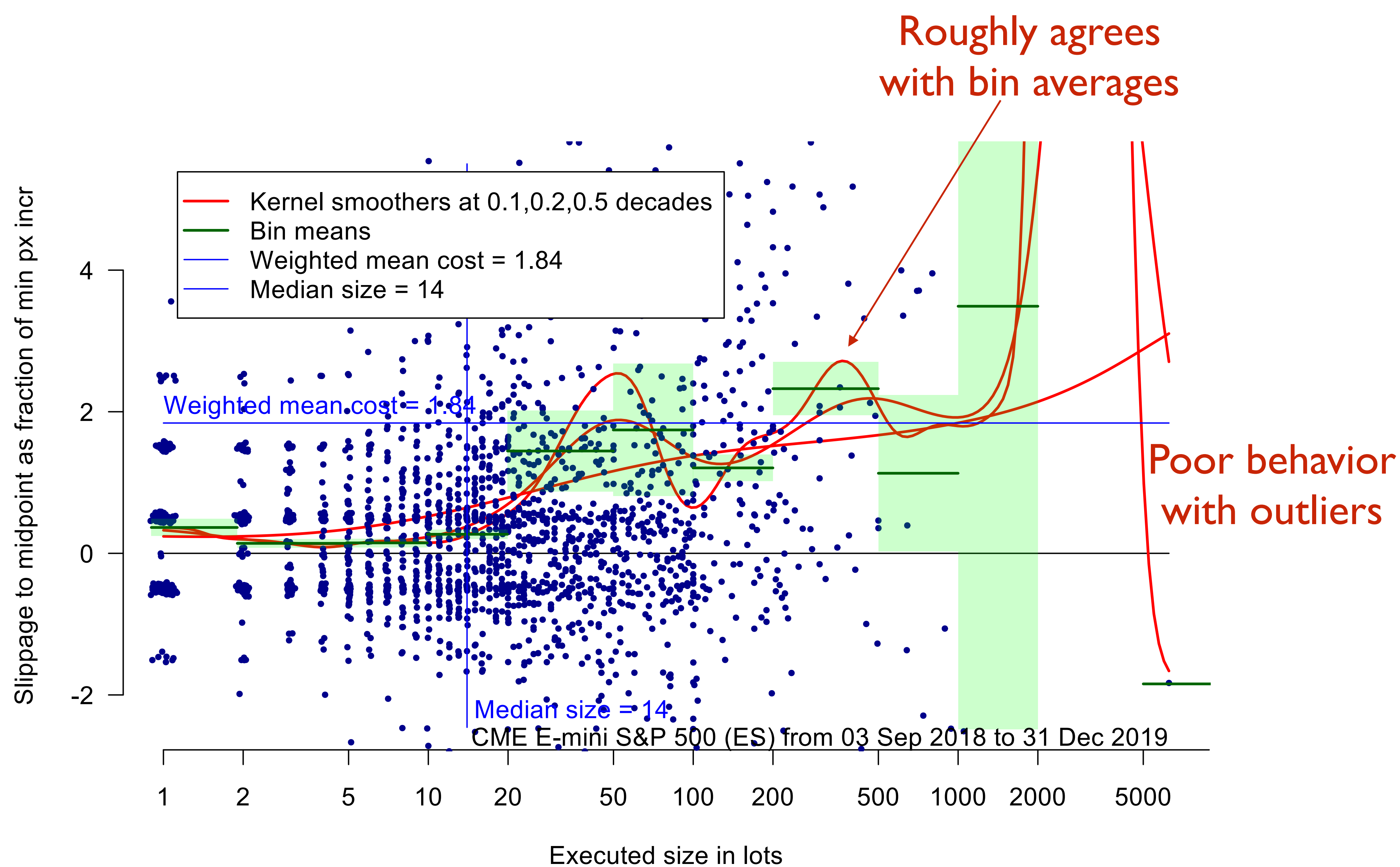
Not at all obvious what kind of model would fit this data

Choice of size as unique independent variable (for now)



Yes, cost increases with size, but error bars are large where it is most interesting
Bin means
Weighted mean cost = 1.84
Median size = 14
Weighted mean cost = 1.84
Median size = 14
CME E-mini S&P 500 (ES) from 03 Sep 2018 to 31 Dec 2019
One standard deviation
Slippage to midpoint as fraction of min px incr
Executed size in lots
4
2
0
-2
1
2
5
10
20
50
100
200
500
1000
2000
5000


Most cost from small
to medium orders
Bin costs
Weighted mean cost = 1.84
Median size = 14
Weighted mean cost = 1.84
Median size = 14
CME E-mini S&P 500 (ES) from 03 Sep 2018 to 31 Dec 2019
Slippage to midpoint as fraction of min px incr
Executed size in lots
4
2
0
-2
1
2
5
10
20
50
100
200
500
1000
2000
5000


Roughly agrees
with bin averages
Poor behavior
with outliers
Kernel smoothers at 0.1,0.2,0.5 decades
Bin means
Weighted mean cost = 1.84
Median size = 14
Weighted mean cost = 1.84
Median size = 14
CME E-mini S&P 500 (ES) from 03 Sep 2018 to 31 Dec 2019
Slippage to midpoint as fraction of min px incr
Executed size in lots
4
2
0
-2
1
2
5
10
20
50
100
200
500
1000
2000
5000

Parametric model

$$y = a + b\,x^{\gamma}$$

$$\min_{a,b,\gamma} \sum_{j=1}^{n} \left(a + b x_j^{\gamma} - y_j\right)^2$$

$$x = \text{order size}$$
$$y = \text{slippage}$$

Two linear coefficients a,b

One exponent γ

Coefficients a, b determined linearly for each exponent γ

γ determined by one-dimensional minimization (easy!)

Fractional-power model is consistent with literature

In all of these cases, the peak impact of a metaorder with volume Q is well described by the relationship:

$$\Im^{\text{peak}}(Q,T) \cong Y\sigma_T\left(\frac{Q}{V_T}\right)^{\delta}, \qquad (Q \ll V_T) \tag{12.6}$$

where Y is a numerical coefficient of order 1 ($Y \cong 0.5$ for US stocks), δ is an exponent in the range 0.4–0.7, σ_T is the contemporaneous volatility on the time horizon T, and V_T is the contemporaneous volume traded over time T.

TRADES, QUOTES AND PRICES

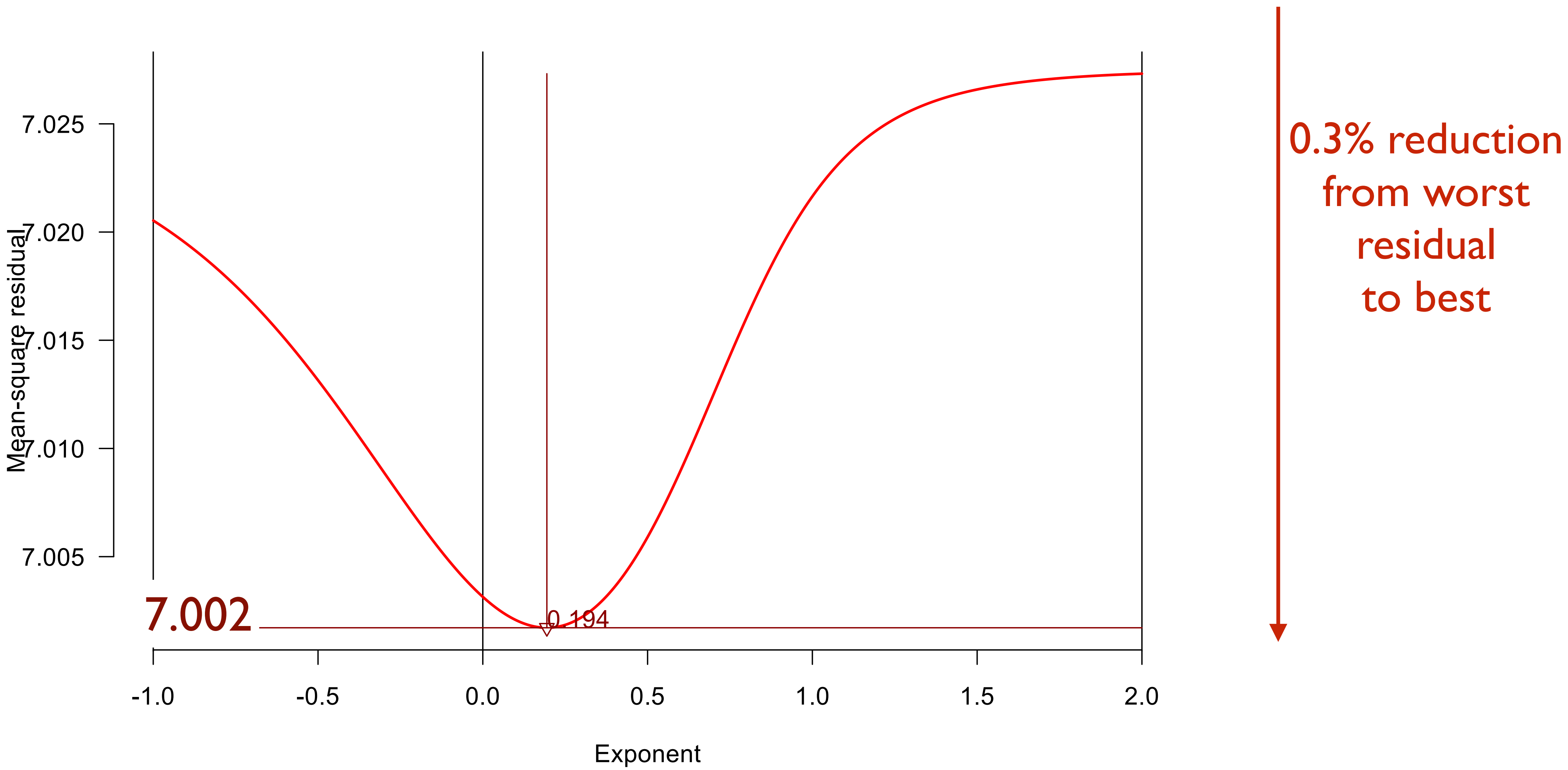
Financial Markets Under the Microscope

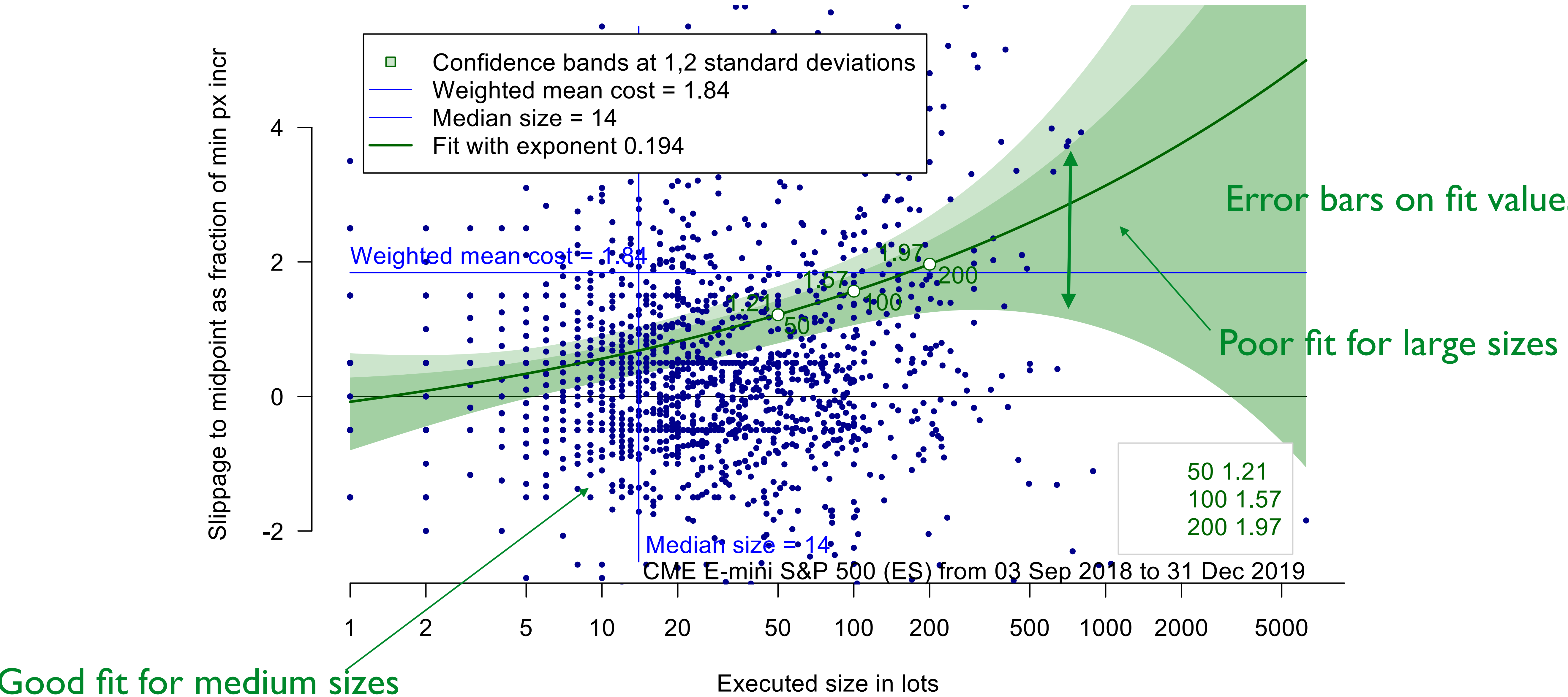
JEAN-PHILIPPE BOUCHAUD
Capital Fund Management, Paris

