Price Signals in Trade Execution

Robert Almgren

SwissQuote
Nov 2019

Trade execution

- Execute order as agent for institutional client
  - QB = futures and interest rate markets
- Goal: "best" final average execution price
- Evaluate relative to benchmark
  - benchmark defines an "ideal" trade
  - different benchmarks give different strategies

Slippage

Difference of final average execution price and benchmark

- execution - benchmark for buys
- benchmark - execution for sells

Positive slippage is bad, negative is good

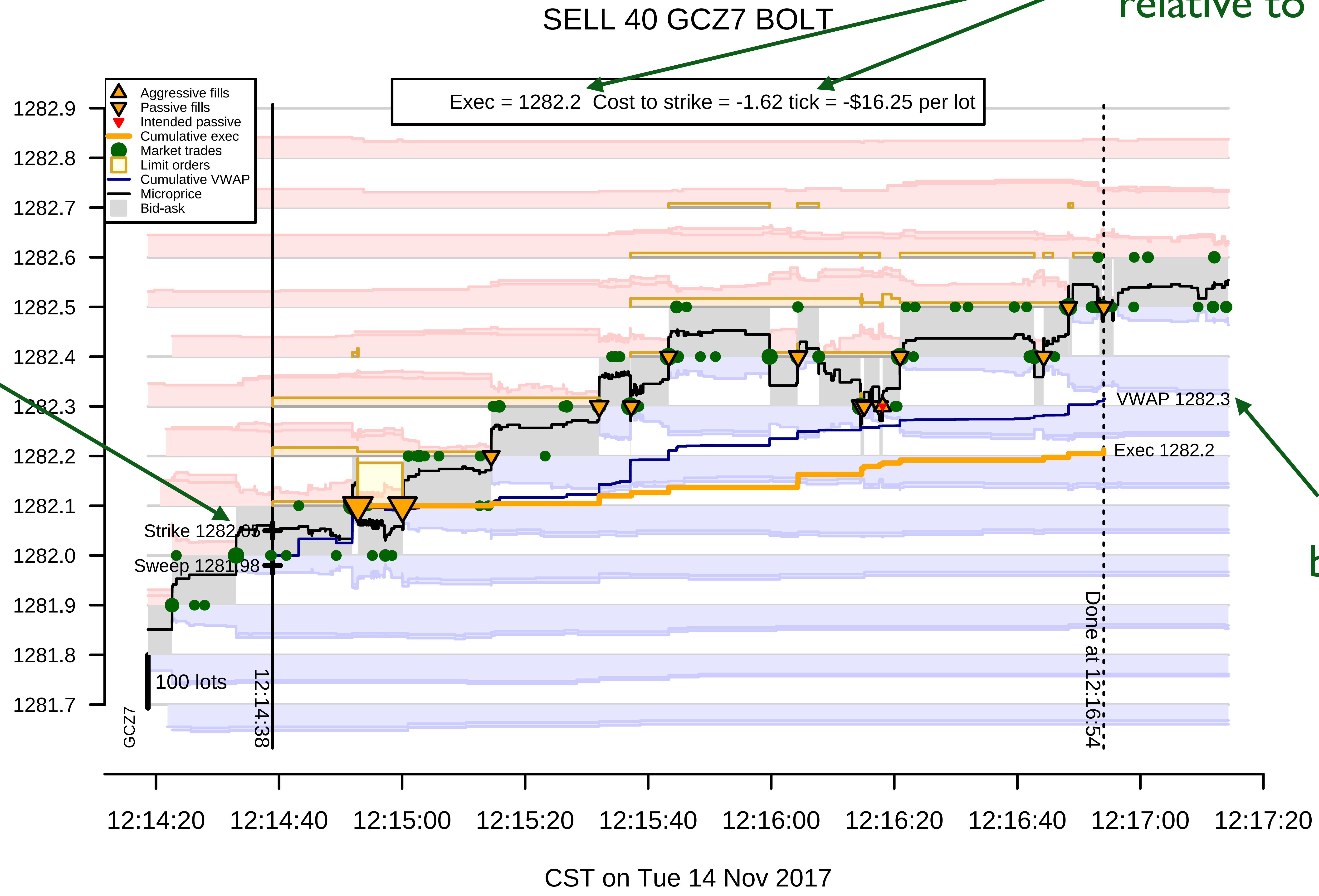
For agency execution, minimize this

Different benchmarks and algorithms

Bolt:	arrival price
Strobe:	average price on interval (TWAP or VWAP)
Closer:	settlement price
Legger:	multi-leg target price
Roll:	multi-day roll benchmark (in progress)

Bolt: arrival price

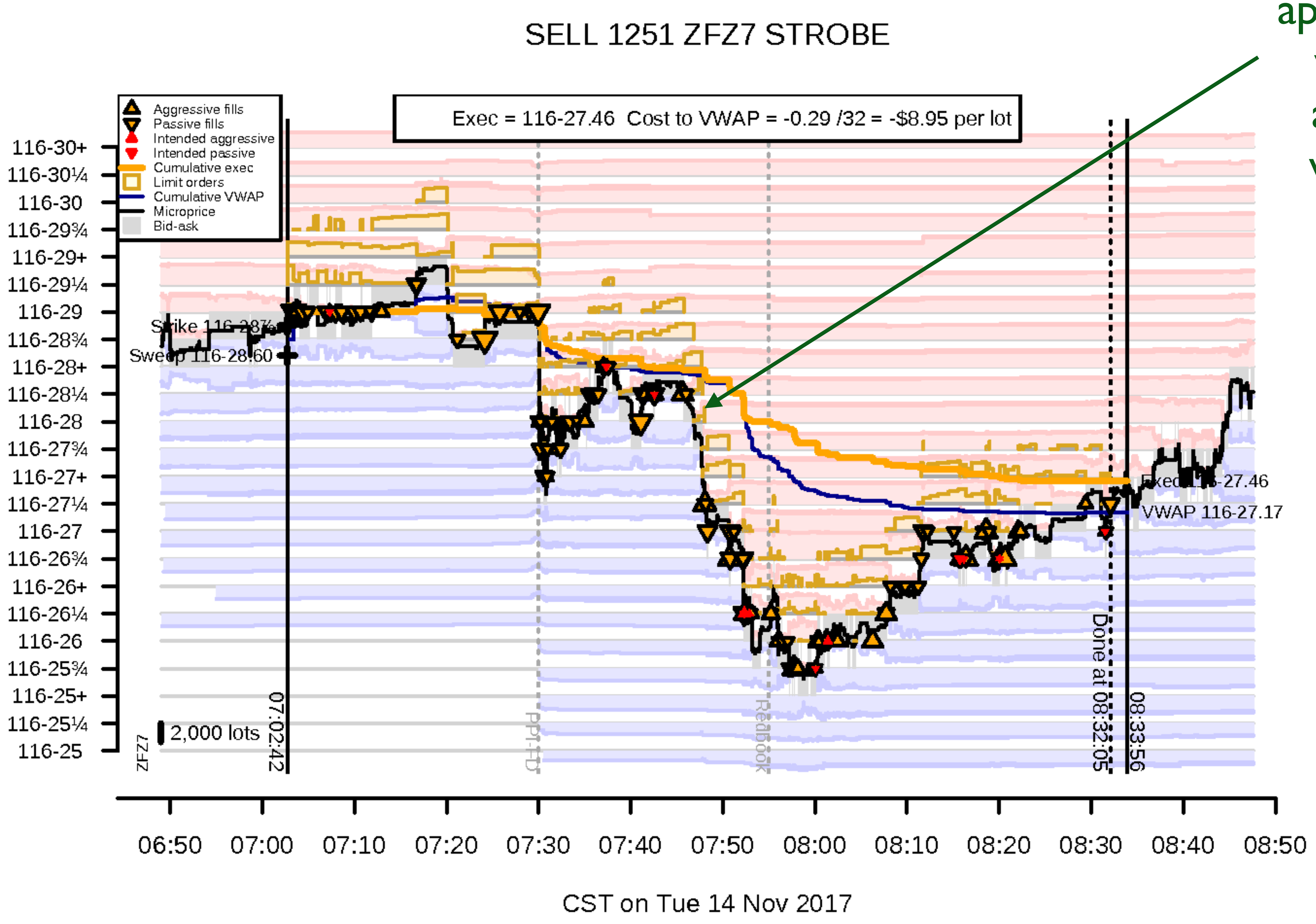
Report execution price and slippage relative to benchmark



Arrival price benchmark ("strike")

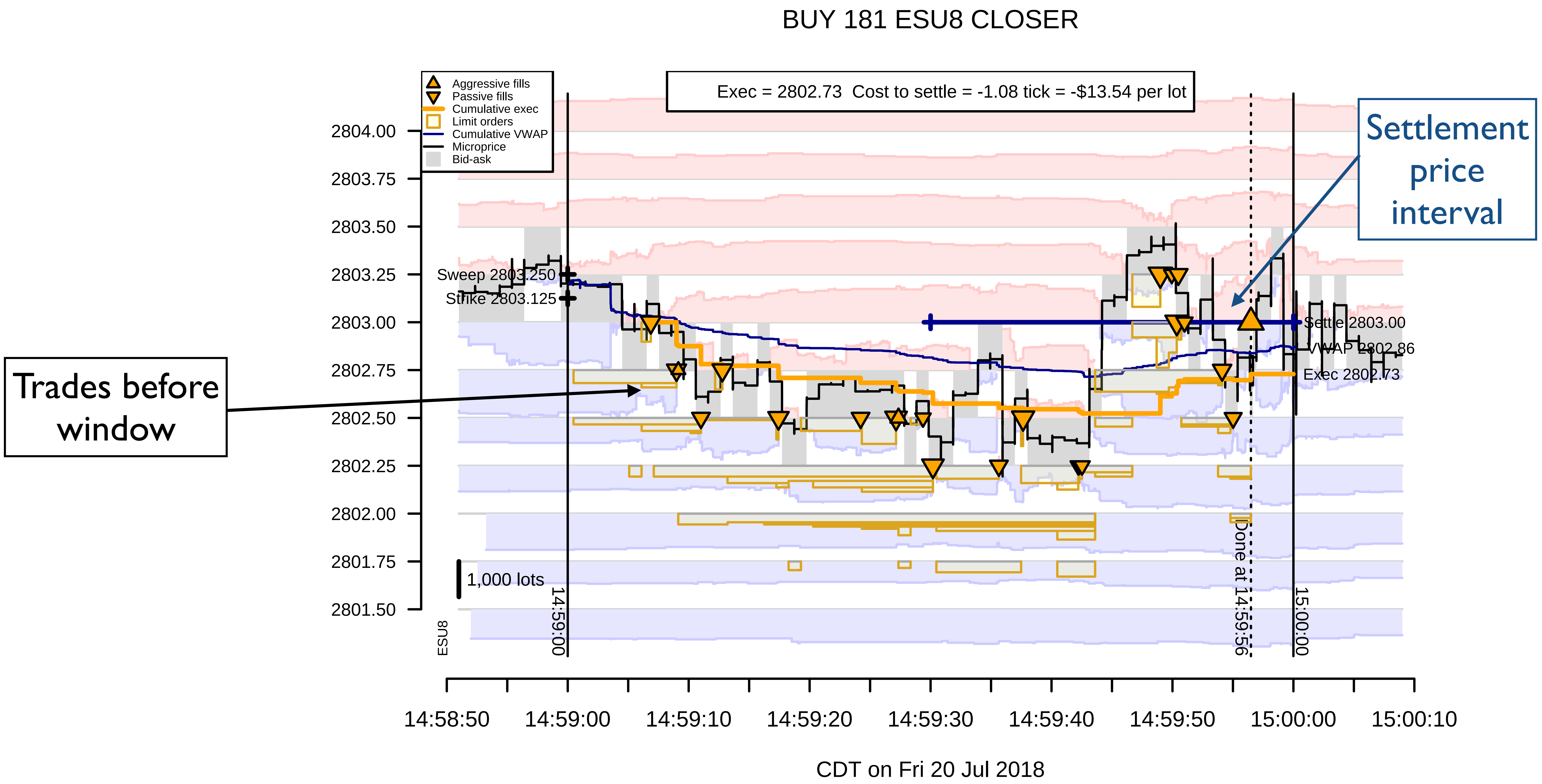
Also report other benchmarks for interest (but these are not targeted by this algo)