JULIUS BONART
University College London

JONATHAN DONIER
Spotify

MARTIN GOULD
Spotify

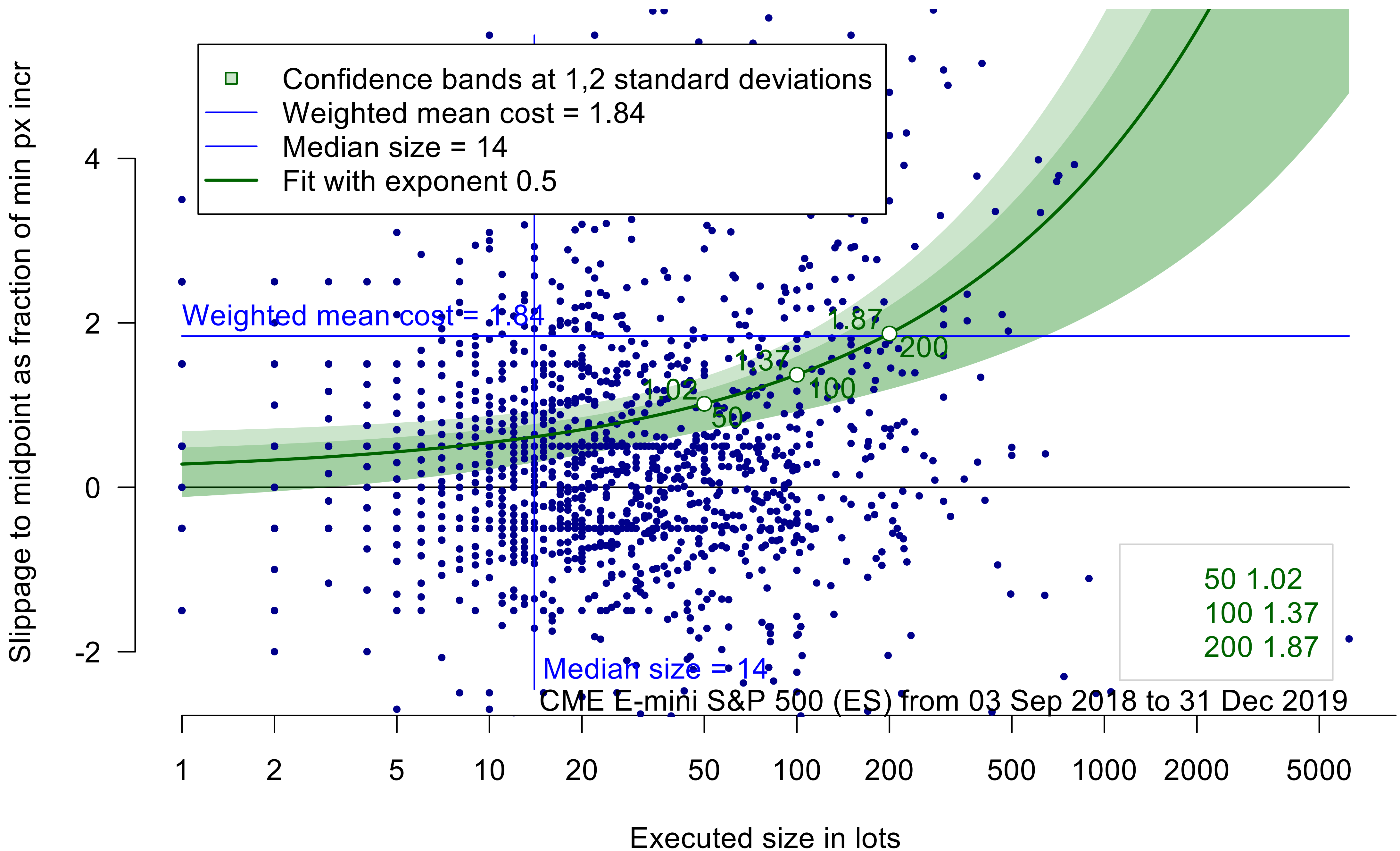

Mean-square residual
7.025
7.020
7.015
7.010
7.005
7.002
0.194
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Exponent
0.3% reduction
from worst
residual
to best



Confidence bands at 1,2 standard deviations
Weighted mean cost = 1.84
Median size = 14
Fit with exponent 0.194
Slippage to midpoint as fraction of min px incr
4
2
0
-2
Weighted mean cost = 1.84
Median size = 14
1.21
50
1.57
100
1.97
200
50 1.21
100 1.57
200 1.97
CME E-mini S&P 500 (ES) from 03 Sep 2018 to 31 Dec 2019
1
2
5
10
20
50
100
200
500
1000
2000
5000
Executed size in lots
Error bars on fit values
Poor fit for large sizes
Good fit for medium sizes

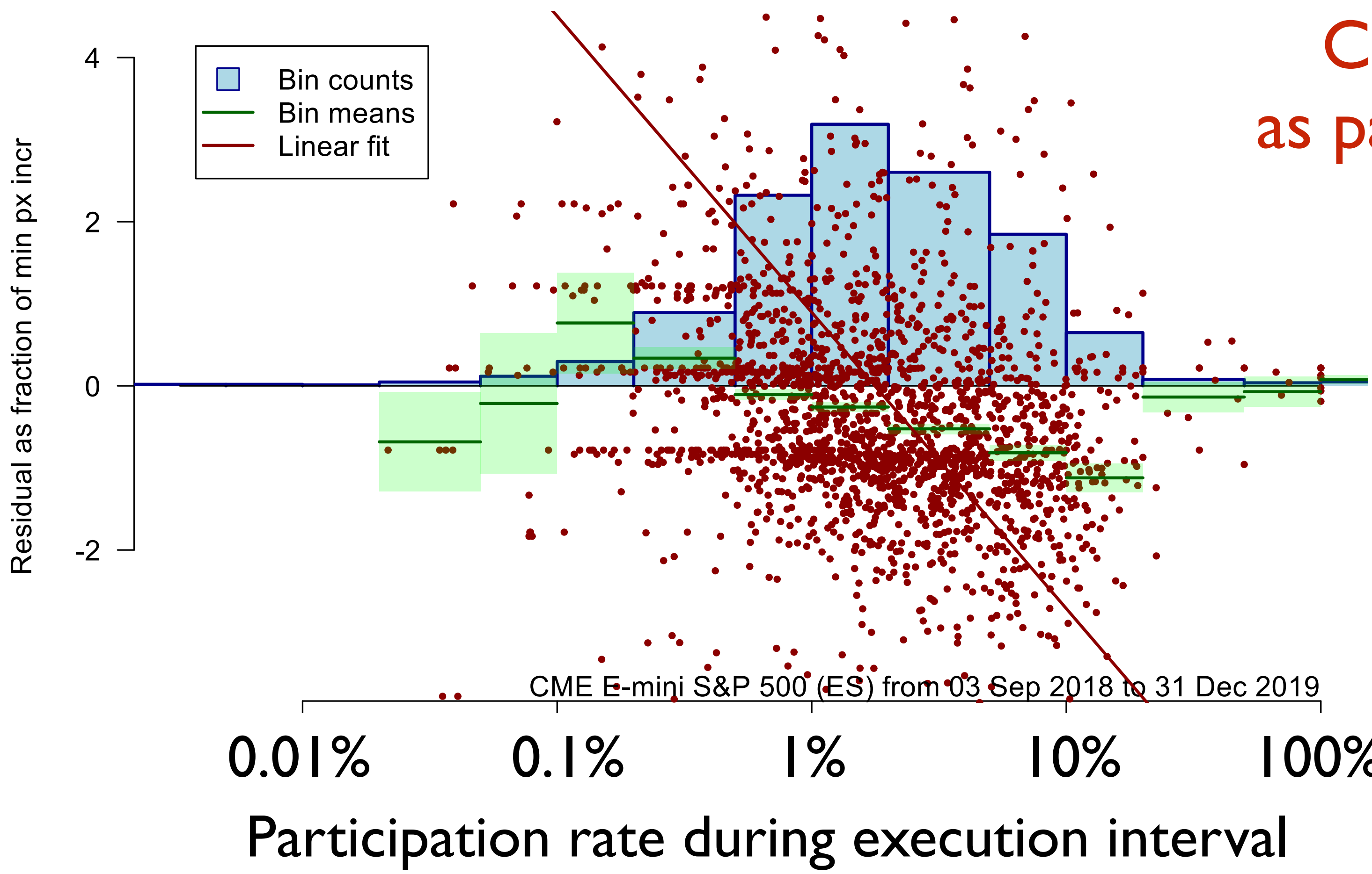
Strong reasons to prefer k = 0.5

(will explain in a few minutes)



Is order duration (or participation) a relevant variable?
Executing faster should cost more

Participation rate = trade size / interval volume

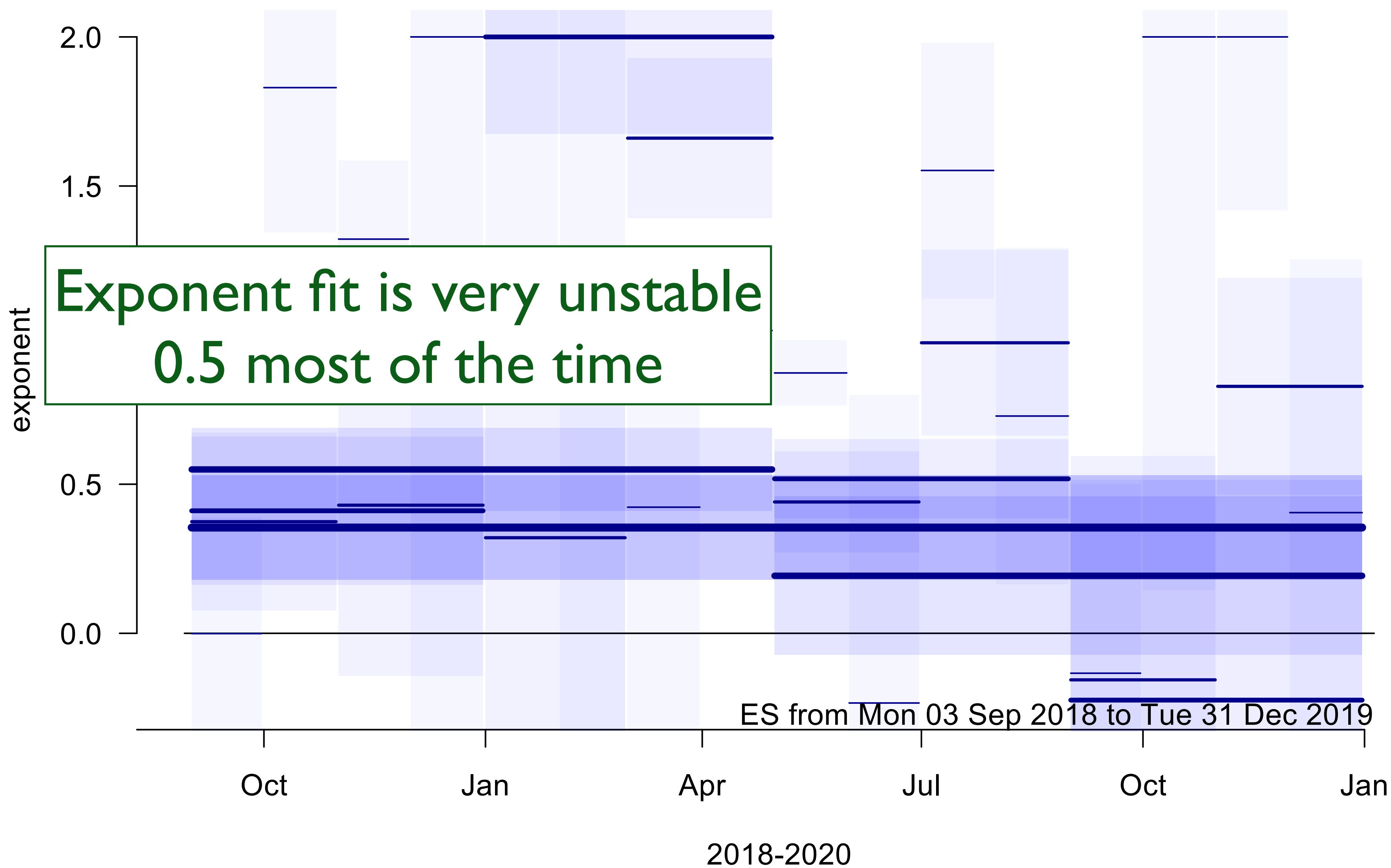


Cost decreases as participation rate increases

Participation rate depends on how quickly the order fills.

Participation rate is a *dependent* variable, not *independent*.

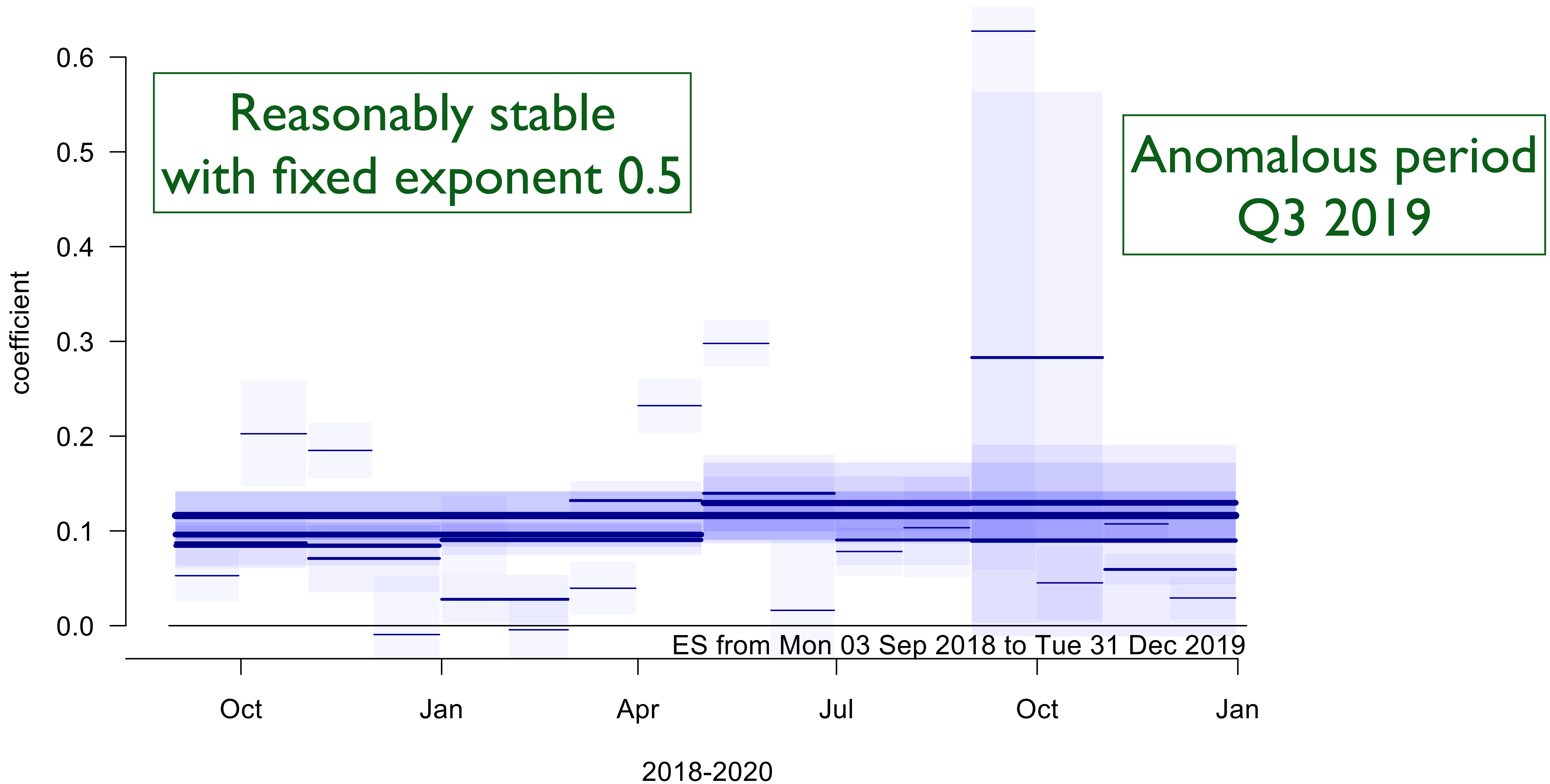
Stability of exponent



Fit values on

- 16 months
- 8 months
- 4 months
- 2 months
- 1 month

Stability of coefficient with fixed exponent



Conclusions of single-asset fitting

- Fractional-power model gives reasonable agreement
  Settle on exponent $k = 0.5$
- Want to fit important part of parameter range (50-100)
- Neglect participation rate

Conclusions of single-asset fitting

- Fractional-power model gives reasonable agreement
- Want to fit important part of parameter range (50-100)
- Residual magnitude depends weakly on exponent
- Neglect participation rate

Comparison across products

- How does exponent vary across products?

  Problem: very noisy fits for each product separately
- Hypothesis: fit properties vary systematically with other measurable properties of products
- large-tick vs small-tick

Multi-asset fitting

- Challenge: wide range of products
- Not enough data for each one to fit individually
- How to group together?

Distribution of number of orders

Number of distinct orders

Futures contract number

Good statistics up here

Not enough data down here-- need to combine with other products

Nondimensionalization: simple incorporation of market properties

$$\frac{C}{\sigma} = f\left(\frac{X}{V}\right) + \text{error} \qquad f(x) = a + bx^{\gamma}$$

Normalize trade size by market volume, impact (price change) by daily volatility

C = slippage

X = trade size

σ = forecast daily volatility

V = forecast daily volume

PHYSICAL REVIEW X **1**, 021006 (2011)

Anomalous Price Impact and the Critical Nature of Liquidity in Financial Markets

B. Tóth, Y. Lempérière, C. Deremble, J. de Lataillade, J. Kockelkoren, and J.-P. Bouchaud
Capital Fund Management, 6, blvd Haussmann 75009 Paris, France
(Received 9 May 2011; published 31 October 2011)

We propose a dynamical theory of market liquidity that predicts that the average supply/demand profile is V shaped and *vanishes* around the current price. This result is generic, and only relies on mild assumptions about the order flow and on the fact that prices are, to a first approximation, diffusive. This naturally accounts for two striking stylized facts: First, large metaorders have to be fragmented in order to be digested by the liquidity funnel, which leads to a long memory in the sign of the order flow. Second, the anomalously small local liquidity induces a breakdown of the linear response and a diverging impact of small orders, explaining the "square-root" impact law, for which we provide additional empirical support. Finally, we test our arguments quantitatively using a numerical model of order flow based on the same minimal ingredients.

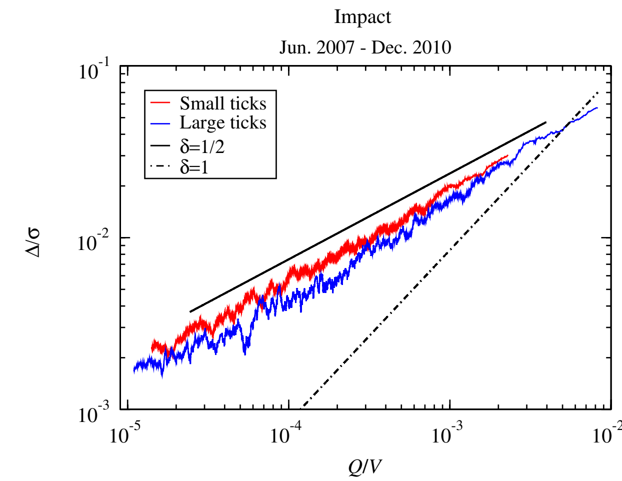


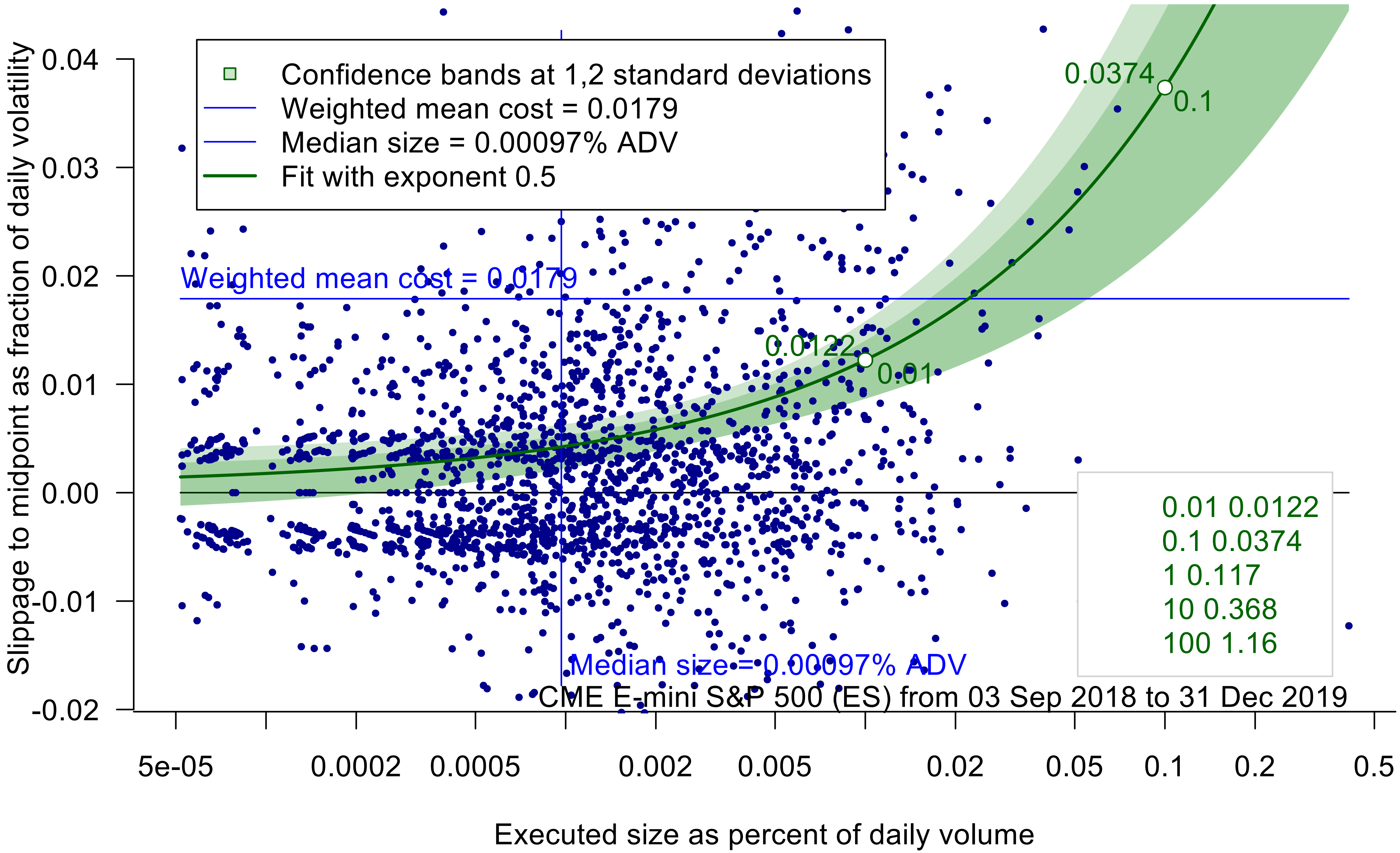
FIG. 1. The impact of metaorders for Capital Fund Management proprietary trades on futures markets, in the period from June 2007 to December 2010. Impact is measured here as the average execution shortfall of a metaorder of size Q.

Preference for exponent 1/2: sqrt(volume) and volatility both scale linearly with time Incorporates changing market conditions

the average relative price change Δ between the first and the last trade of a metaorder of size Q is well described by the so-called "square-root" law:

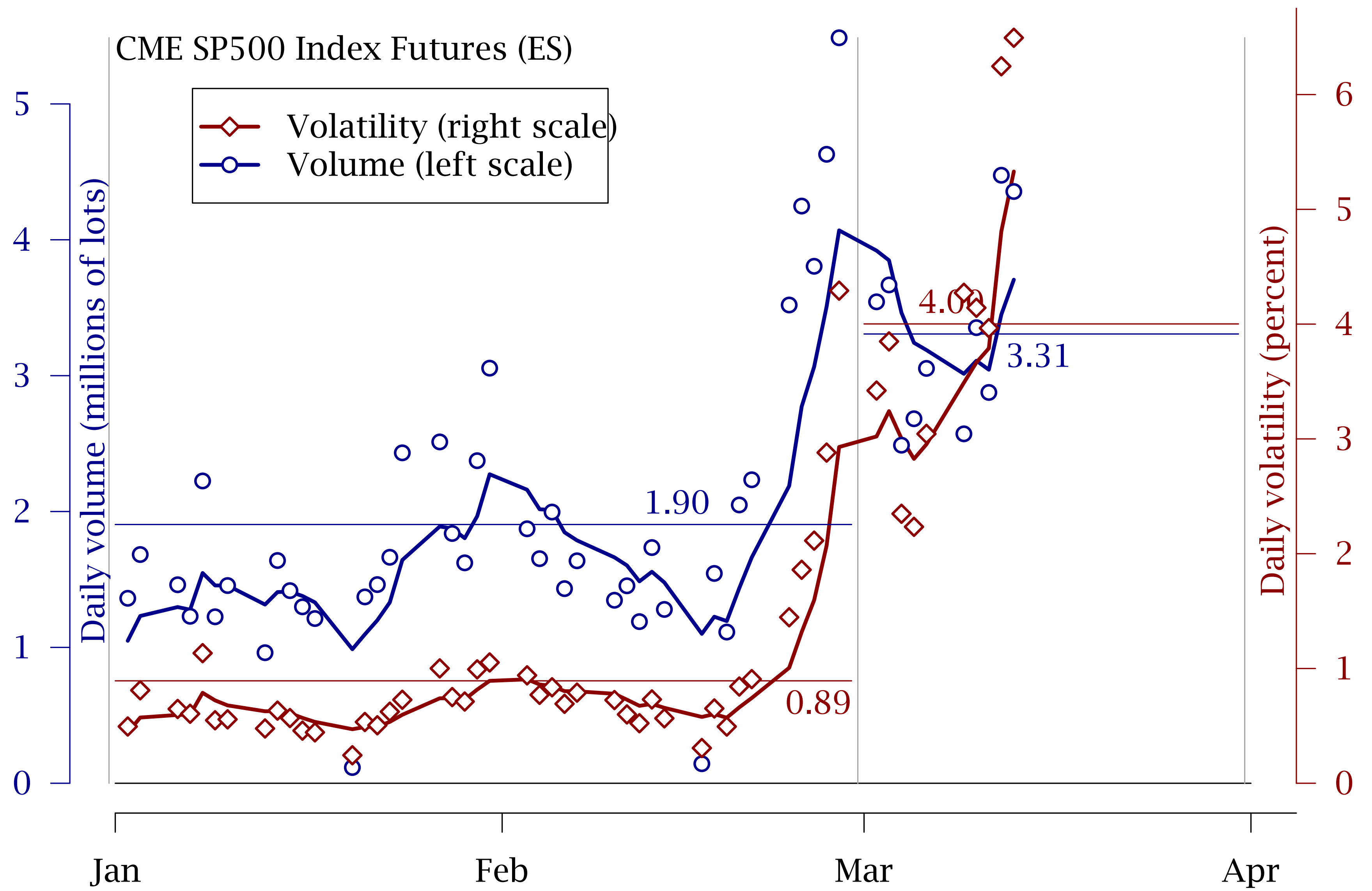
$$\Delta(Q) = Y\sigma\sqrt{\frac{Q}{V}}, \qquad (1)$$

where σ is the daily volatility of the asset and V is the daily traded volume, and both quantities are measured contemporaneously to the trade. The numerical constant Y is of order unity. Published and unpublished data suggest slightly different versions of this law; in particular, the $\sqrt{Q}$ dependence is more generally described as a power-law relation $\Delta(Q) \propto Q^{\delta}$, with δ in the range 0.4 to 0.7

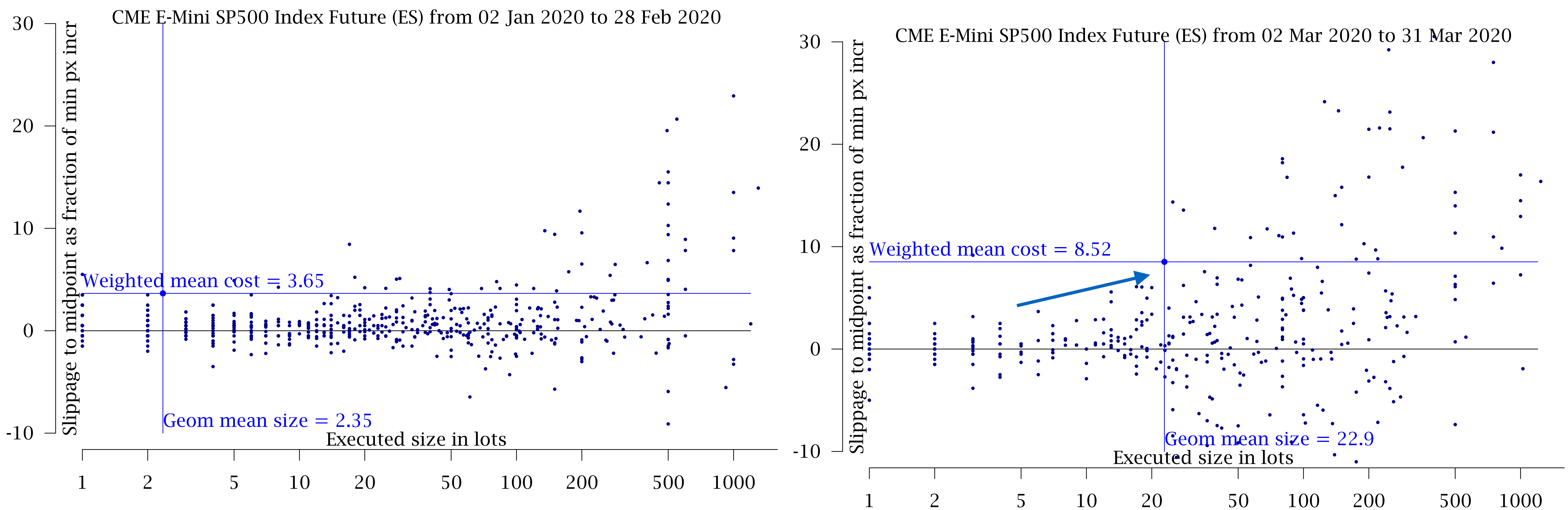

Confidence bands at 1,2 standard deviations
Weighted mean cost = 0.0179
Median size = 0.00097% ADV
Fit with exponent 0.5
Weighted mean cost = 0.0179
Median size = 0.00097% ADV
0.0374
0.1
0.0122
0.01
0.01 0.0122
0.1 0.0374
1 0.117
10 0.368
100 1.16
CME E-mini S&P 500 (ES) from 03 Sep 2018 to 31 Dec 2019
Slippage to midpoint as fraction of daily volatility
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
5e-05
0.0002
0.0005
0.002
0.005
0.02
0.05
0.1
0.2
0.5
Executed size as percent of daily volume