Strobe: average price on interval

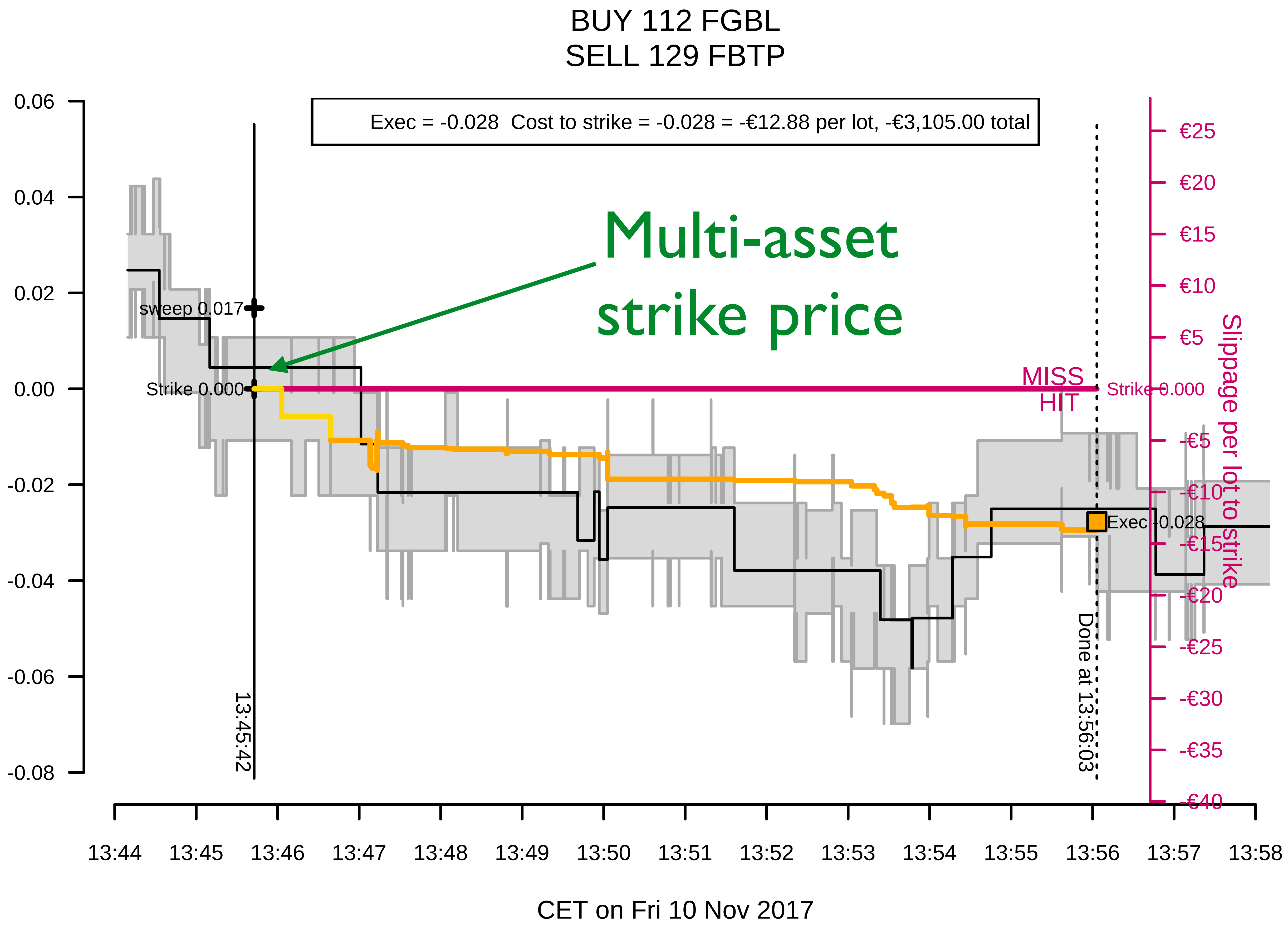


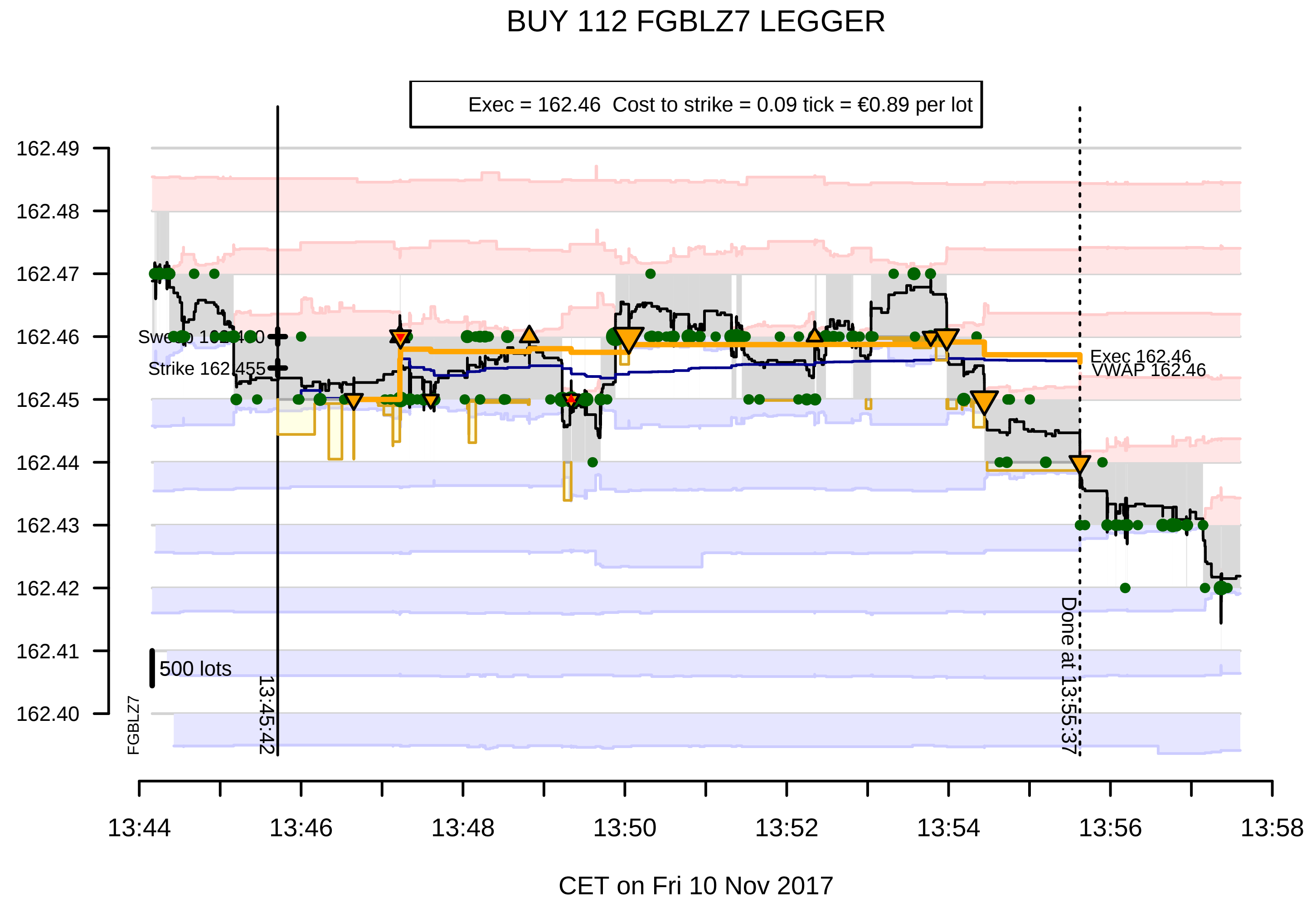
For Strobe, execution approximately follows volume curve, but also opportunistic when can improve performance

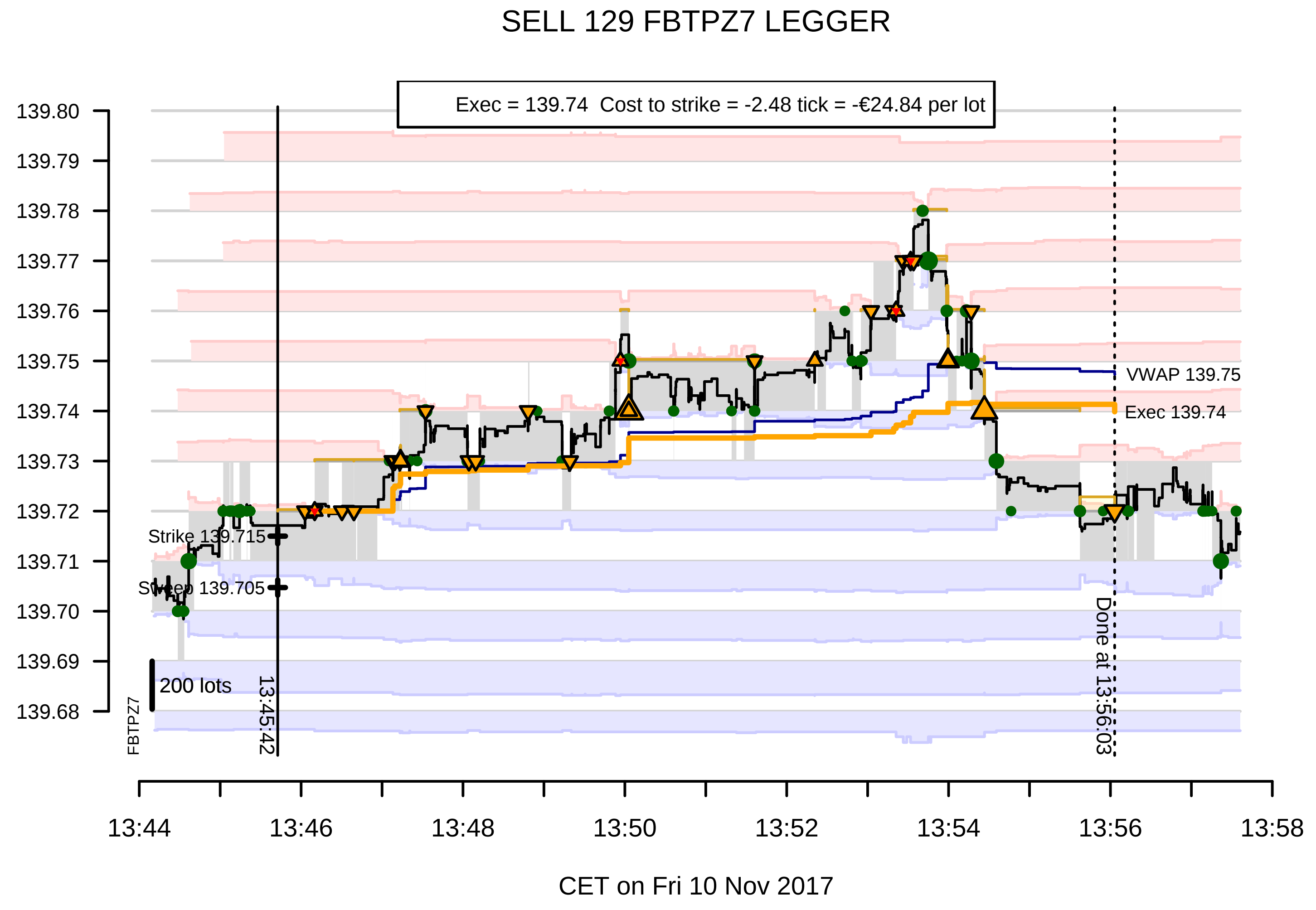
Settlement price algorithm



Legger: multi-asset strike price







Business drivers

Good average execution price relative to benchmark

- Also manage risk relative to benchmark

Reliable systems and broad global coverage

- large investments in data and technology, and support

Transparent processes and algorithms

- Must be able to explain to clients
- Pictures are very helpful

Correlation and regression

Nick Patterson

[30:06] "...I joined a hedge fund, Renaissance Technologies. ... our most important statistical tool was simple regression with one target and one independent variable. ... nobody tells you what the variables you should be regressing [are]. What's the target? Should you do a nonlinear transform before you regress? What's the source? Should you clean your data? Do you notice when your results are obviously rubbish?"

http://www.thetalkingmachines.com/episodes/ai-safety-and-legacy-bletchley-park

Outline

- What is performance? Best execution
- How do we achieve performance? Signals and infrastructure
- Signal framework and signals
- Three particular topics in semi-detail
  - Smart order router using machine learning
  - Y-means consensus framework
  - Treasury roll forecasting