Fit across time

March 2020: much increased volume and volatility compared with Jan/Feb

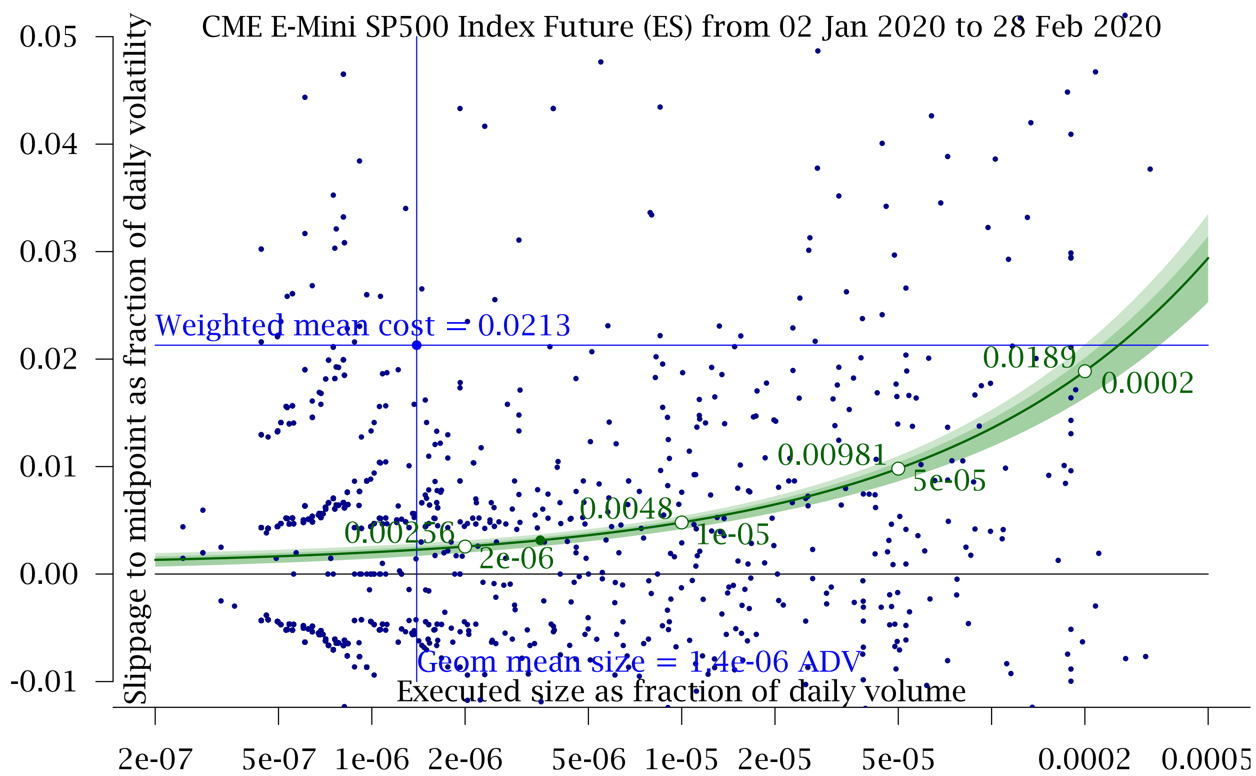


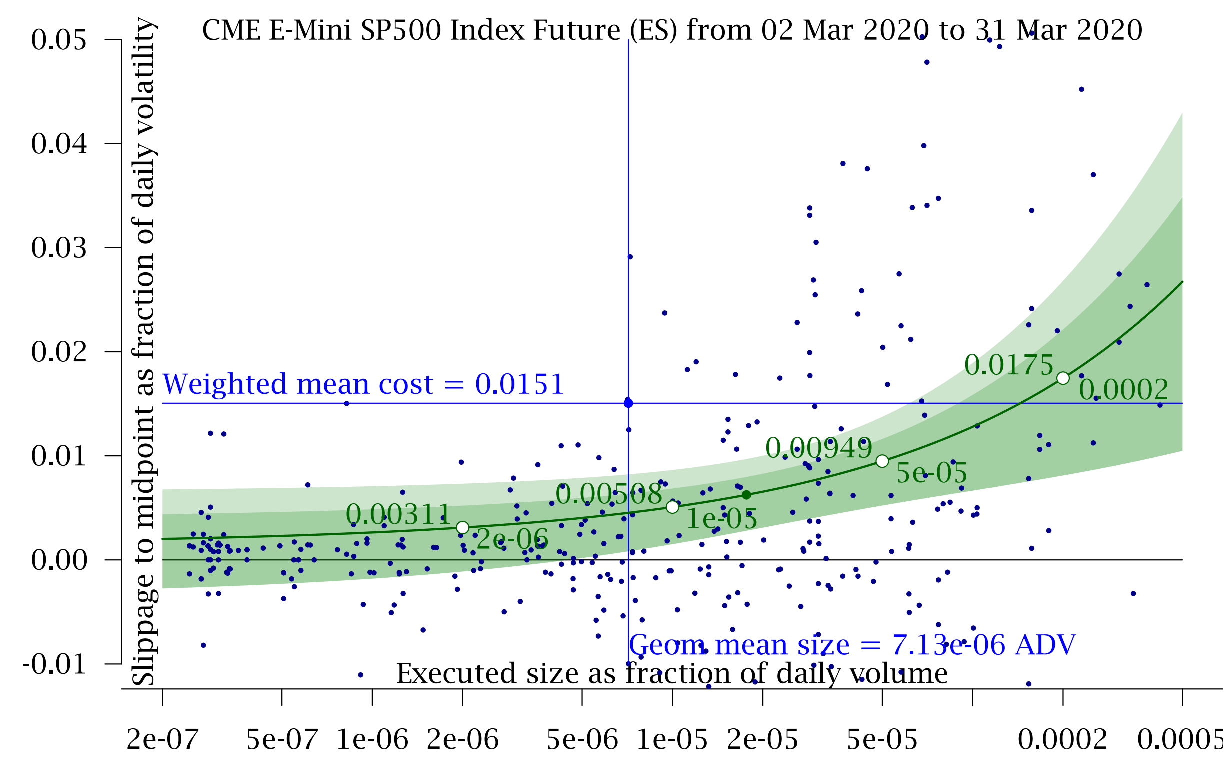
Order size and slippage distribution: Jan/Feb → March