What matters for performance

Passive fills

- many futures products are large-tick

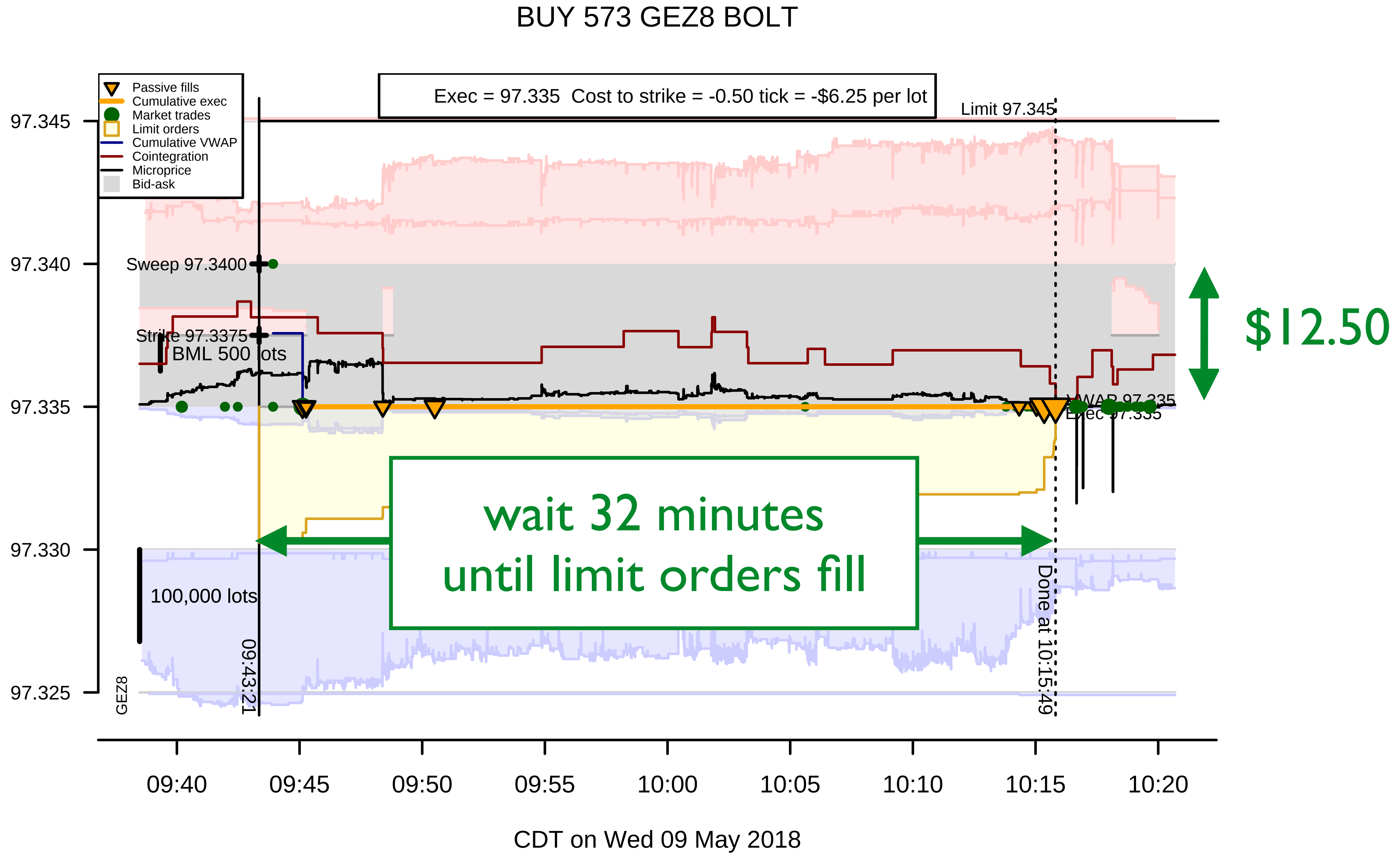
Short-term price prediction

- aggress or pull back based on price forecast

Use simulator to evaluate algorithm improvements

- simulator uses real data to capture fills and signals

Determinants of slippage



Passive fills

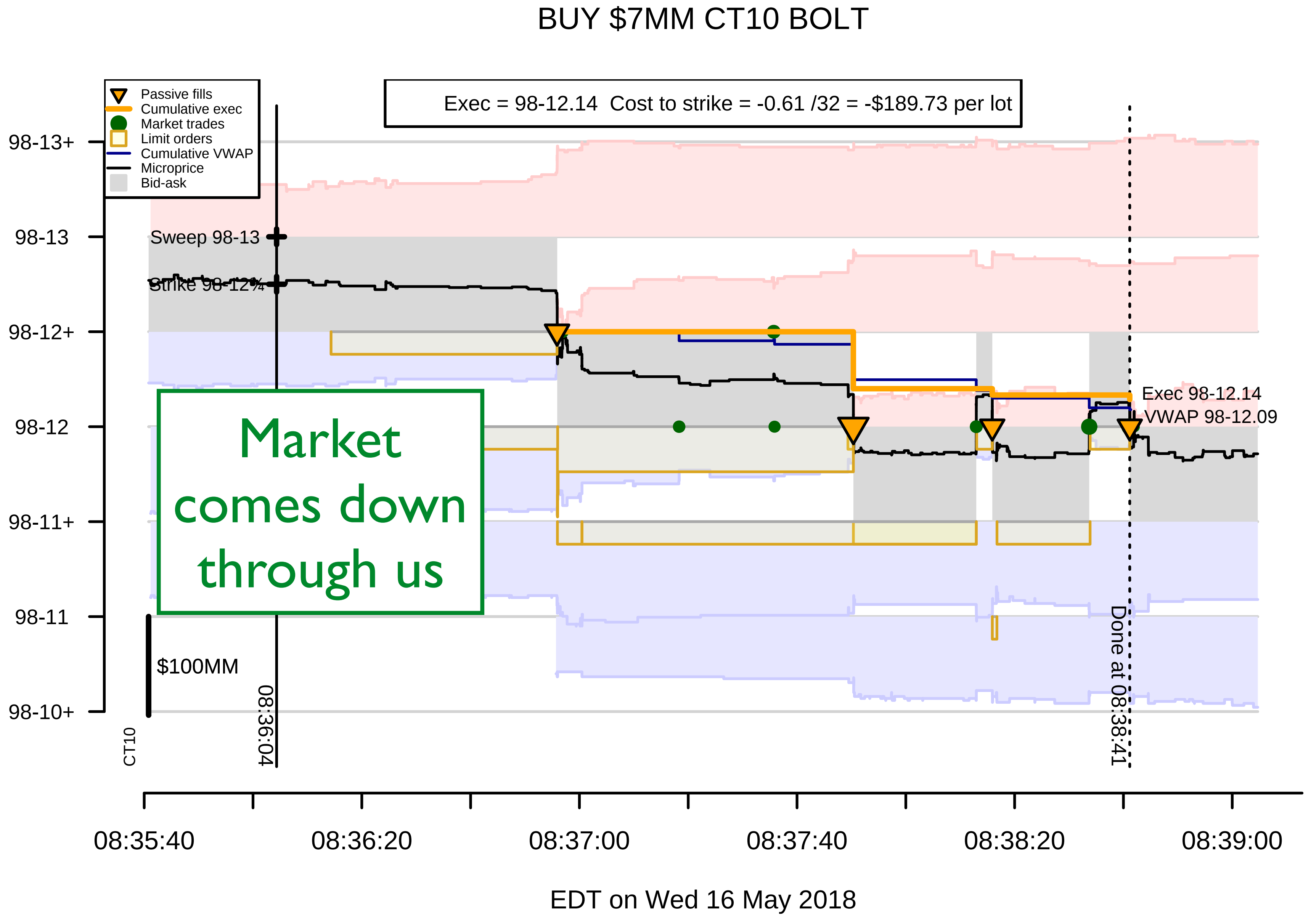
buy at bid, sell at ask

be patient, unless price will move away

*** Short term pricing signals

price will go up or down?

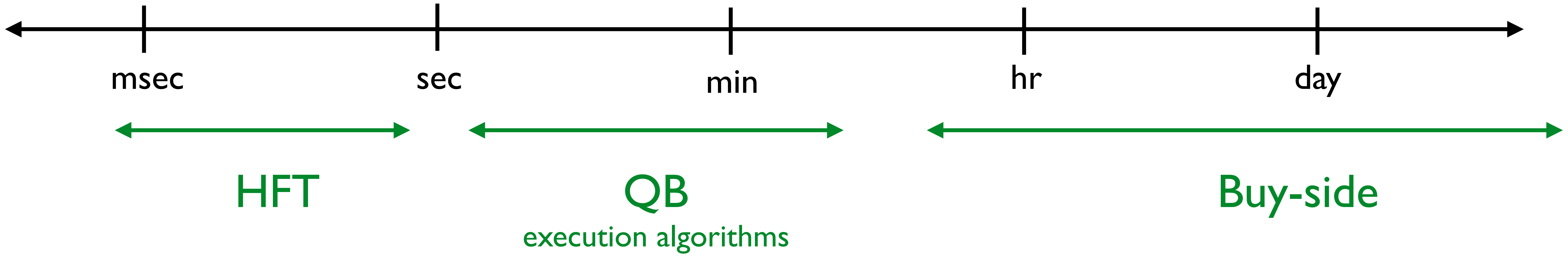
pick when to execute



What is a signal?

- Signal = short-term price forecast
  - Computed from past market data
  - Forecast on time horizons seconds to minutes
  - Use them conditional on market state variables
- Signals are independent of order being executed
  - objective statement of market properties
- Biggest ingredient in execution performance
  - Speed up or slow down depending on direction

Time frames of signals



Bar is lower for execution signals than for alpha trading

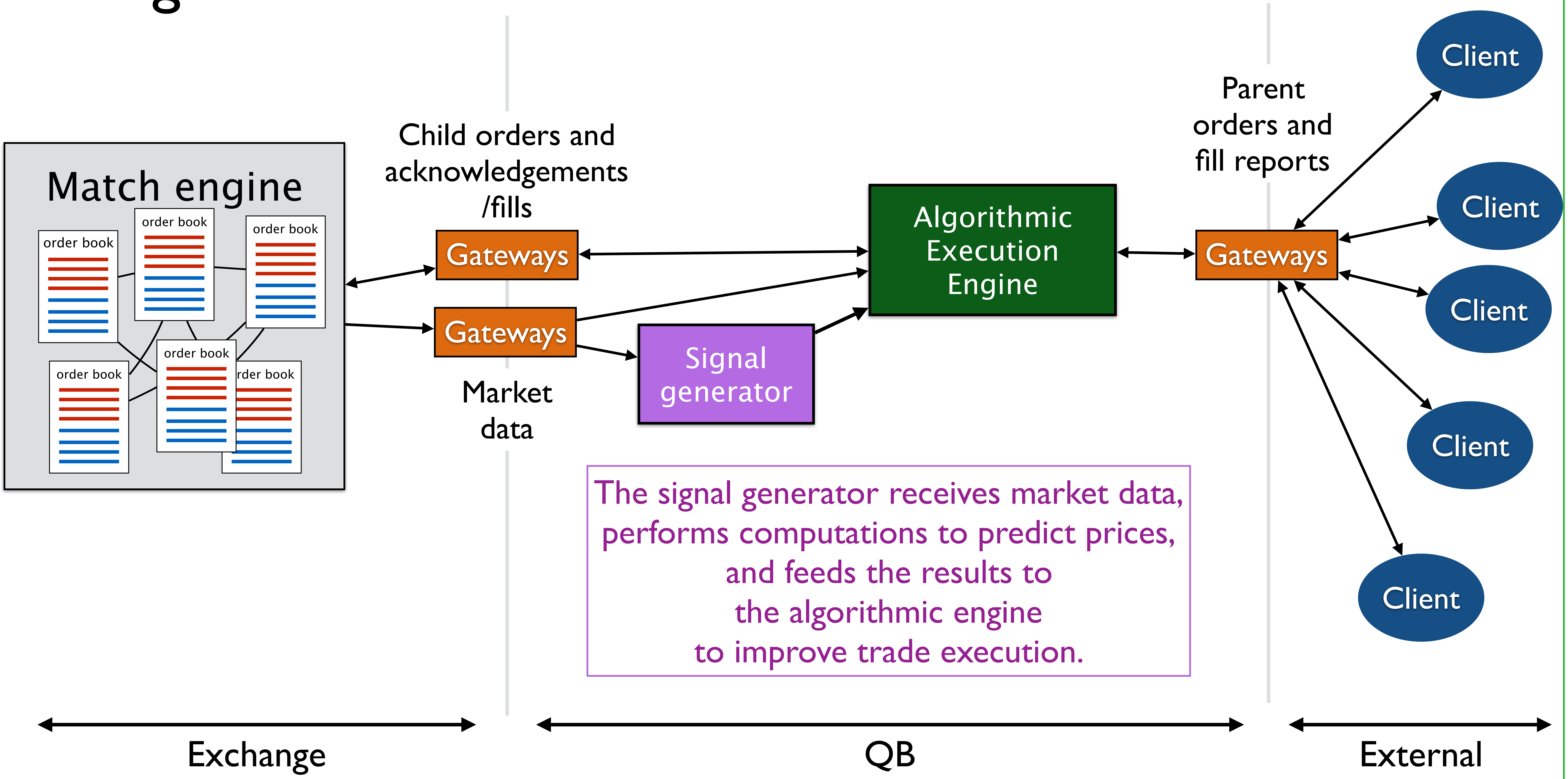
- not competing with HF firms
- no round-trip trading, so small signals add value

How do we compute signals?

- Computed in real time from streaming market data
- Latency is important

  not to get signals extremely rapidly
  but to not fall behind
- May be complex calculations
- Rest on simple ingredients
- Need flexible platform to develop new signals

Trading architecture



The signal generator receives market data, performs computations to predict prices, and feeds the results to the algorithmic engine to improve trade execution.

What does not work?

Master Thesis - Luca Rona
***S&P500* Short-Term Price Prediction using Machine Learning**

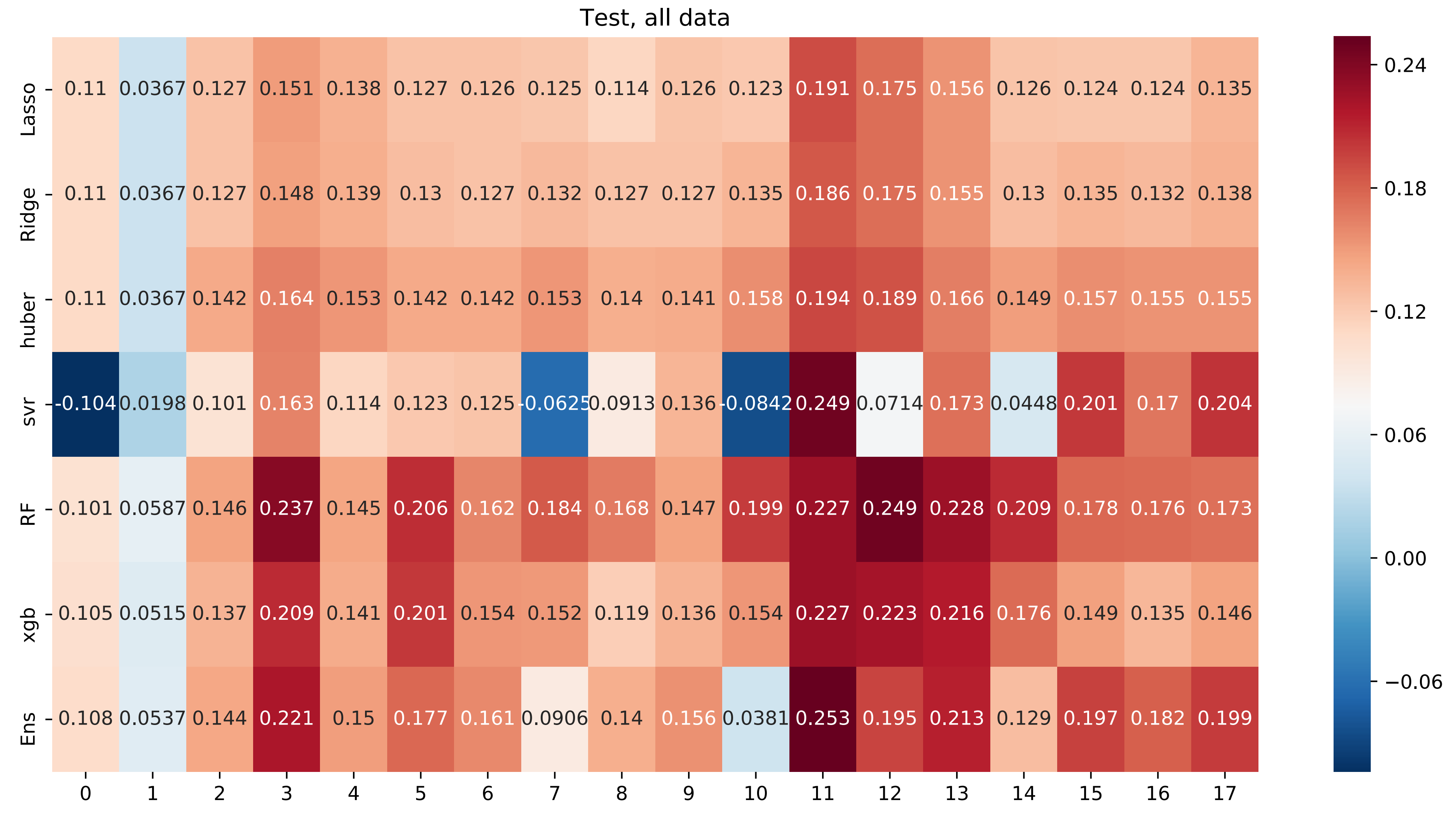
Spring 2018

Luca Rona
Master in Finance
Princeton University
lrona@princeton.edu

In this paper we investigate whether S&P500 mean reverts after sharp moves over different time horizons ranging from 10 seconds to 5 minutes. After verifying that statistically significant mean-reversion properties which are too small for active trading exist, we find that that Machine Learning methods obtain increased forecasting power over forward returns when combined with a rich enough feature set. We notice that including too many variables results in sub-optimal models and that a *forward* variable selection method works better than *backward*. Linear Methods with Shrinkage provide good baseline, but have overall lower accuracy than SVR, Random Forest and Gradient Boosting in the testing set. Ensembling predictions from different models makes the model more stable, but does not provide substantial accuracy gains. A simple trading strategy based on the predictions is developed and proves profitable in the testing set. However, we are cautious about these findings as they are not statistically significant and based on a test-set that is not large enough to be representative of different trading regimes.

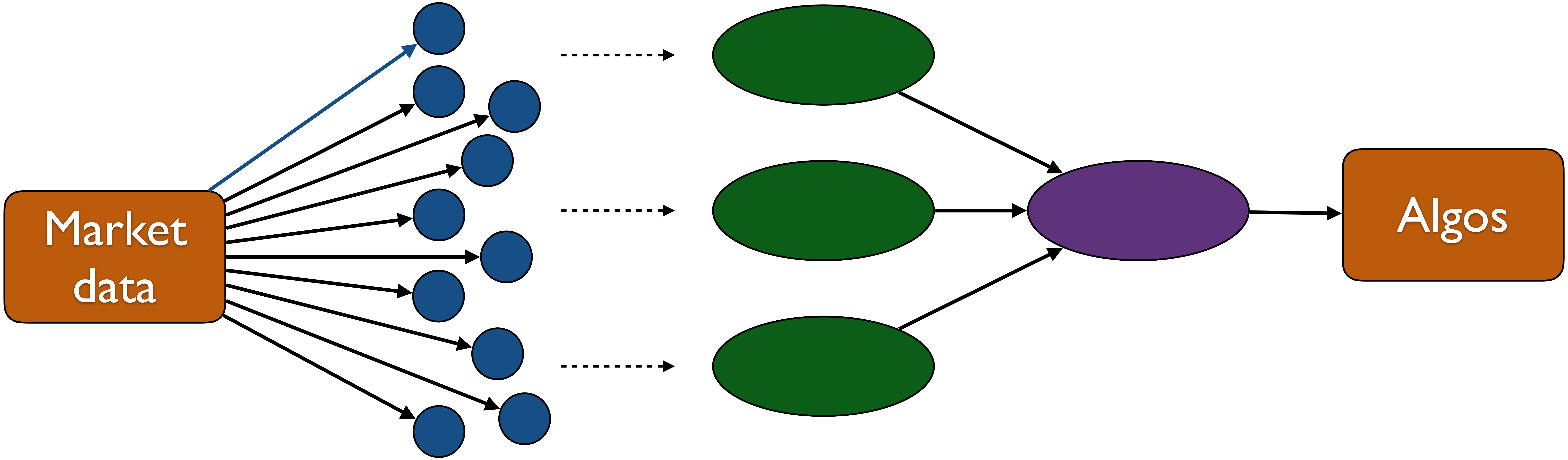
At the end of each bin we record the following quantities:

- **TimeStamp**: date and time
- **Bid/Ask**: bid and ask price
- **VWAP**: volume weighted average price in the previous bin
- **Volume**: quantity of asset traded in the previous bin
- **Number of trades**: number of separate trades in the previous bin
- **Volume buy/sell**: volume of trades marked as buy and sell respectively by the exchange
- **Number of buy/sell**: number of separate buy/sell trades in the previous bin



Data does not automatically tell you: need to construct signals using reasoning.

Signal architecture



"Features"
small, quick and widely useful

"Signals"
complex calculations

"Consensus"
combination of signals

Implemented in Kdb+

Features

- Features are simple computations of market data that are useful to a variety of signals
- Are computed synchronously--must be fast
- Examples:

  Average quote size
  Traded volume
  Volatility
  Average price

Signals

Trade-at-Settlement (Kenan Si)

useful for Closer (settlement price)

Cointegration for Treasuries (Reza Gholizadeh)

more complex than for short-term rates

Variance Risk Premium (Shankar Narayanan)

compare VIX with realized volatility

Sweep (whole team)

rapid directional motions will revert

Bubble (Shankar Narayanan)

directional motions will persist

Smart Order Routing (Isaac Carruthers)

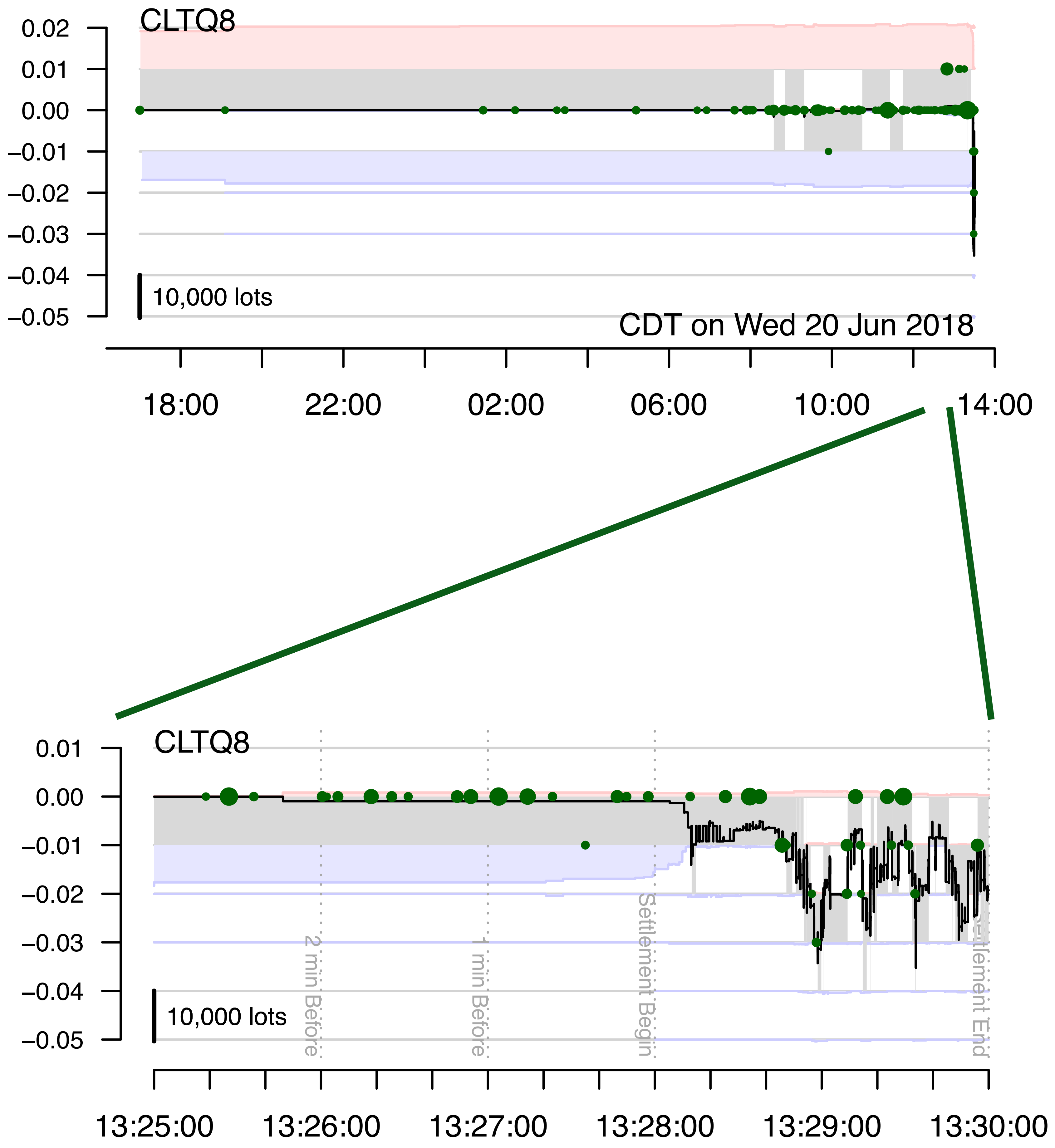
Trade at Settlement

CME Group | Trading | Clearing | Regulation | Data | Technology | Education | About

A flexible and transparent way to manage settlement price uncertainty

Trading at Settlement (TAS) is an order type that allows a market participant to buy or sell futures contracts during the trading day equal to the yet-to-be determined settlement price, or at a price up to four ticks above or below that price.

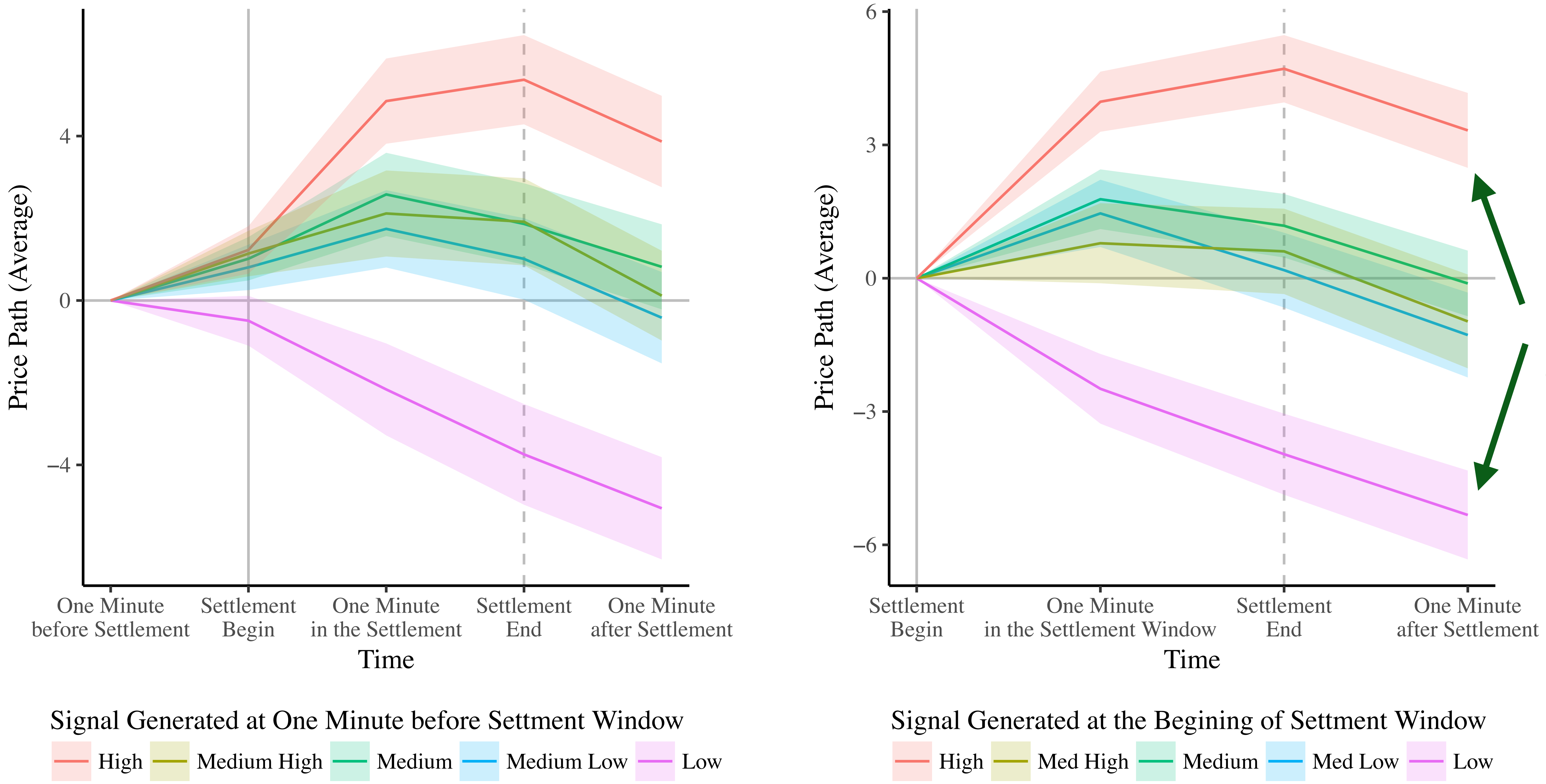
TAS for Crude Oil



- TAS contracts have their own order book
- Trade through whole trading day, though more active before settlement
- Give information about order imbalance, and price direction during settlement (QB Closer algorithm)

Signal validity during settlement window

Mean price trajectory



Extreme values of signal predict price motions during settlement window (correlation is low)