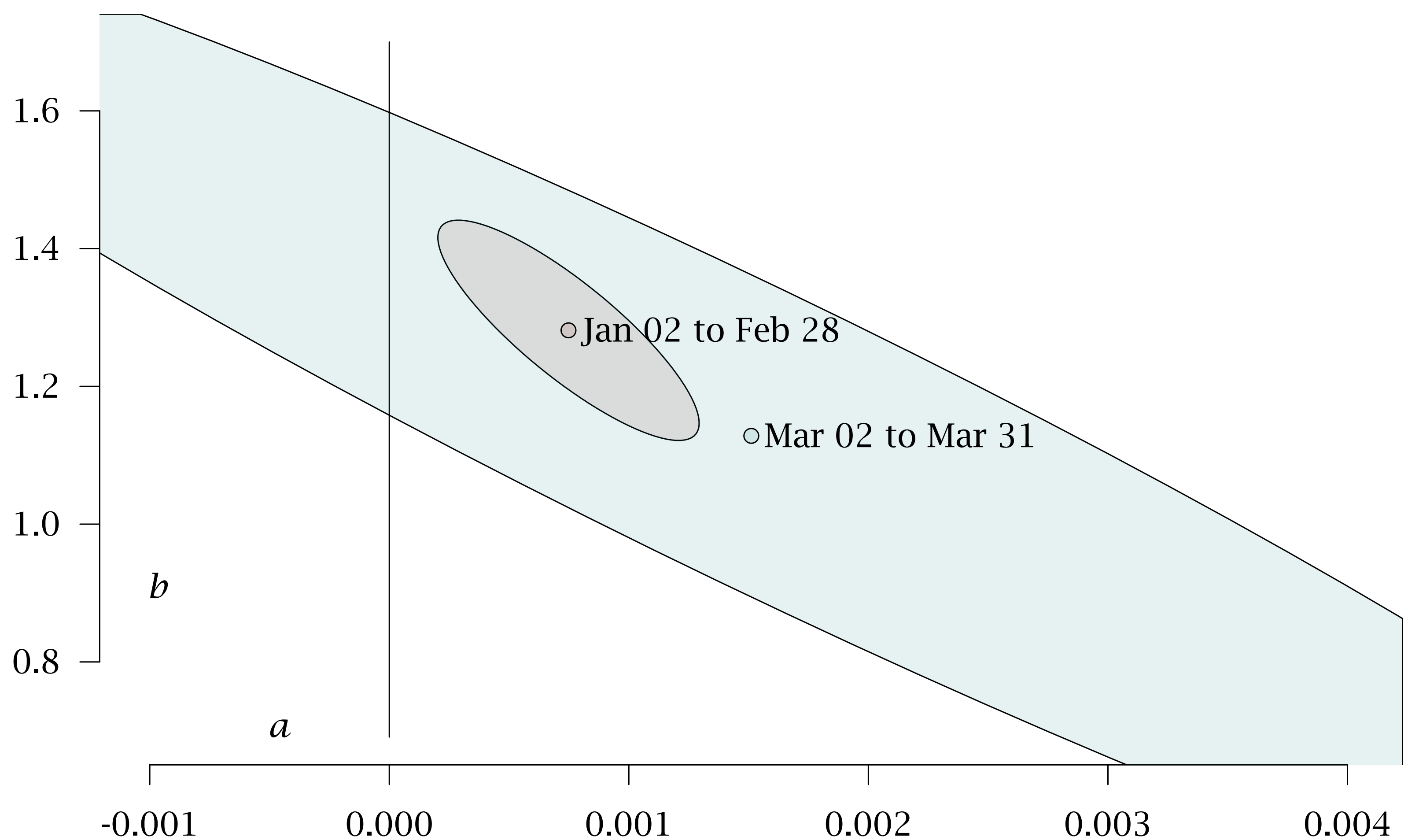
Larger orders, higher slippage

Nondimensional model is stable





Coefficient comparison



How to group products in general

1. Based on intrinsic properties of product:
   - tick size
   - liquidity
   - etc

This does not work

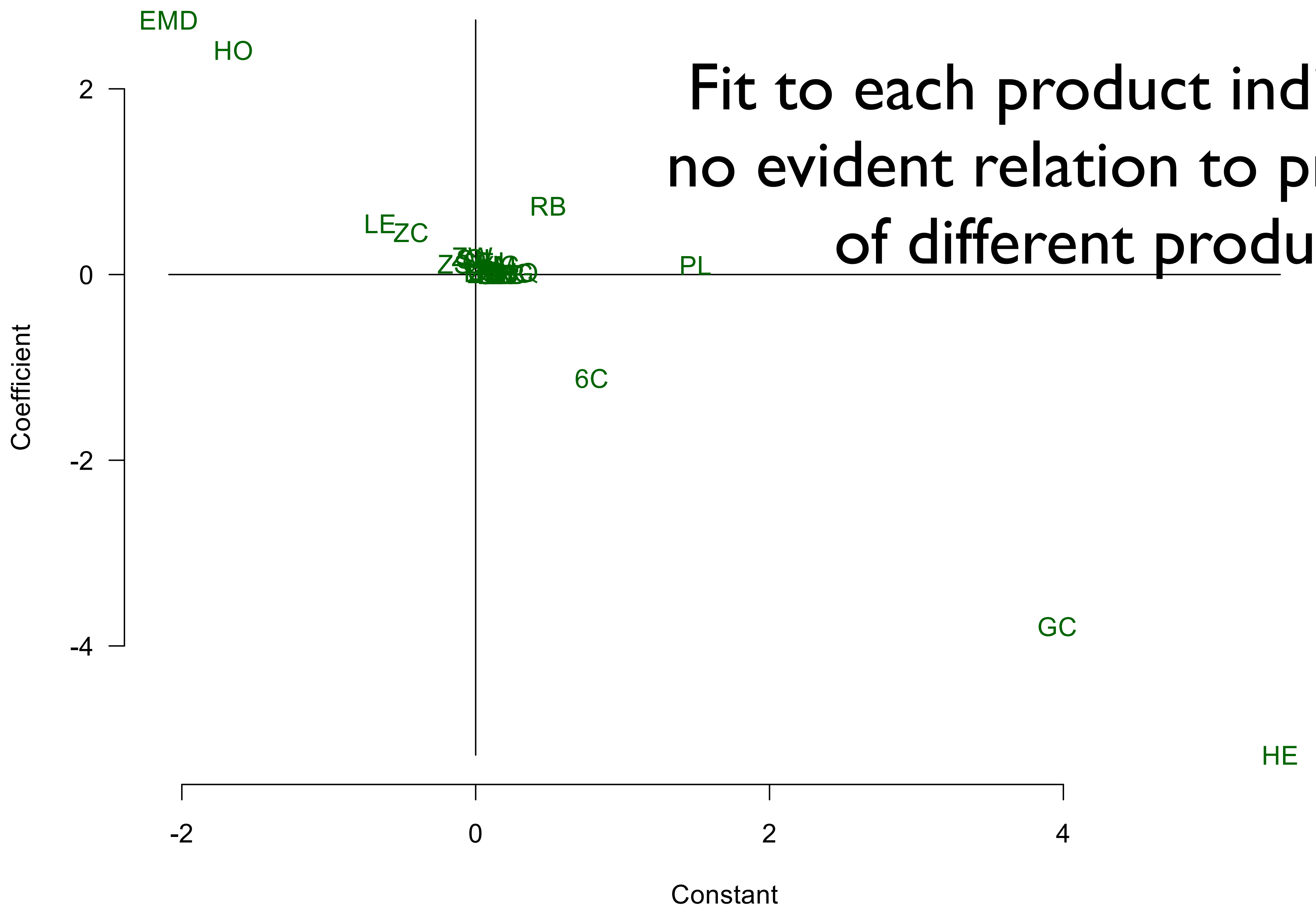
This works

2. Based on regression fit itself:
   - mean and variance of coefficients

Expectation

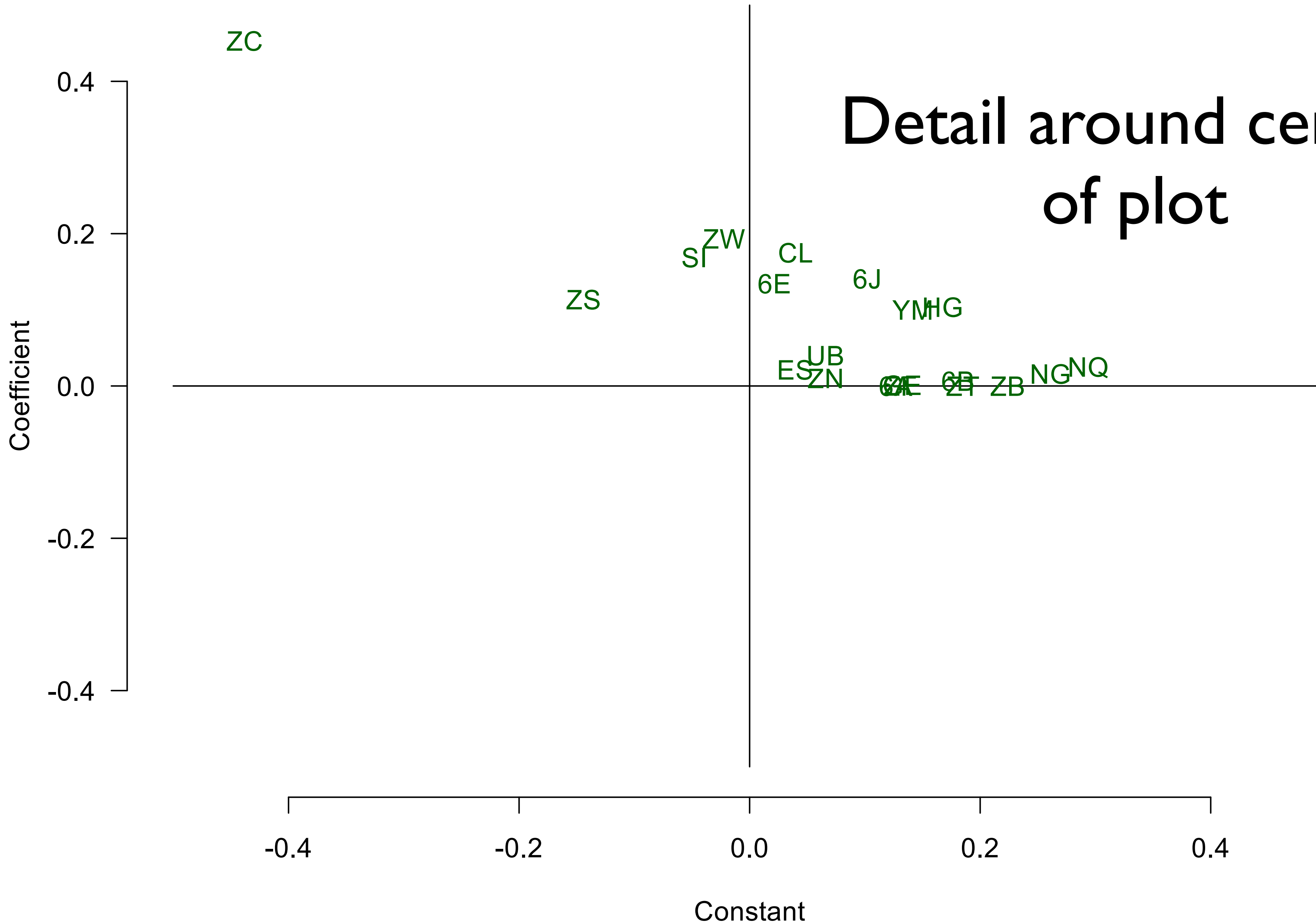
- For different products, normalizing by market values is enough to get same model for all products
- This more or less works across stocks (all are same things, but different traded volumes and volatilities)
- Does not work for futures

Fit across different futures products



Fit to each product individually: no evident relation to properties of different products

Detail around center of plot



Large tick vs small tick

Journal of Financial Econometrics, 2011, Vol. 9, No. 2, 344–366

A New Approach for the Dynamics of Ultra-High-Frequency Data: The Model with Uncertainty Zones

CHRISTIAN Y. ROBERT
CREST and ENSAE Paris Tech

MATHIEU ROSENBAUM
CMAP-École Polytechnique Paris UMR CNRS 7641

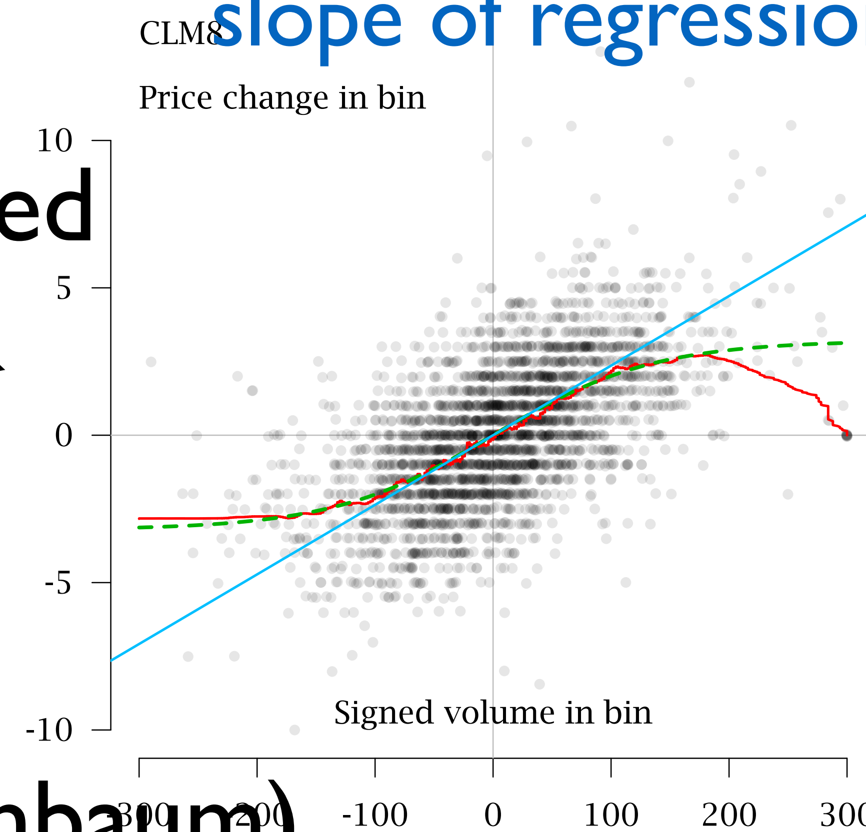
ABSTRACT
In this paper, we provide a model which accommodates the assumption of a continuous efficient price with the inherent properties of ultra-high-frequency transaction data (price discreteness, irregular temporal spacing, diurnal patterns...). Our approach consists in designing a stochastic mechanism for deriving the transaction prices from the latent efficient price. The main idea behind the model is that, if a transaction occurs at some value on the tick grid and leads to a price change, then the efficient price has been close enough to this value shortly before the transaction. We call uncertainty zones the bands around the mid-tick grid where the efficient price is too far from the tick grid to trigger a price change. In our setting, the width of these uncertainty zones quantifies the aversion to price changes of the market participants. Furthermore, this model enables us to derive approximated values of the efficient price at some random times, which is particularly useful for building statistical procedures. Convincing results are obtained through a simulation study and the use of the model over 10 representative stocks.

One can also see the parameter η as a measure of the relevance of the tick size on the market. Indeed, if $\eta < 1/2$, market participants are convinced they have to trade at a new price before the efficient price crosses this new price on the tick grid. So, it means that the tick size appears too large to them. Conversely, a large η ($\eta > 1/2$) means that the tick size appears too small. From the tick size perspective, an ideal market is consequently a market where η is equal to $1/2$.

A natural estimation procedure for the parameter η is given in Robert and Rosenbaum (2010a). We define an alternation (resp. continuation) of one tick as a price jump of one tick whose direction is opposite to (resp. the same as) the one of the preceding price jump. Let $N^{(a)}_{\alpha,t}$ and $N^{(c)}_{\alpha,t}$ be respectively the number of alternations and continuations of one tick over the period $[0,t]$. An estimator of η over $[0,t]$ is given by

$$\hat{\eta}_{\alpha,t} = \frac{N^{(c)}_{\alpha,t}}{2N^{(a)}_{\alpha,t}}.$$

Illiquidity =
slope of regression line



Liquidity
= price change per volume traded

Tick size
= average spread in terms of minimum price increment
= reversion ratio (Robert/Rosenbaum)
= average quote size / average trade size

Large tick vs small tick

Journal of Financial Econometrics, 2011, Vol. 9, No. 2, 344–366

A New Approach for the Dynamics of Ultra-High-Frequency Data: The Model with Uncertainty Zones

Christian Y. Robert
CREST and ENSAE Paris Tech
Mathieu Rosenbaum
CMAP-École Polytechnique Paris UMR CNRS 7641

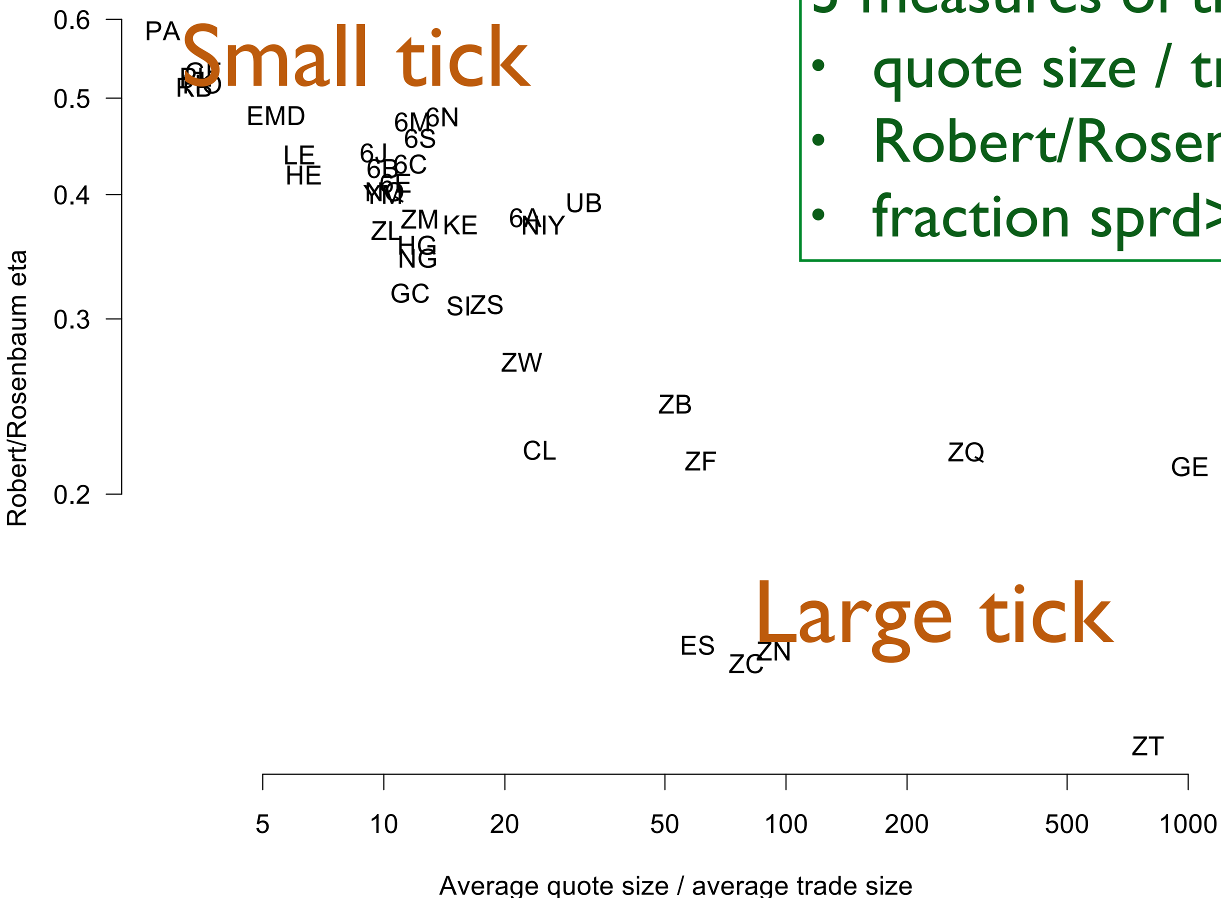
ABSTRACT
In this paper, we provide a model which accommodates the assumption of a continuous efficient price with the inherent properties of ultra-high-frequency transaction data (price discreteness, irregular temporal spacing, diurnal patterns...). Our approach consists in designing a stochastic mechanism for deriving the transaction prices from the latent efficient price. The main idea behind the model is that, if a transaction occurs at some value on the tick grid and leads to a price change, then the efficient price has been close enough to this value shortly before the transaction. We call uncertainty zones the bands around the mid-tick grid where the efficient price is too far from the tick grid to trigger a price change. In our setting, the width of these uncertainty zones quantifies the aversion to price changes of the market participants. Furthermore, this model enables us to derive approximated values of the efficient price at some random times, which is particularly useful for building statistical procedures. Convincing results are obtained through a simulation study and the use of the model over 10 representative stocks.

One can also see the parameter η as a measure of the relevance of the tick size on the market. Indeed, if $\eta < 1/2$, market participants are convinced they have to trade at a new price before the efficient price crosses this new price on the tick grid. So, it means that the tick size appears too large to them. Conversely, a large η ($\eta > 1/2$) means that the tick size appears too small. From the tick size perspective, an ideal market is consequently a market where η is equal to $1/2$.

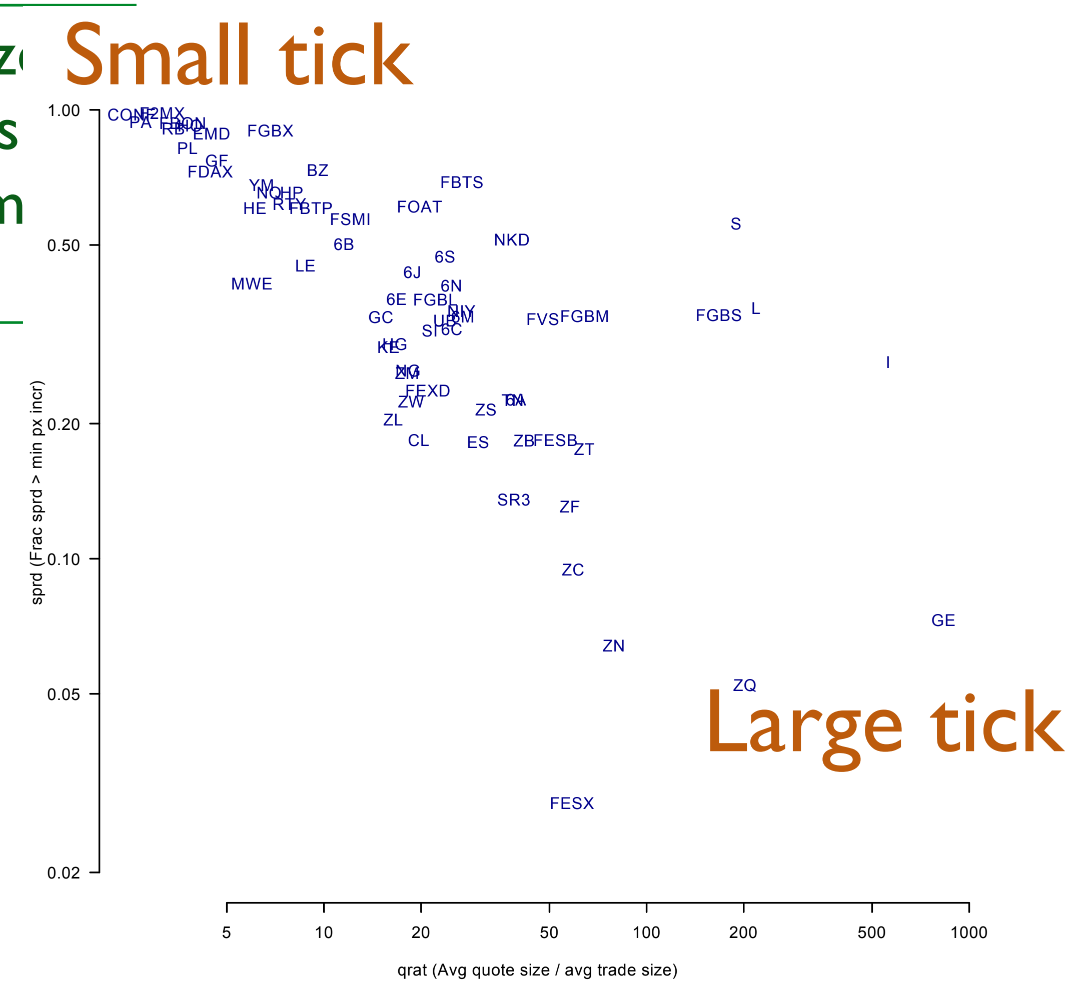
A natural estimation procedure for the parameter η is given in Robert and Rosenbaum (2010a). We define an alternation (resp. continuation) of one tick as a price jump of one tick whose direction is opposite to (resp. the same as) the one of the preceding price jump. Let $N_{\alpha,t}^{(a)}$ and $N_{\alpha,t}^{(c)}$ be respectively the number of alternations and continuations of one tick over the period $[0,t]$. An estimator of η over $[0,t]$ is given by

$$\hat{\eta}_{\alpha,t} = \frac{N_{\alpha,t}^{(c)}}{2N_{\alpha,t}^{(a)}}.$$

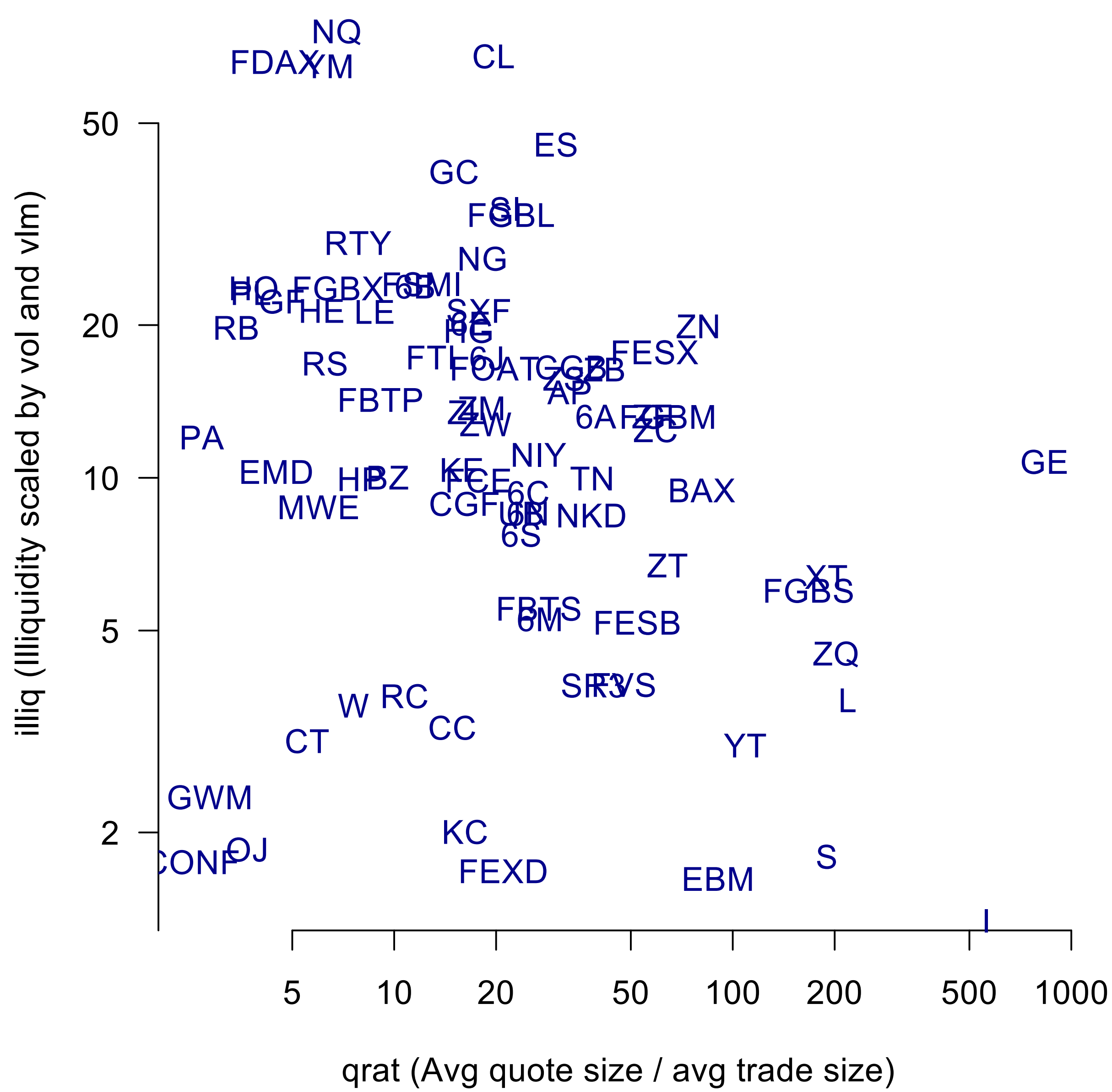
Tick size spectrum



3 measures of tick siz
- quote size / trade s
- Robert/Rosenbaum
- fraction sprd>tick