Signal = difference in microprice at two times before settlement

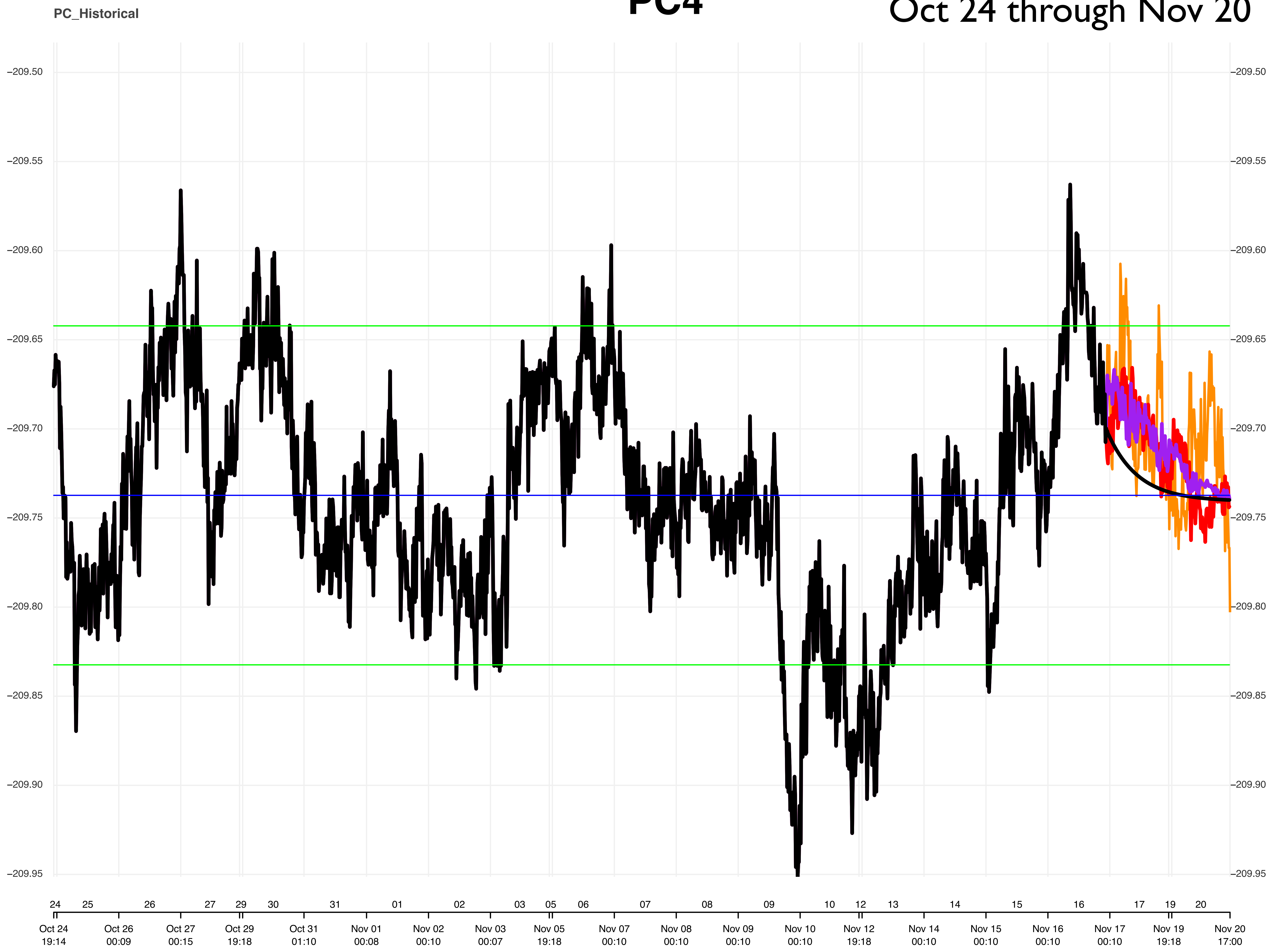
Easy to compute based on preimplemented features

Cointegration for Treasury futures

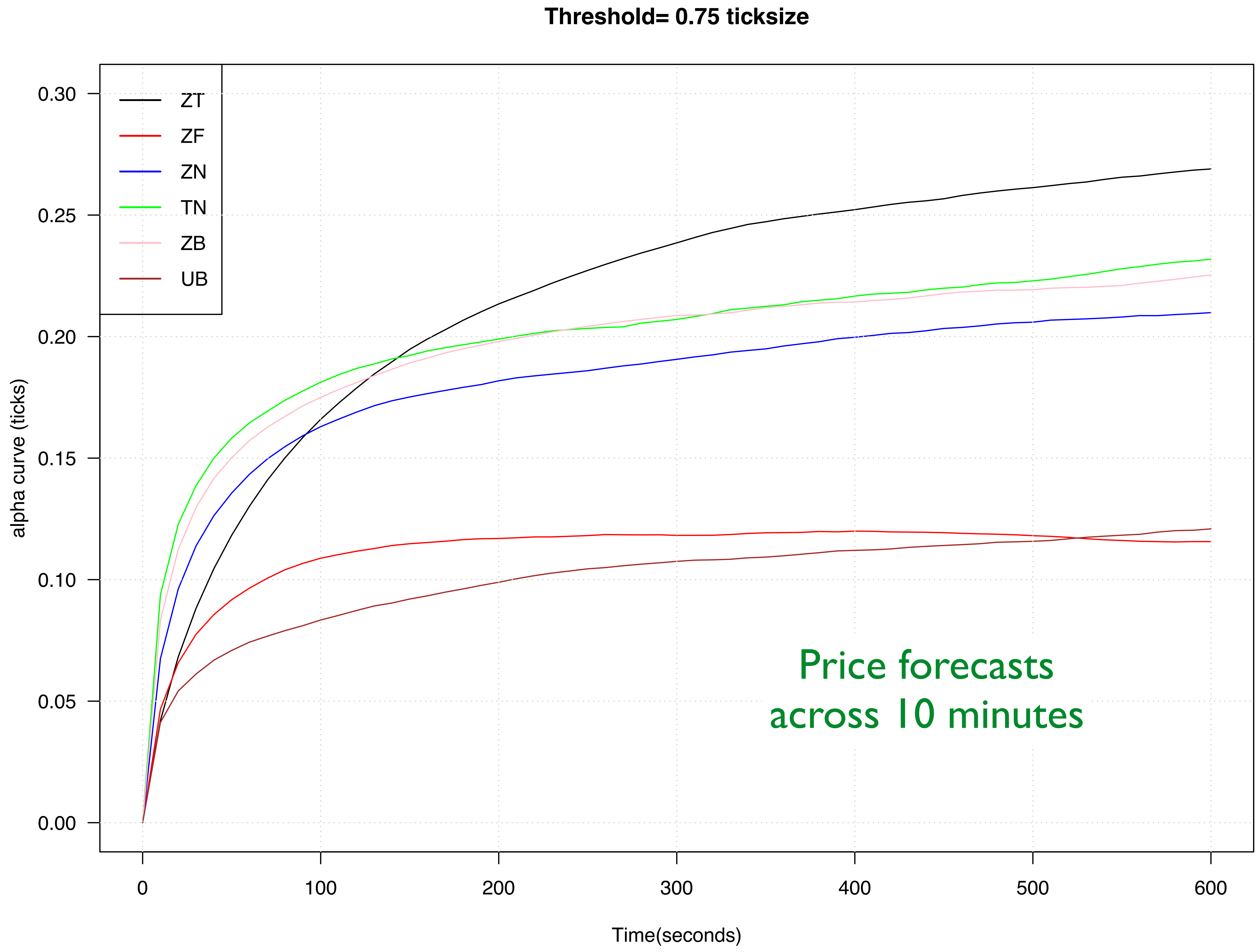
For STIRS, we use an intraday rolling average
For Treasuries, we need a longer-term calculation
Look at 6 Treasury futures across 20 previous days
Store principal components overnight

PC4

Oct 24 through Nov 20



Price forecast for each Treasury futures



Variance Risk Premium

VRP = (Implied vol)^2 - (Realized vol)^2

VRP is forecast of price changes

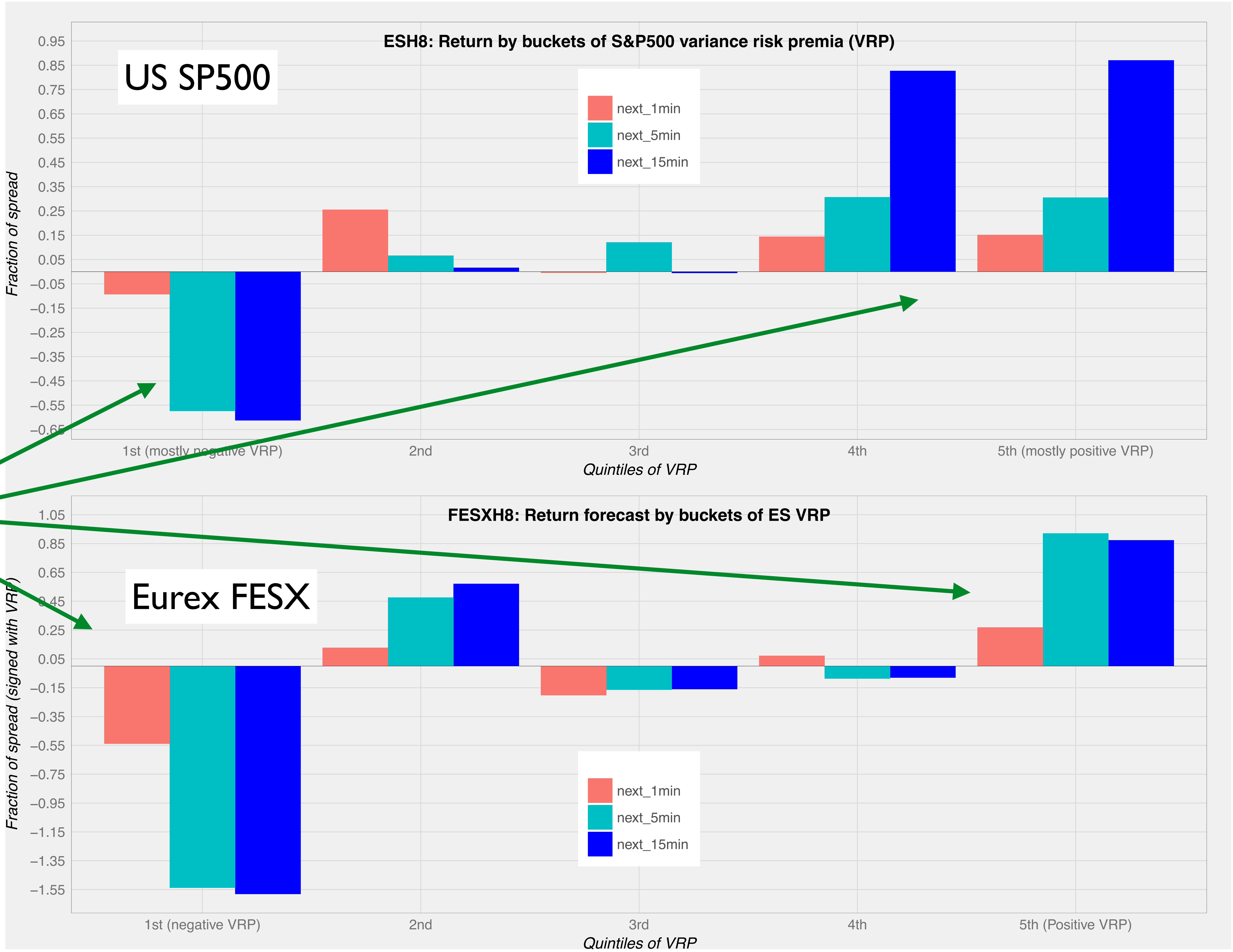
- Well-known at daily and slower time scales
- Novel at intraday trading

Data sources:

- Implied Vol from CBOE VIX futures (or traded options)
- Real-time realized vol from new QB indicator

Use for SP500 futures, and other products

VRP alone as signal



Extreme values predict forward price change

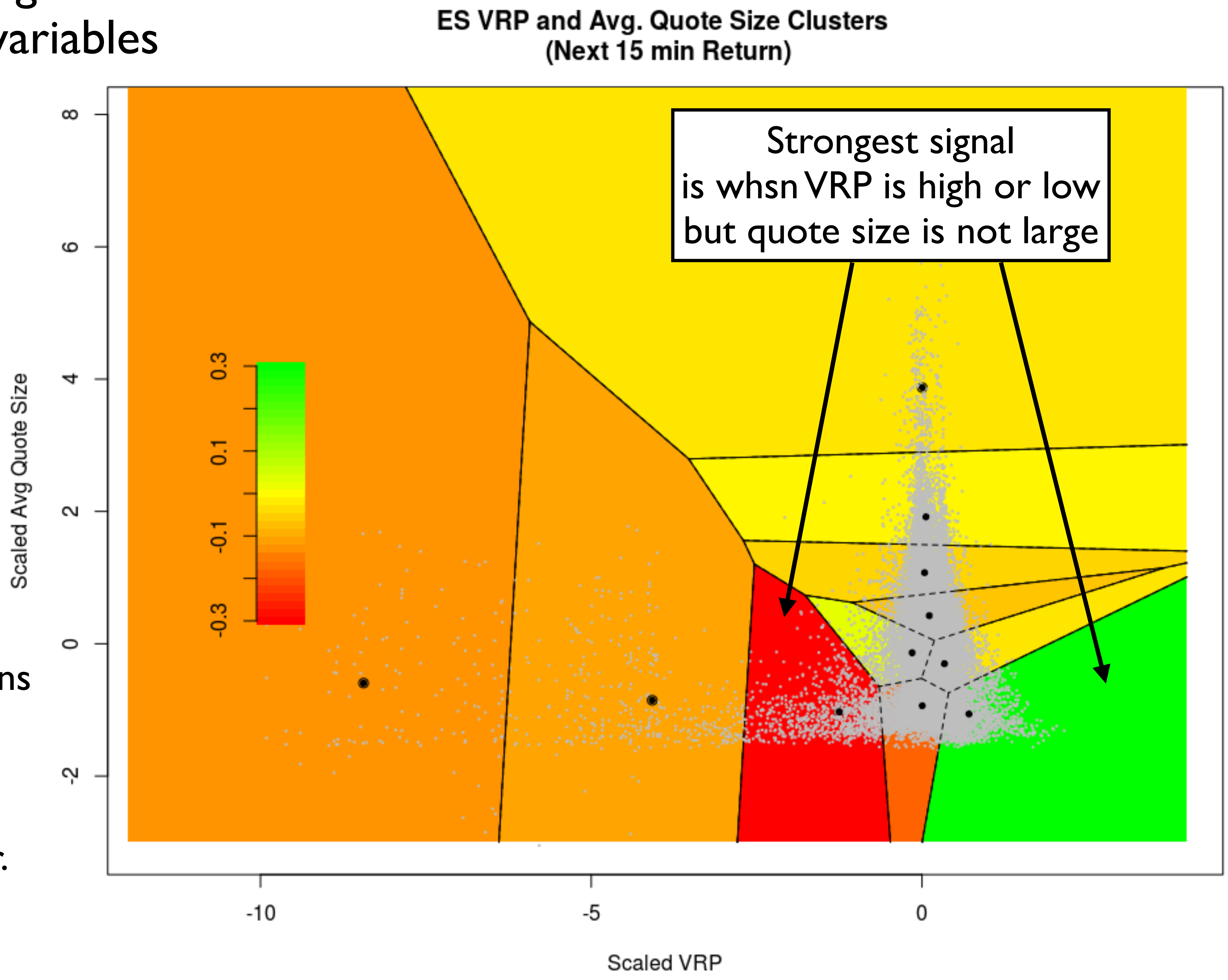
Combining with other variables (features) increases significance