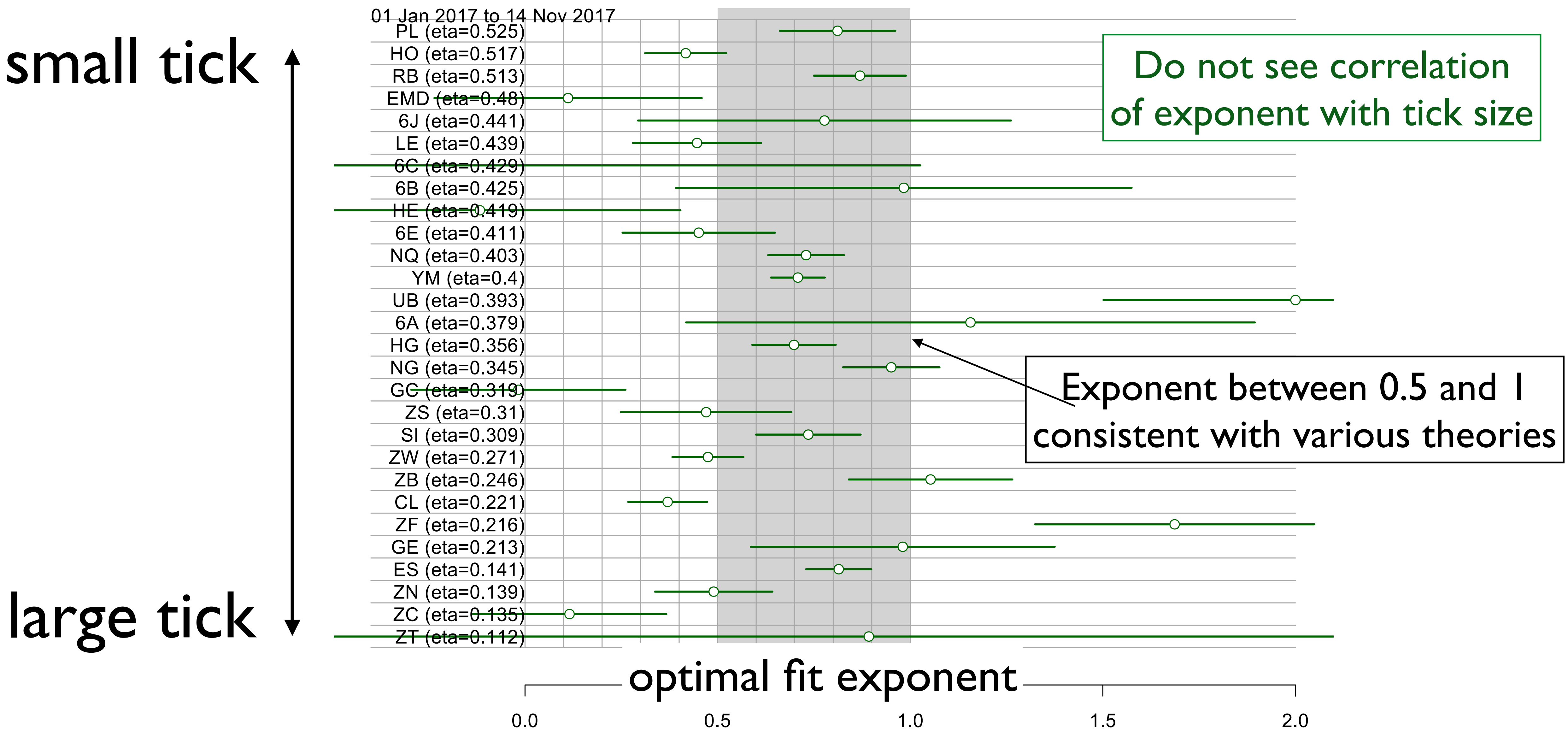


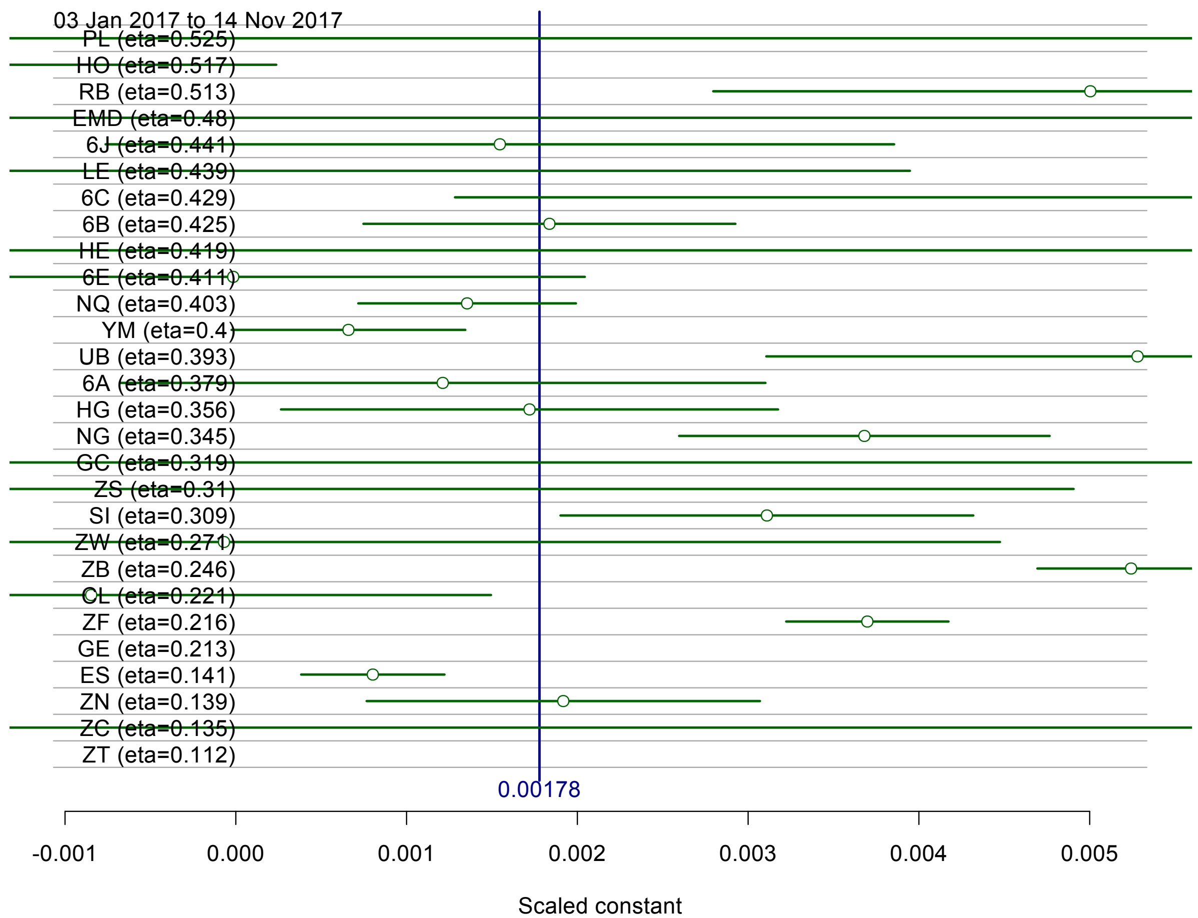
Tick size vs nondimensional liquidity

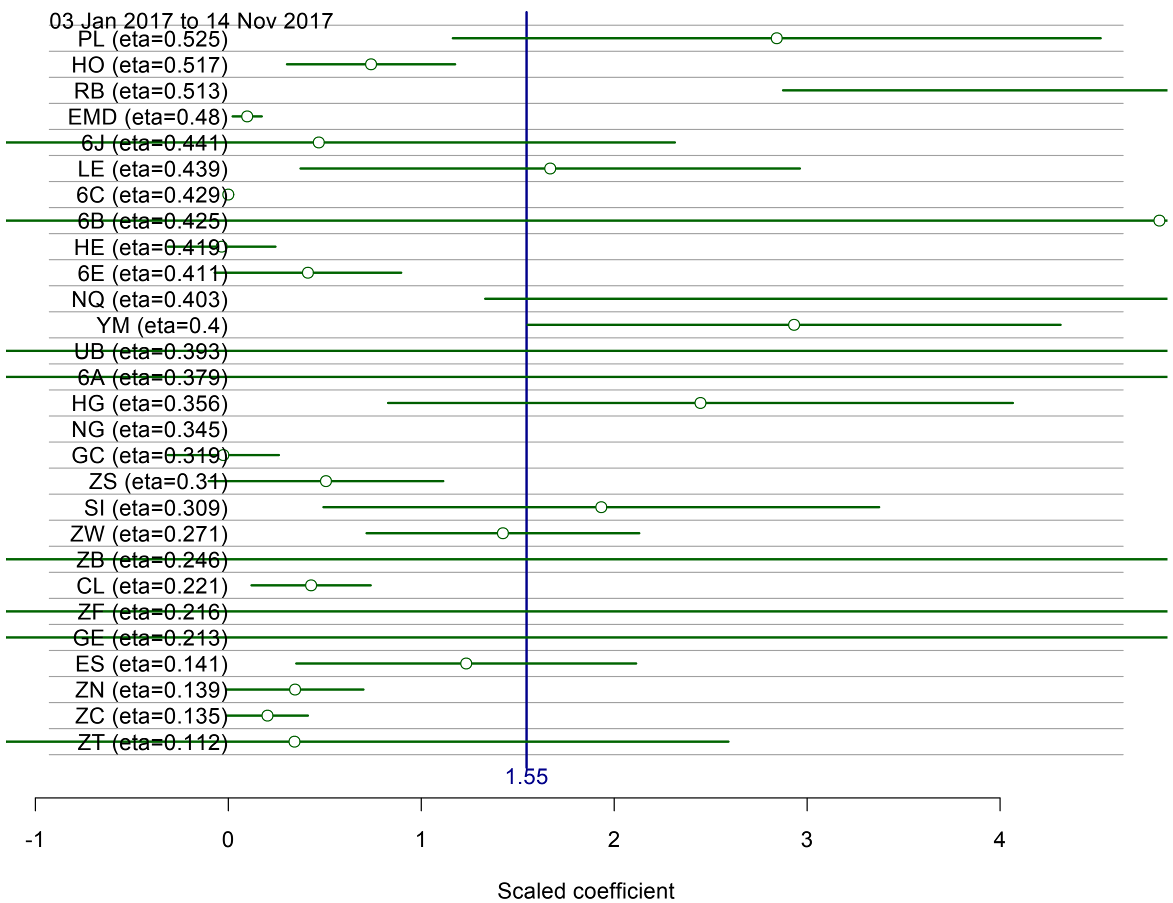


Subdivide based on these parameters

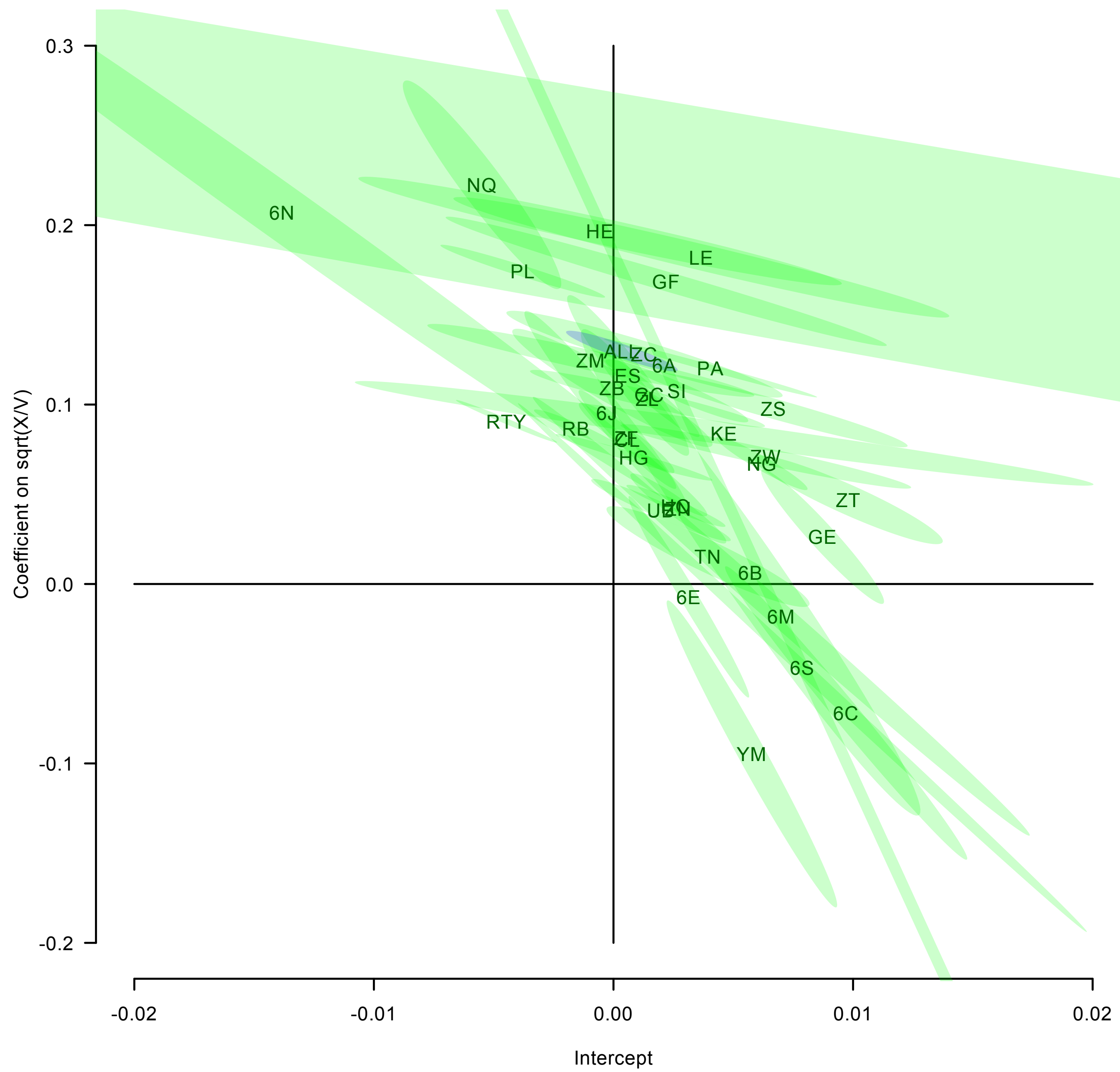
- Does not work (does not give meaningful results)
- because points that are close in parameters
  are not close in cost models
-
  Problem: market impact model depends on
  properties
  that are not part of market data
  for example, size of underlying asset.

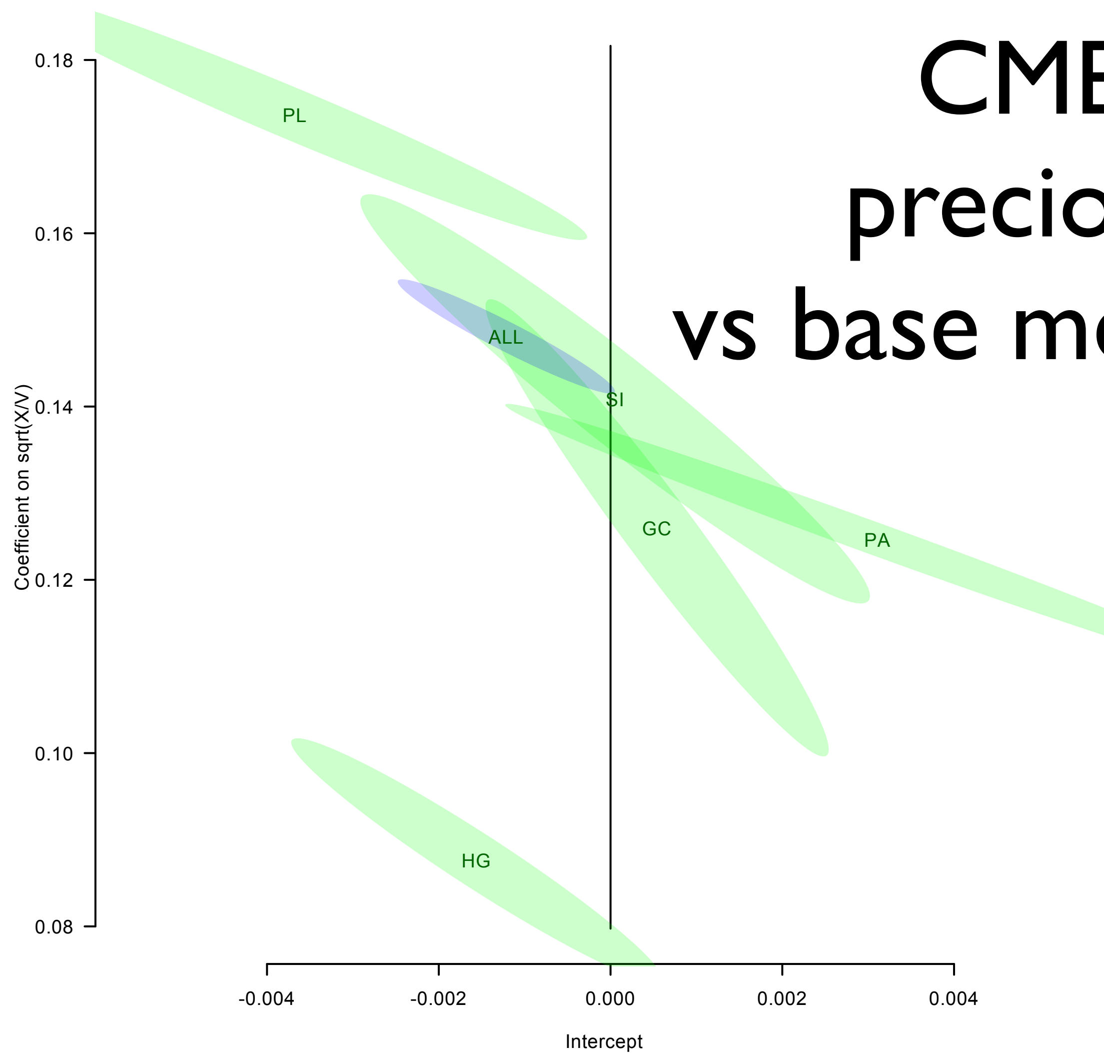

small tick
large tick
01 Jan 2017 to 14 Nov 2017
PL (eta=0.525)
HO (eta=0.517)
RB (eta=0.513)
EMD (eta=0.48)
6J (eta=0.441)
LE (eta=0.439)
6C (eta=0.429)
6B (eta=0.425)
HE (eta=0.419)
6E (eta=0.411)
NQ (eta=0.403)
YM (eta=0.4)
UB (eta=0.393)
6A (eta=0.379)
HG (eta=0.356)
NG (eta=0.345)
GC (eta=0.319)
ZS (eta=0.31)
SI (eta=0.309)
ZW (eta=0.271)
ZB (eta=0.246)
CL (eta=0.221)
ZF (eta=0.216)
GE (eta=0.213)
ES (eta=0.141)
ZN (eta=0.139)
ZC (eta=0.135)
ZT (eta=0.112)
optimal fit exponent
0.0
0.5
1.0
1.5
2.0
Do not see correlation of exponent with tick size
Exponent between 0.5 and 1 consistent with various theories