Conditioning: significance of signal depends on other market state variables

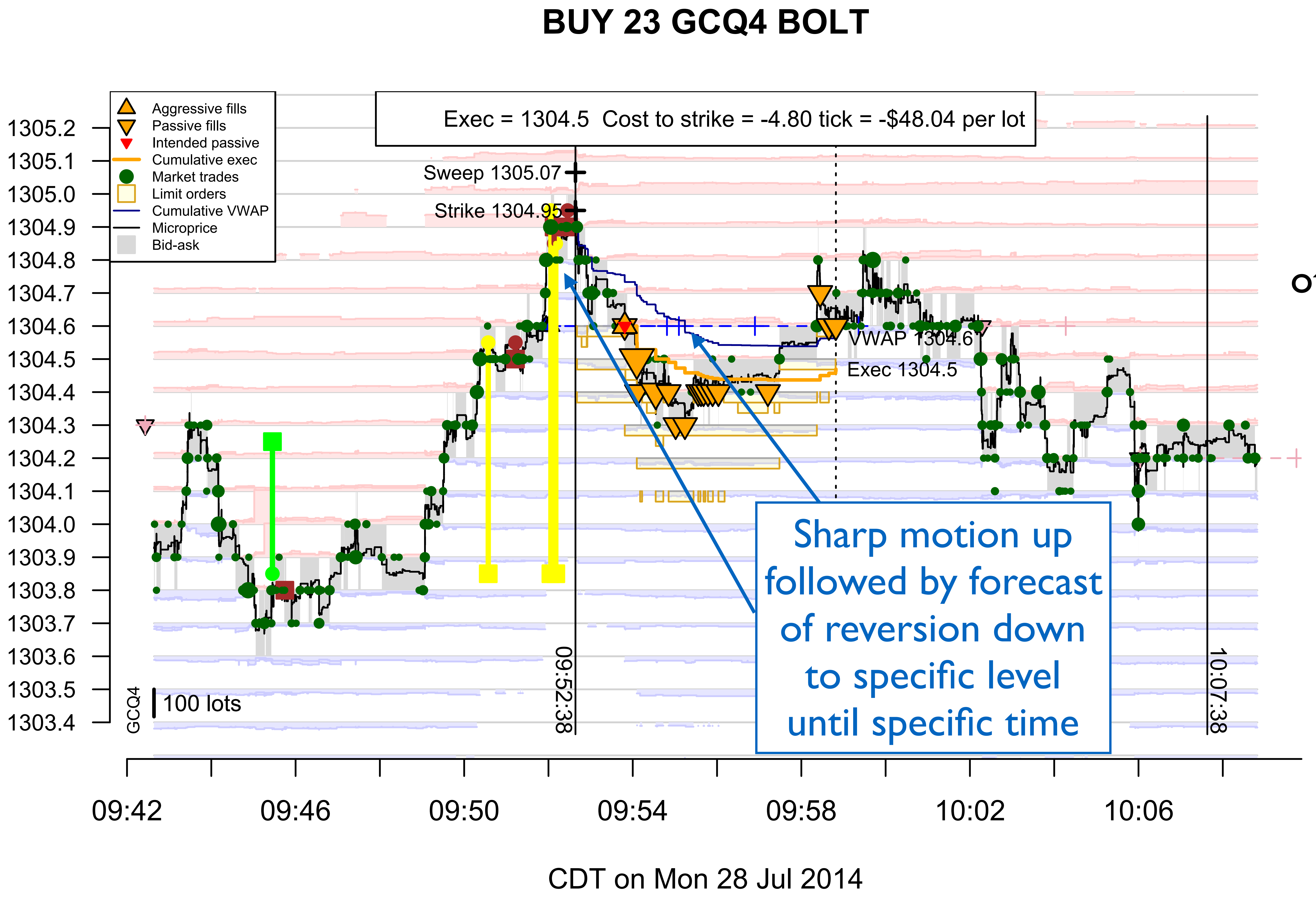
Use average quote size (a feature) as conditioning variable

Use average quote size (a feature) as conditioning variable

- Cluster (k-means) historical observations based on these two variables
- Compute average forward return in each cluster
- Substantially increases predictive power.



Sweep (reversion) signal



To make this work: condition on several other variables describing market state

Intraday bubbles

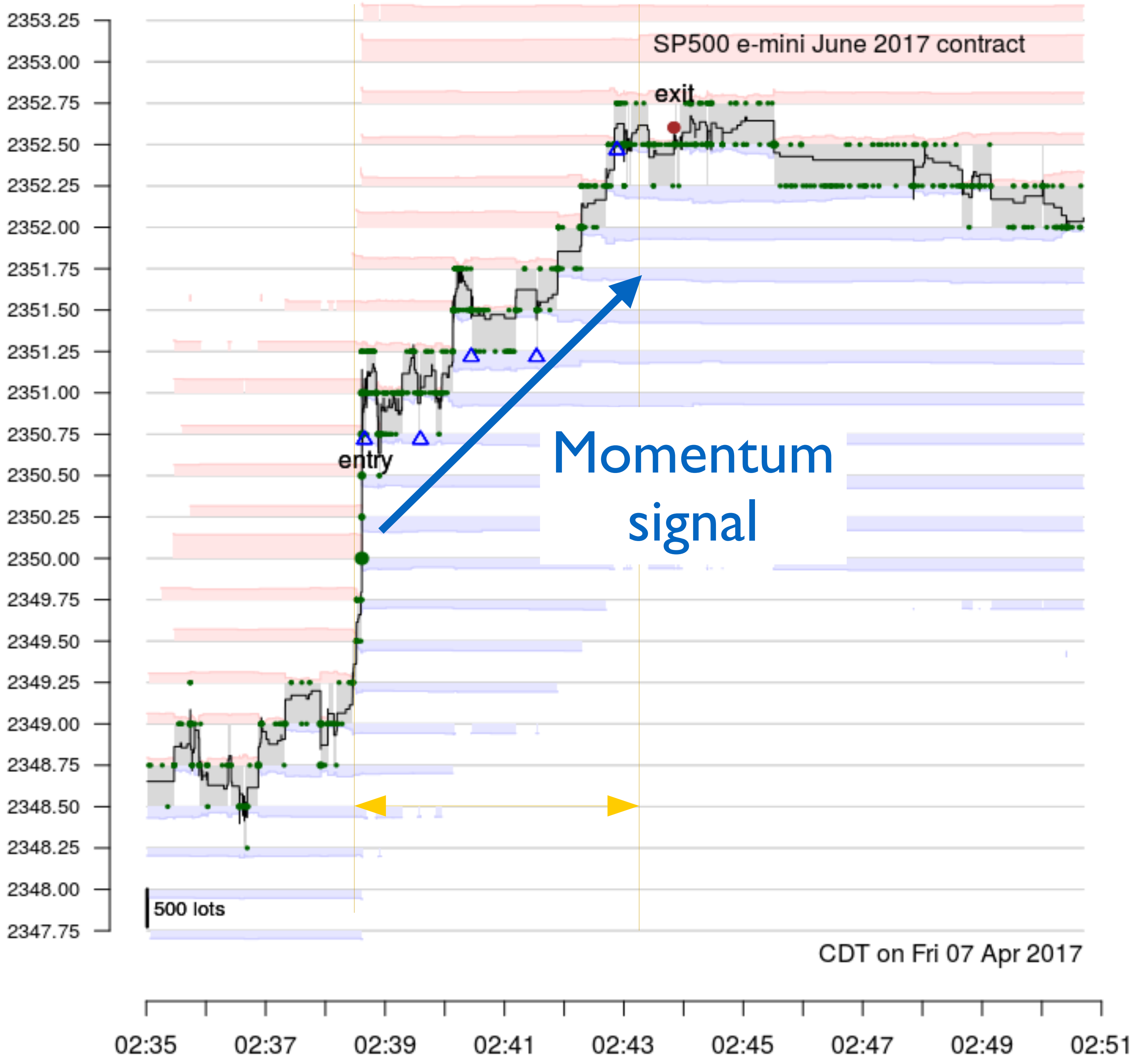
The Detection of Intra-Day Bubbles

- Test is a generalized version of Augmented-Dickey Fuller test of unit root
- The prototypical model takes the following form:

$$y_t = \rho\,(y_{t-1} - \bar{y}) + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

$$H_0 : \hat{\rho} = 1$$
$$H_1 : \hat{\rho} > 1$$

- When $\hat{\rho} > 1$ the price is believed to be in an explosive state.



Example Buy Signal

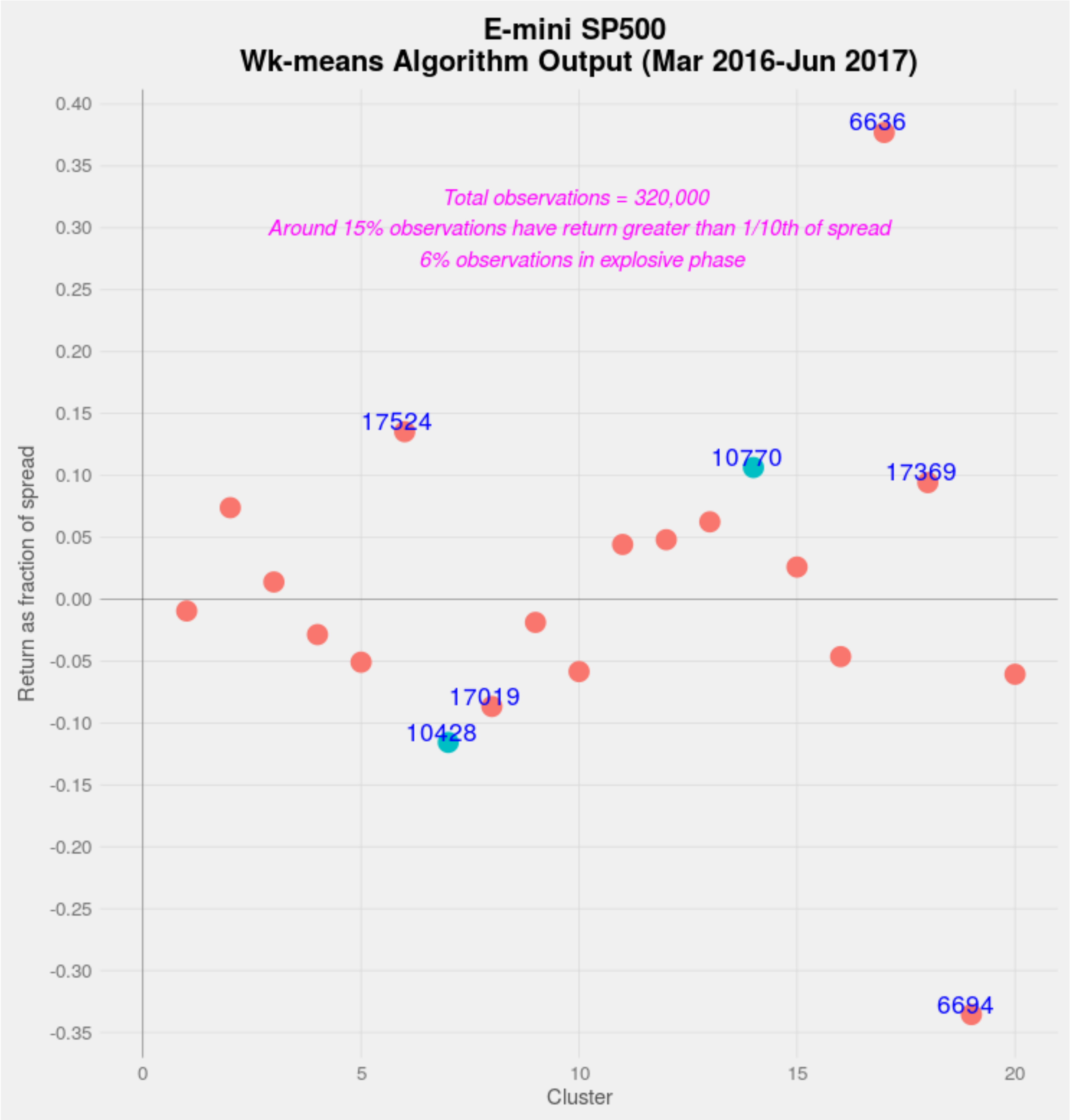
- The market was trending up
- Our model correctly identified this and produced a signal about 2 minutes after the rally started (around 2:39 am)
- The signal expired after the price flattened out (around 2:44 am).

Shankar Narayanan, Quantitative Brokers

To make this work: condition on several other variables describing market state

Condition on 5 different features to improve performance

Return by cluster

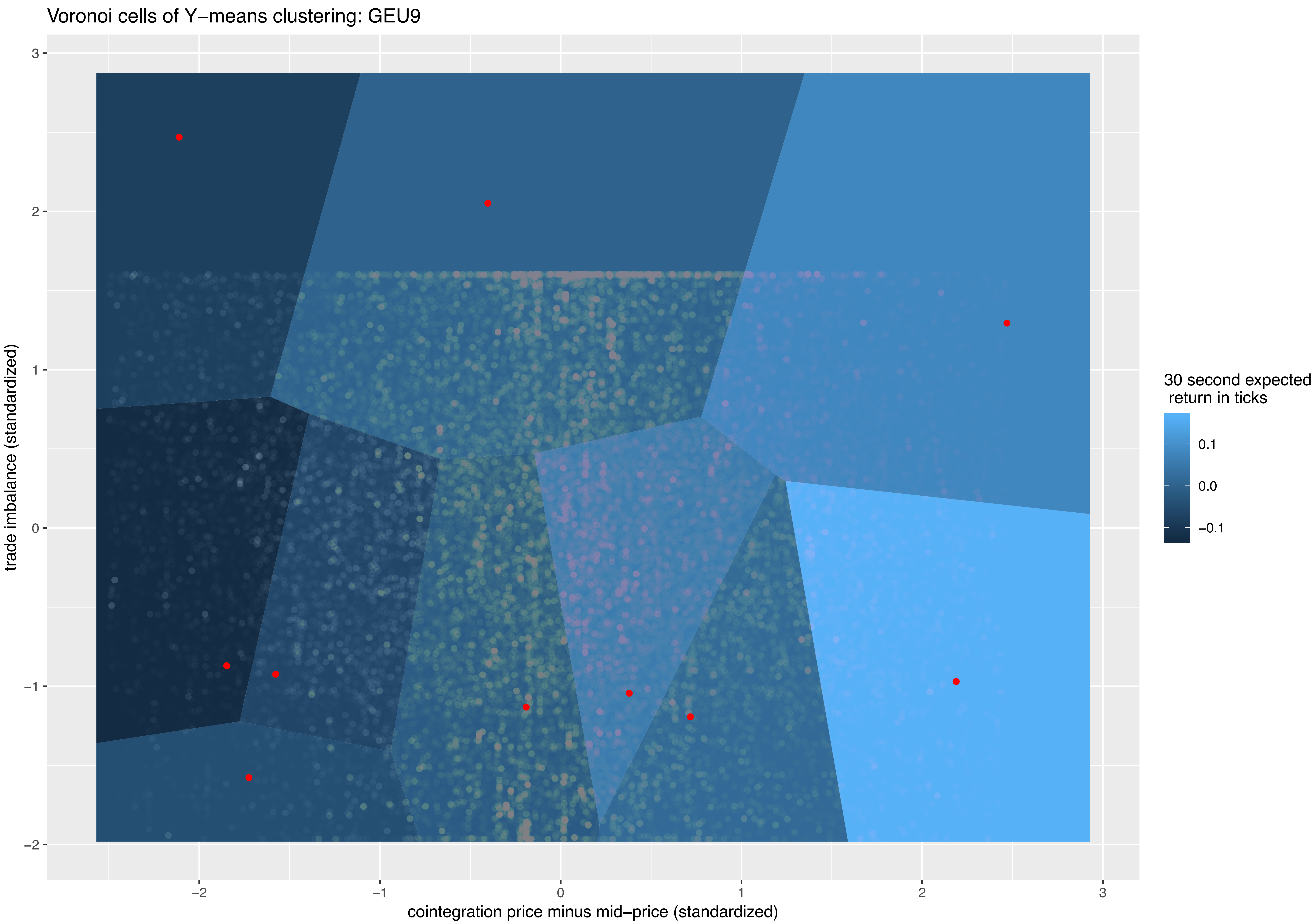


Cluster 7 auxiliary features
(Voronoi cells in 7 dimensions)

Sweep vs bubble

Sweep = reversion
Bubble = momentum

Importance of "consensus" layer, to make specific prediction to algorithm.



Consensus framework for signal combination

Yiming Peng,
QB and Northwestern

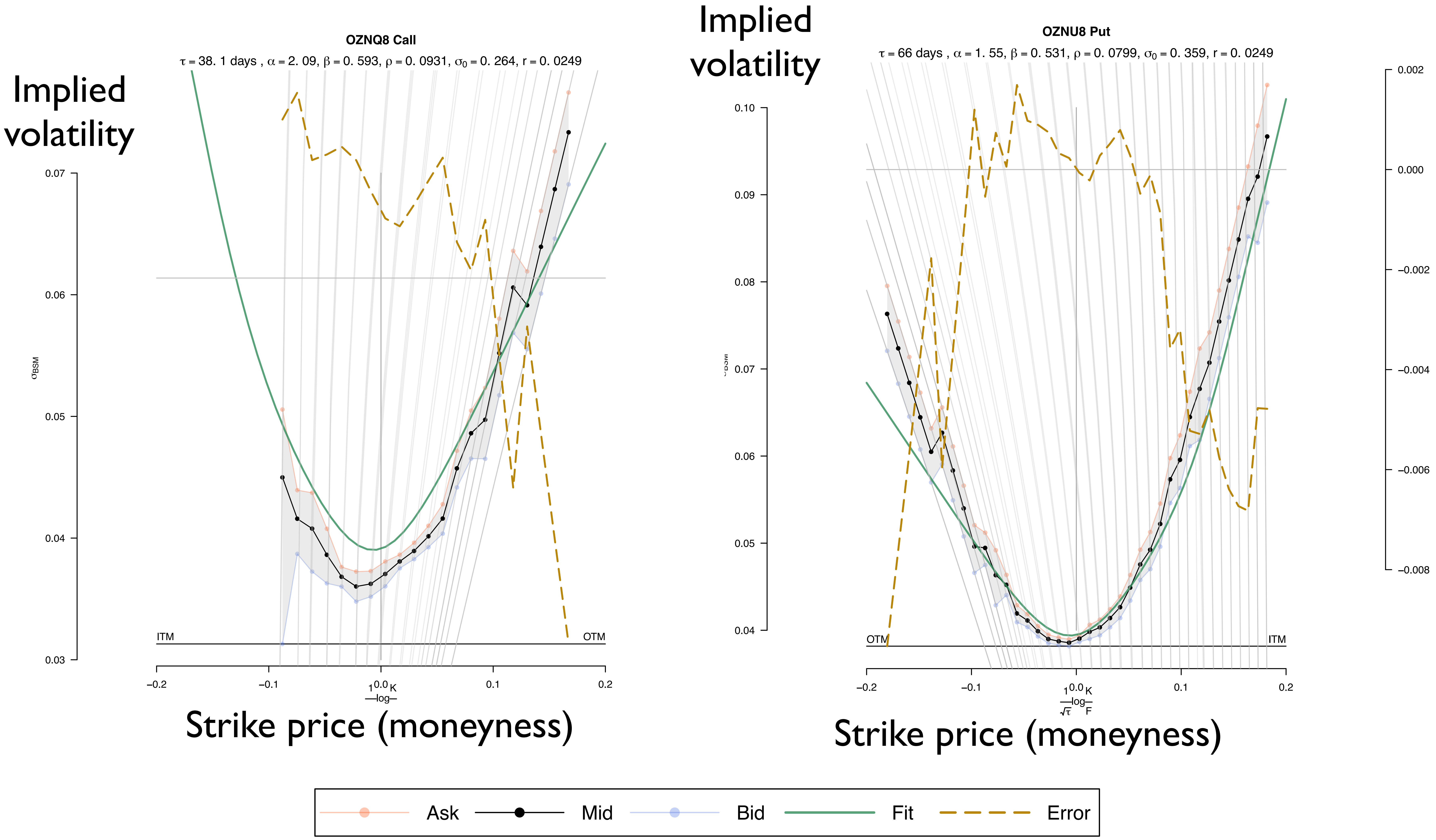
"Y-means" algorithm:
Like K-means, but cluster based on dependent variable (supervised learning)

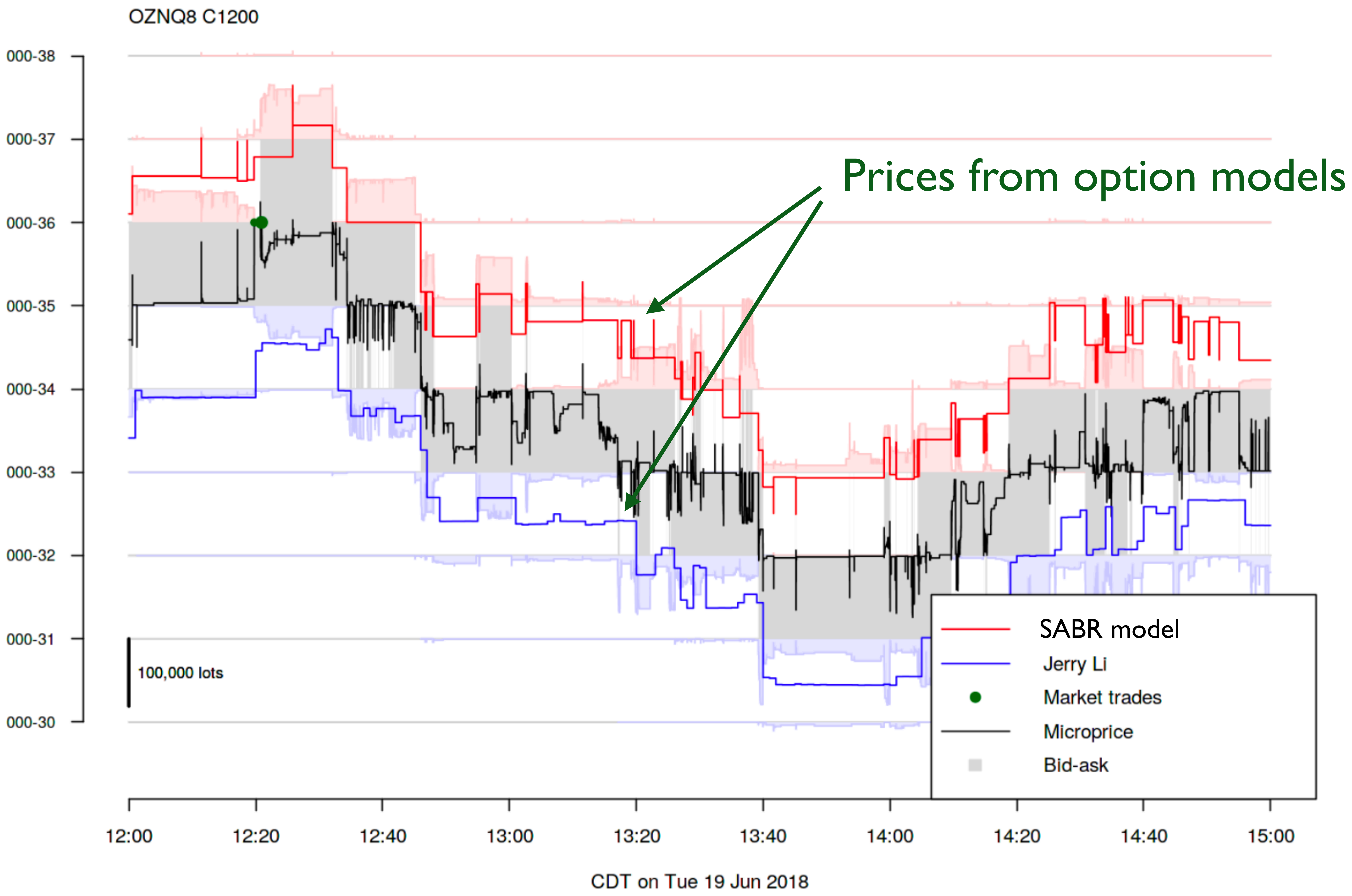
Option implied prices

Options trade in wide range of strikes
Complex combinations also have bid-ask quotes
Arithmetic relationships give indicative prices

Option pricing methods have persistent errors

SABR model has consistent errors at different parts of strike curve

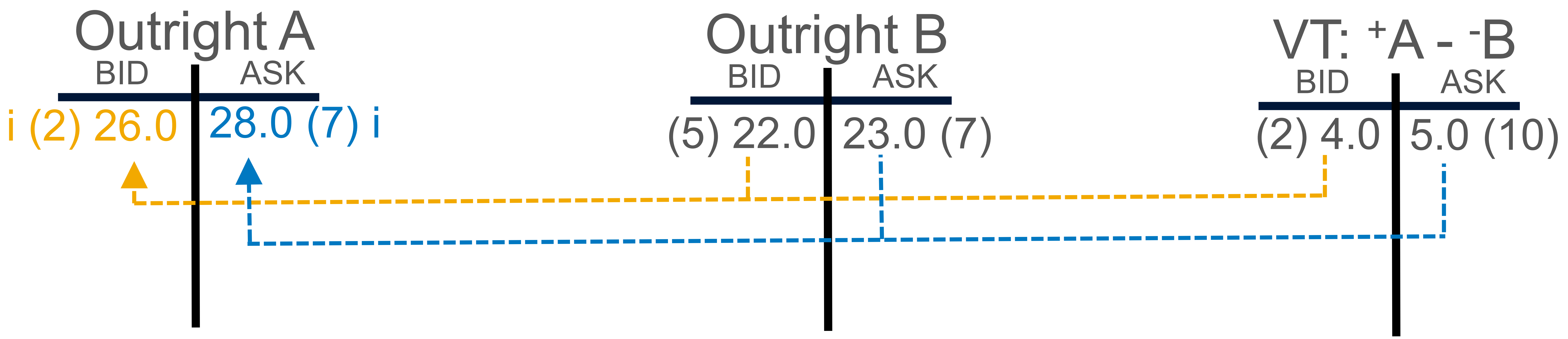



OZNQ8 C1200
000-38
000-37
000-36
000-35
000-34
000-33
000-32
000-31
000-30
100,000 lots
Prices from option models
SABR model
Jerry Li
Market trades
Microprice
Bid-ask
12:00
12:20
12:40
13:00
13:20
13:40
14:00
14:20
14:40
15:00
CDT on Tue 19 Jun 2018

Implied pricing

Familiar in futures contracts based on calendar spreads

Implied OUT: Real spread and outright orders create an implied order in an outright book