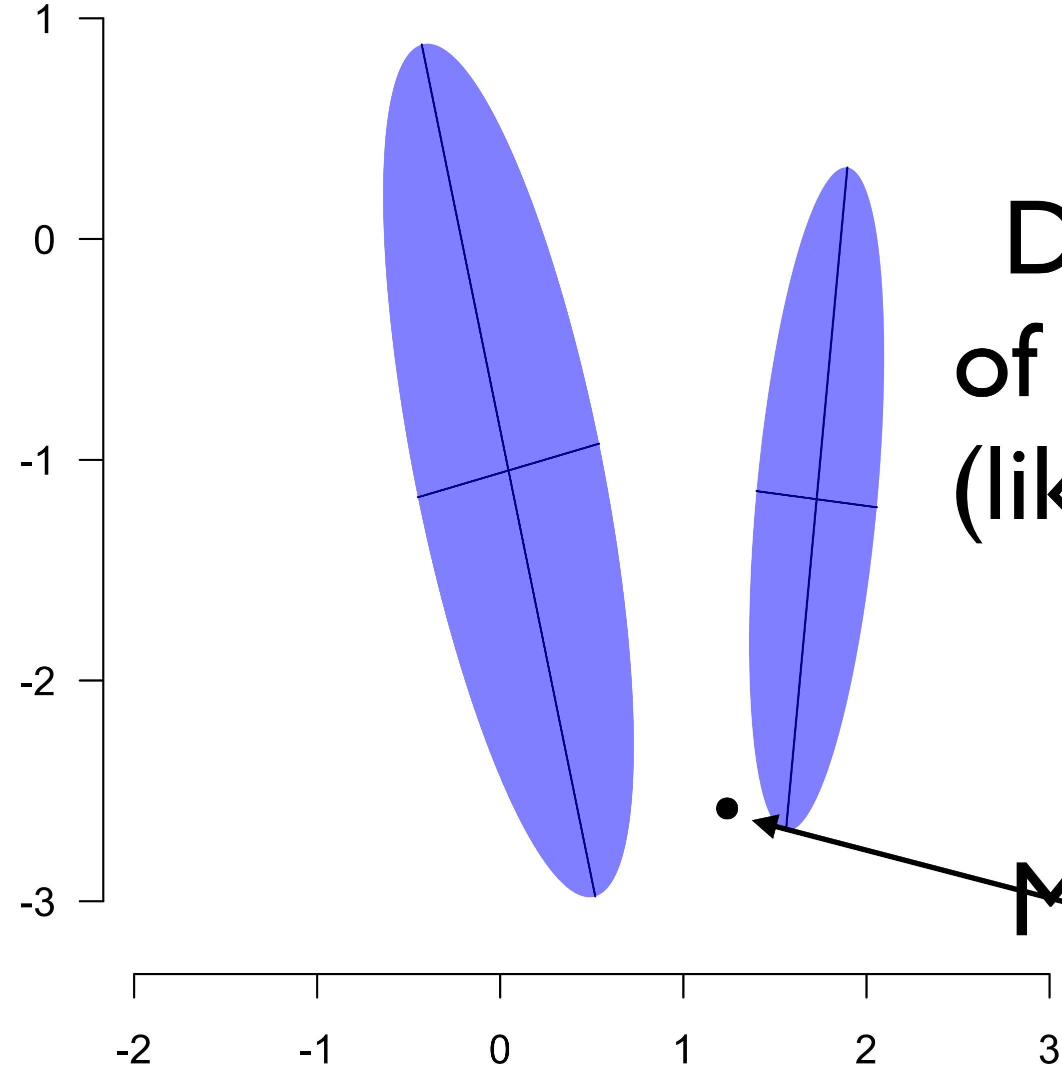


Coefficients do not depend on tick size in consistent way


0.3
0.2
0.1
0.0
-0.1
-0.2
Coefficient on sqrt(X/V)
-0.02
-0.01
0.00
0.01
0.02
Intercept
NQ
6N
HE
LE
PL
GF
ZM
ES
6A
PA
ZB
SI
ZS
6J
ZL
RTY
RB
KE
CL
HG
NG
ZT
GE
TN
6B
6E
6M
6S
6C
YM



CME metals: precious metals vs base metals (copper)



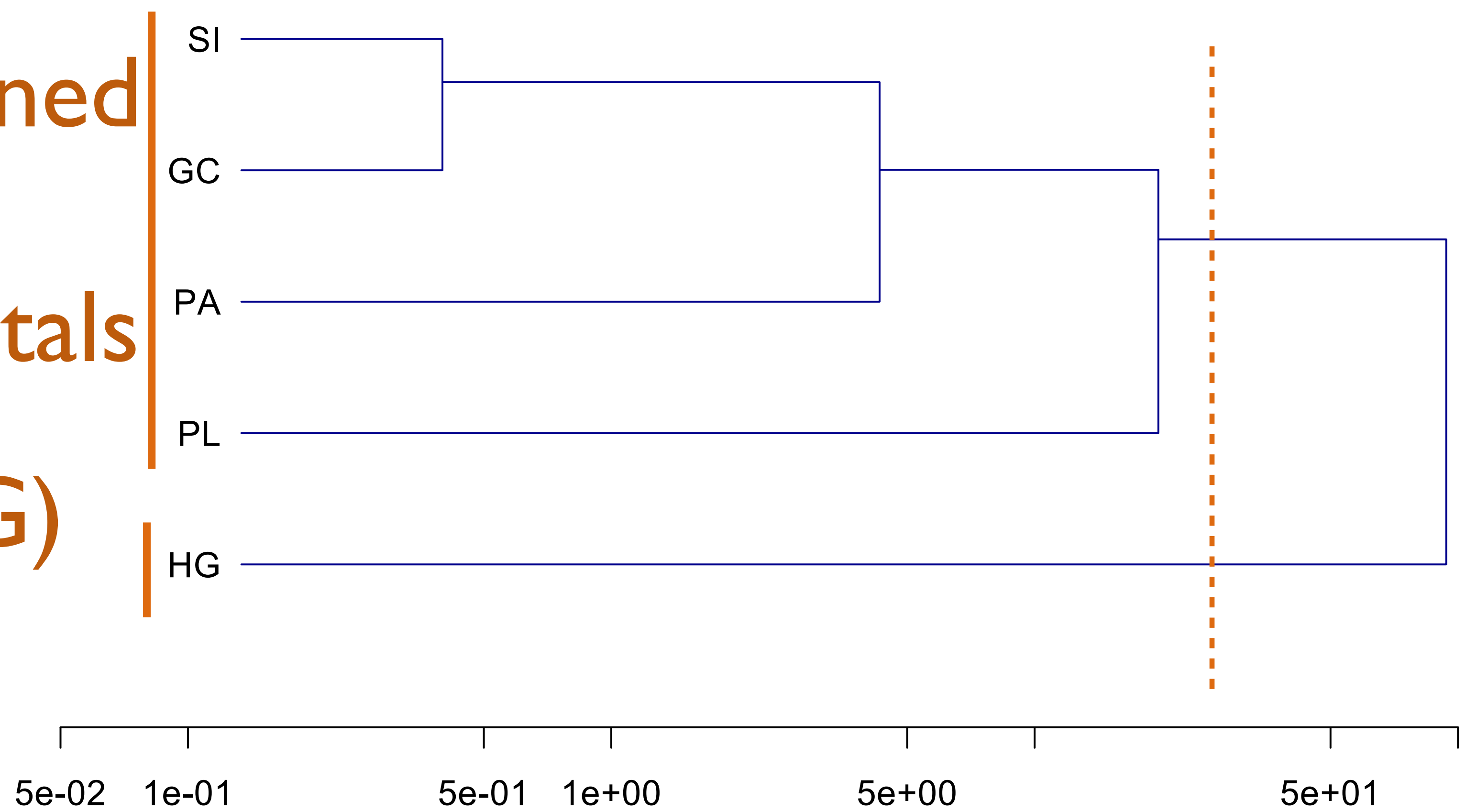
Distance = -log prob of most probable point (like a 2-variable t-test)

Most probable point

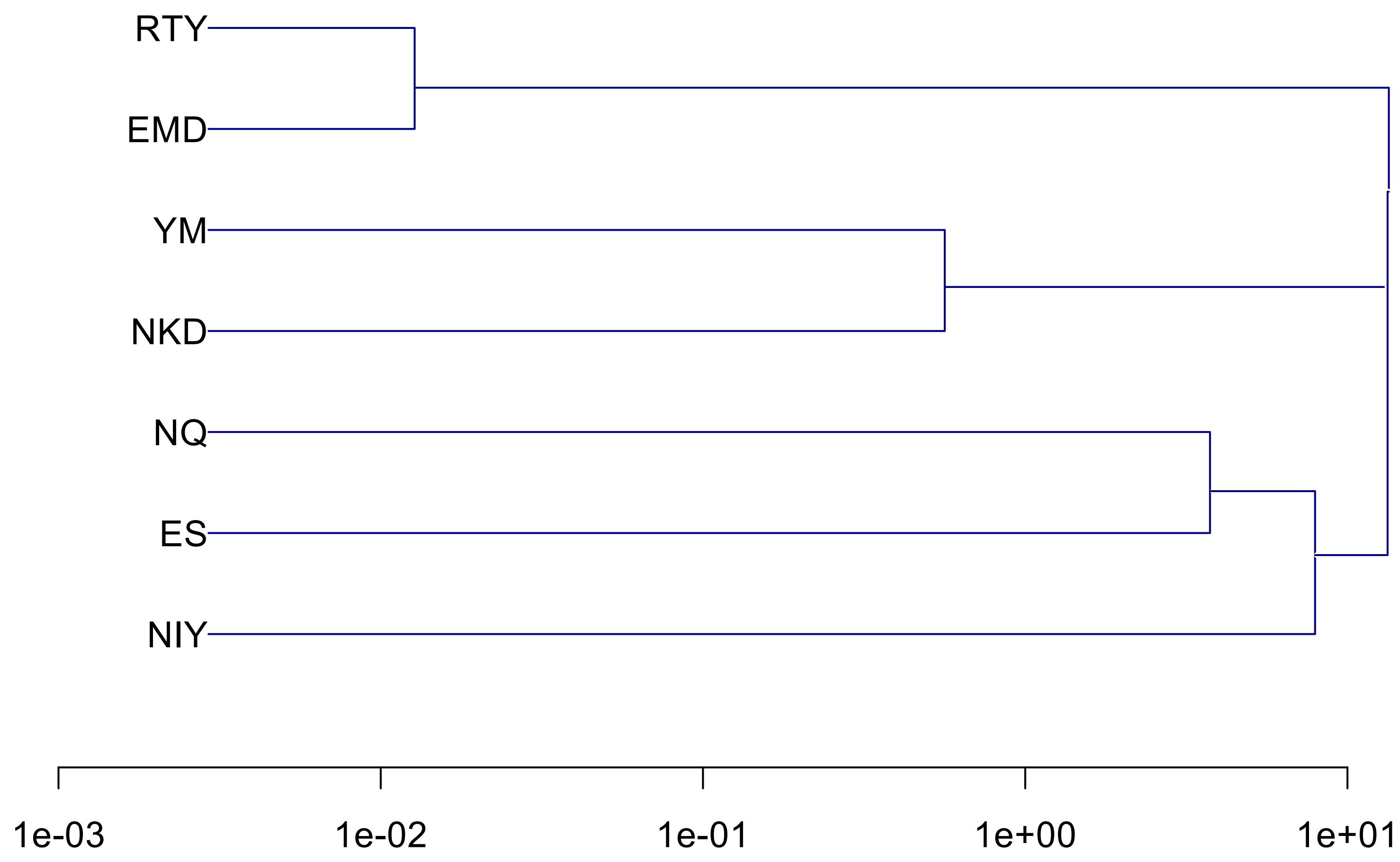
CME metals

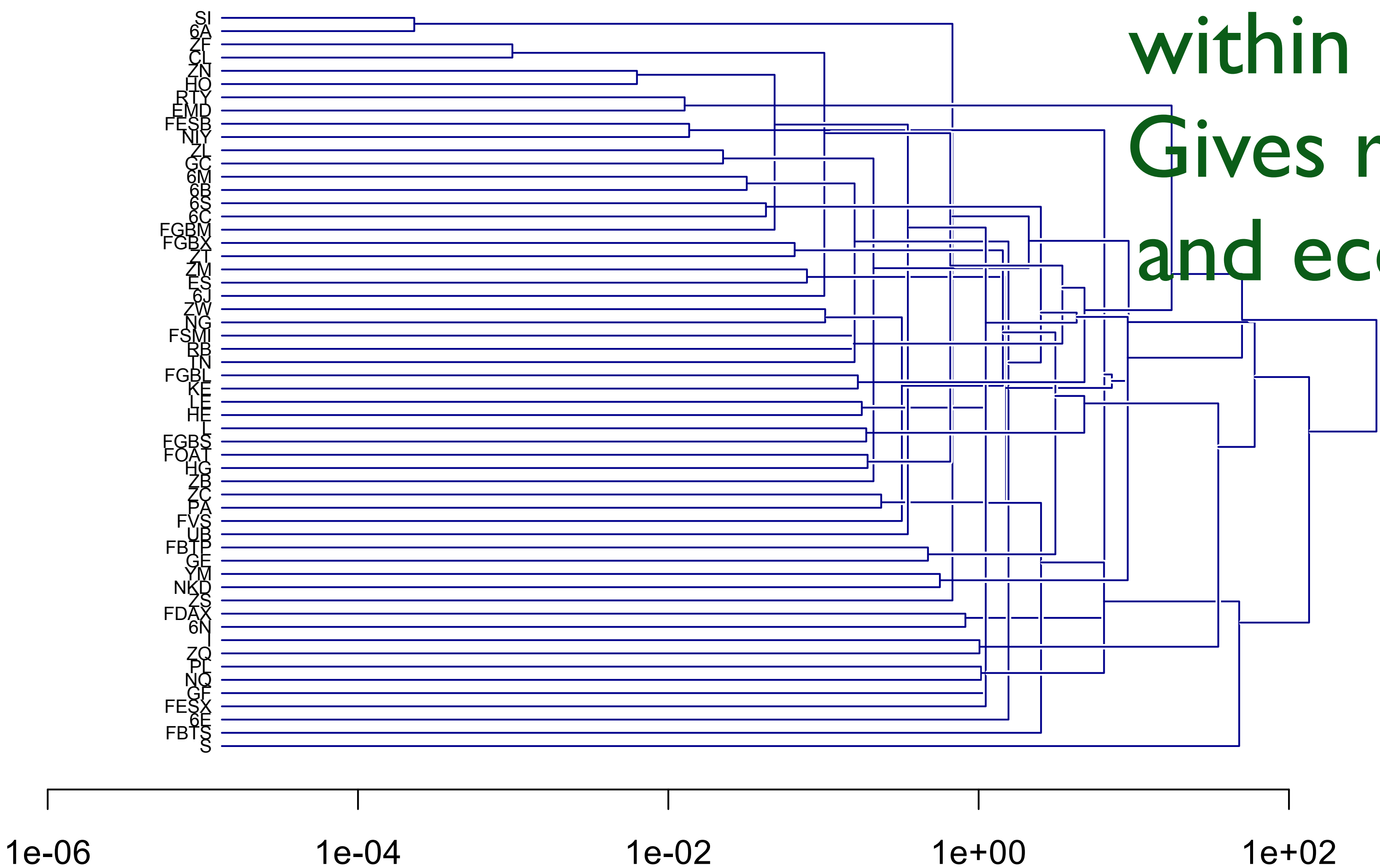
Make combined fit for precious metals

Copper (HG) on its own



CME equity index futures





Conclusion: cluster within exchange and class. Gives reasonable accuracy and economically sensible

Conclusions

- Market impact modeling is noisy
  - R^2 terrible, t-stats good
  - ability to predict any particular trade is poor
- Need to use physical reasoning and ad hoc decisions
  - focus on parameter ranges that are economically important
- Results should be interpreted intelligently