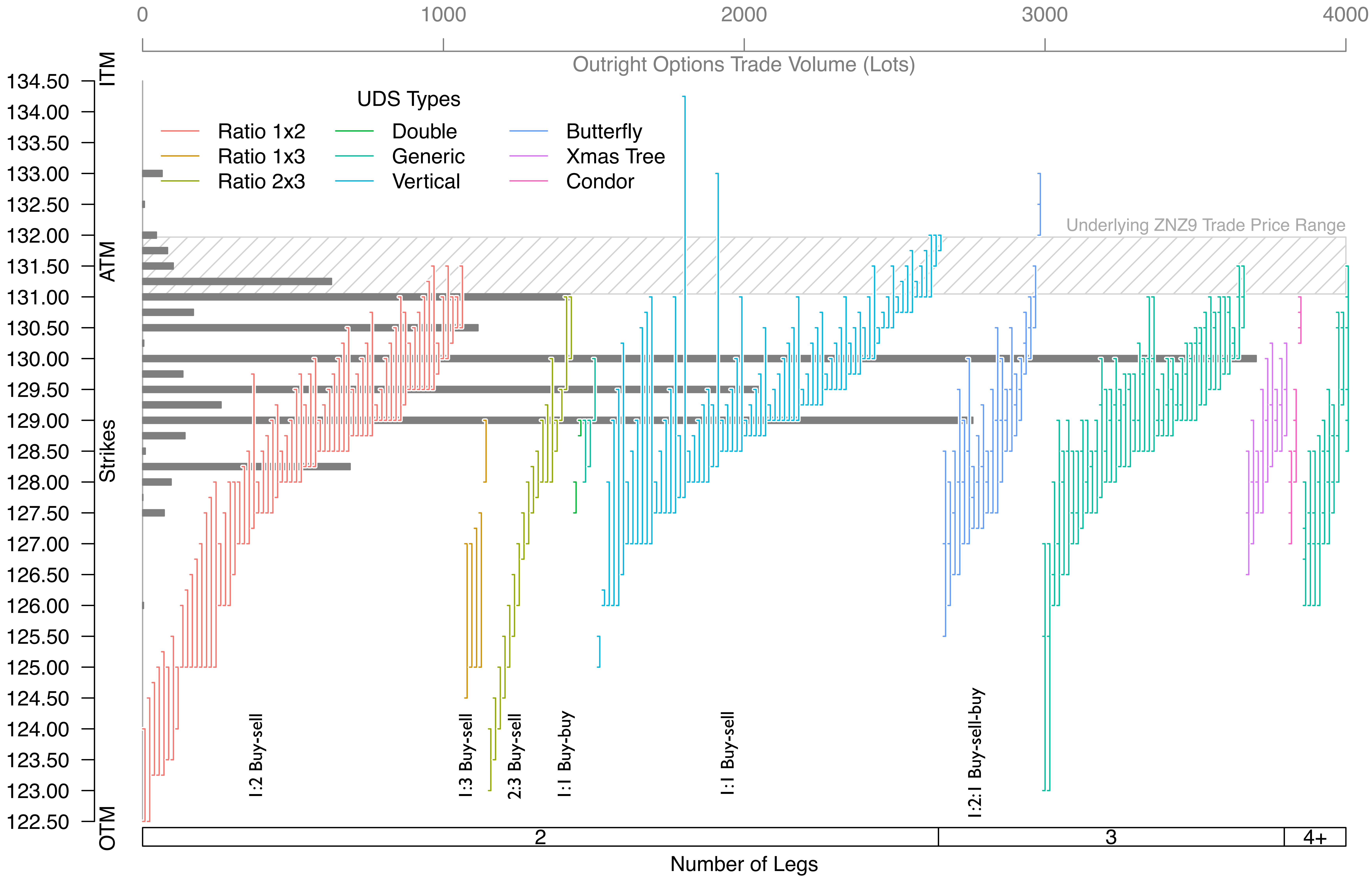


[A] = [B] + [A-B]

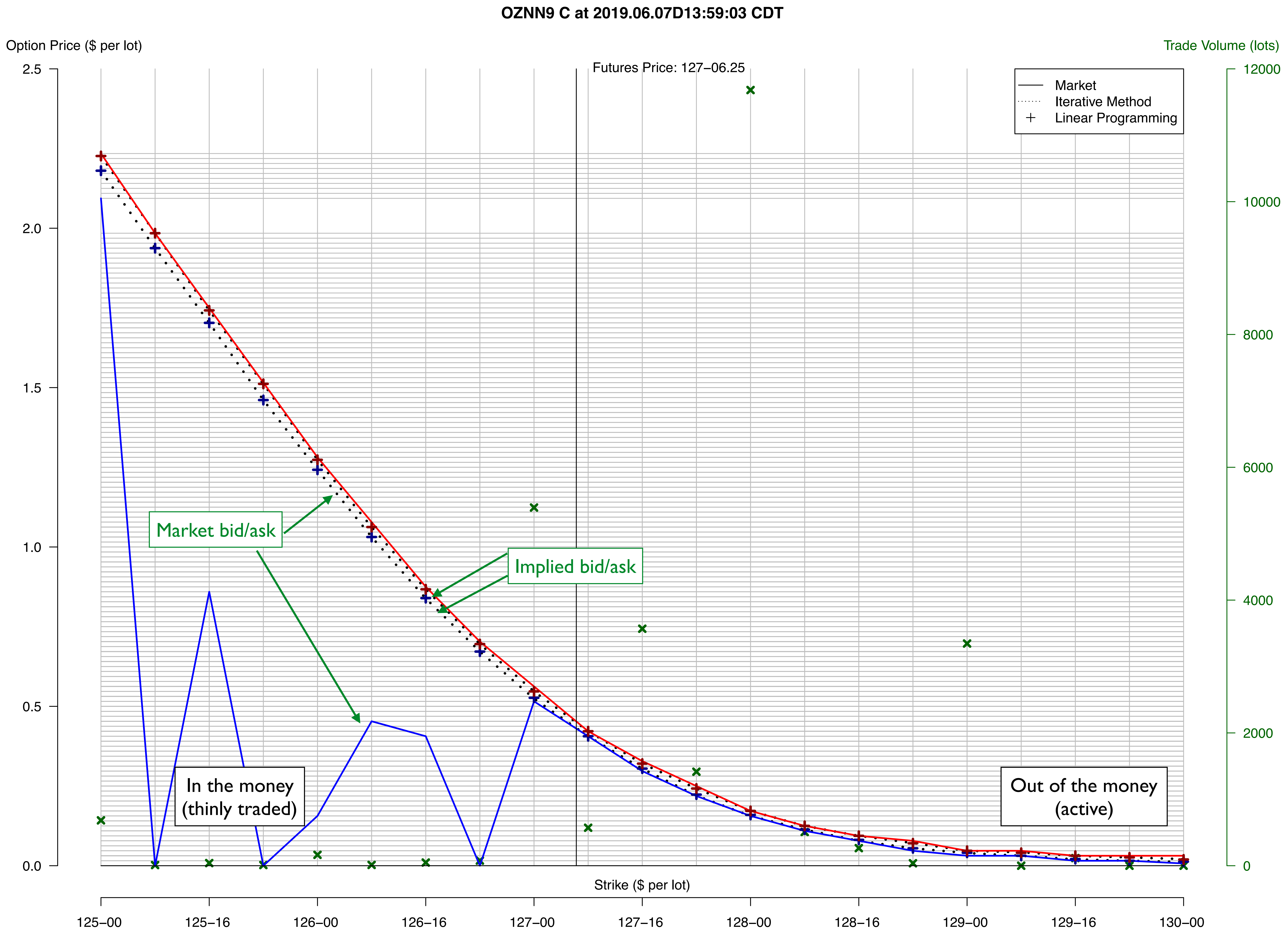
Calendar spreads are 1:1, so prices just add and subtract: prices are always on grid.

CME displays some implied quotes but not all. Important to compute independently for best prices

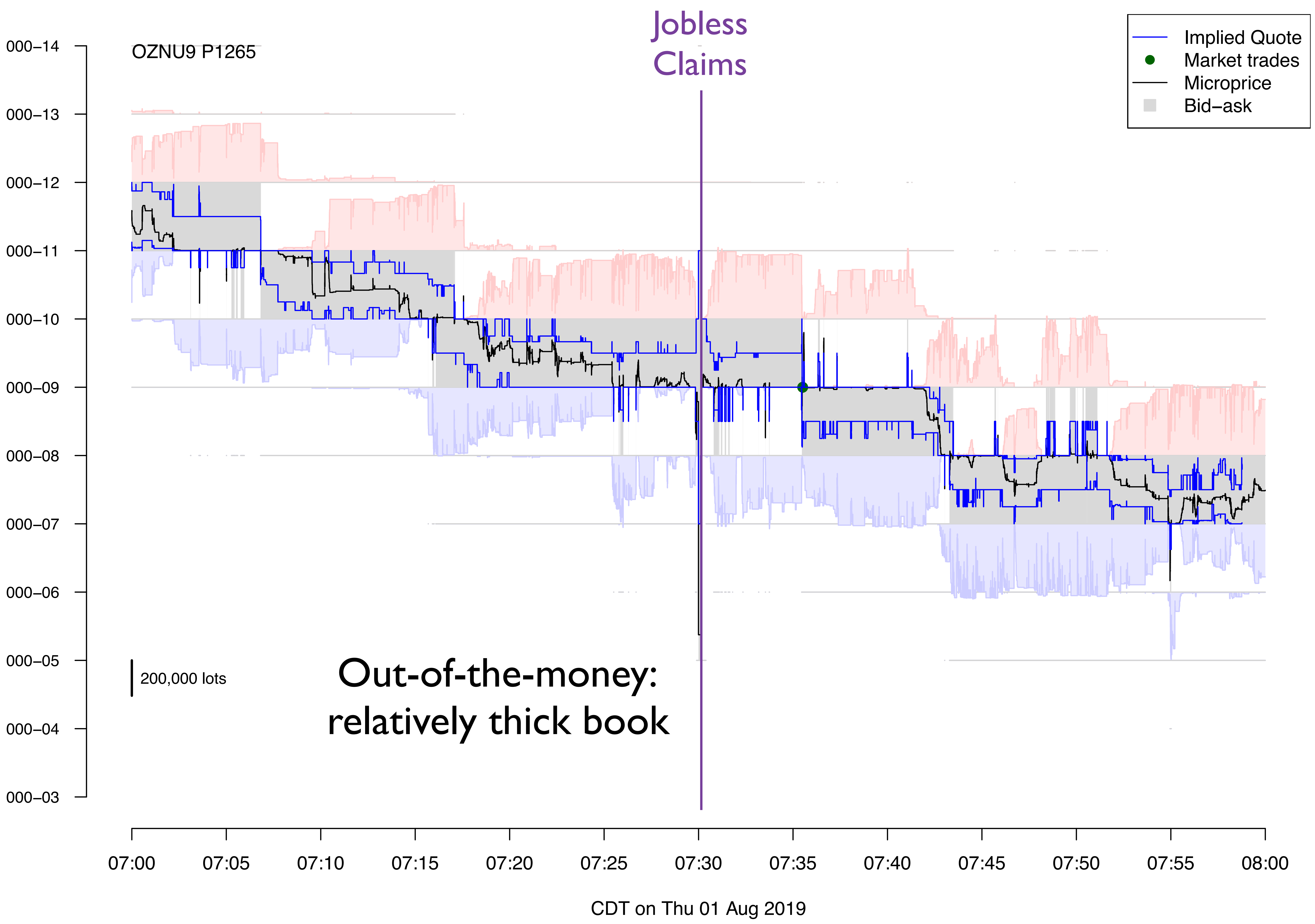
Option user-defined spreads



User Defined Spreads of OZNX9 Put Options on 2019−10−03


OZNN9 C at 2019.06.07D13:59:03 CDT
Option Price ($ per lot)
Trade Volume (lots)
Futures Price: 127−06.25
Market
Iterative Method
Linear Programming
Market bid/ask
Implied bid/ask
In the money
(thinly traded)
Out of the money
(active)
Strike ($ per lot)
2.5
2.0
1.5
1.0
0.5
0.0
12000
10000
8000
6000
4000
2000
0
125−00
125−16
126−00
126−16
127−00
127−16
128−00
128−16
129−00
129−16
130−00

Implied price compared with direct



Two examples

Smart Order Routing

Renyuan Xu, Isaac Carruthers

Y-means clustering approximation algorithm

Yiming Peng, Mengya Hu

Smart Order Routing

Multiple venues to trade same security

- Equities: dozens
- US Treasuries: BrokerTec, eSpeed, FENICS, + a few

All have same bid-ask quotes -- where to send limit order

- Maximise probability of fill in short time.

Order is filled when queue depletes

Optimal order placement in limit order markets

RAMA CONT*†‡ and ARSENIY KUKANOV§

†Department of Mathematics, Imperial College, London, UK
‡Laboratoire de Probabilités et Modèles Aléatoires, CNRS - Université Pierre & Marie Curie, Paris, France
§AQR Capital Management LLC, Greenwich, CT, USA

Quantitative Finance, 2016

(*Received 5 May 2015; accepted 28 April 2016; published online 17 June 2016*)

To execute a trade, participants in electronic equity markets may choose to submit limit orders or market orders across various exchanges where a stock is traded. This decision is influenced by characteristics of the order flows and queue sizes in each limit order book, as well as the structure of transaction fees and rebates across exchanges. We propose a quantitative framework for studying this *order placement* problem by formulating it as a convex optimization problem. This formulation allows the study of how the optimal order placement decision depends on the interplay between the state of order books, the fee structure, order flow properties and the aversion to execution risk. In the case of a single exchange, we derive an explicit solution for the optimal split between limit and market orders. For the general case of order placement across multiple exchanges, we propose a stochastic algorithm that computes the optimal routing policy and study the sensitivity of the solution to various parameters. Our algorithm does not require an explicit statistical model of order flow but exploits data on recent order fills across exchanges in the numerical implementation of the algorithm to acquire this information through a supervised learning procedure.

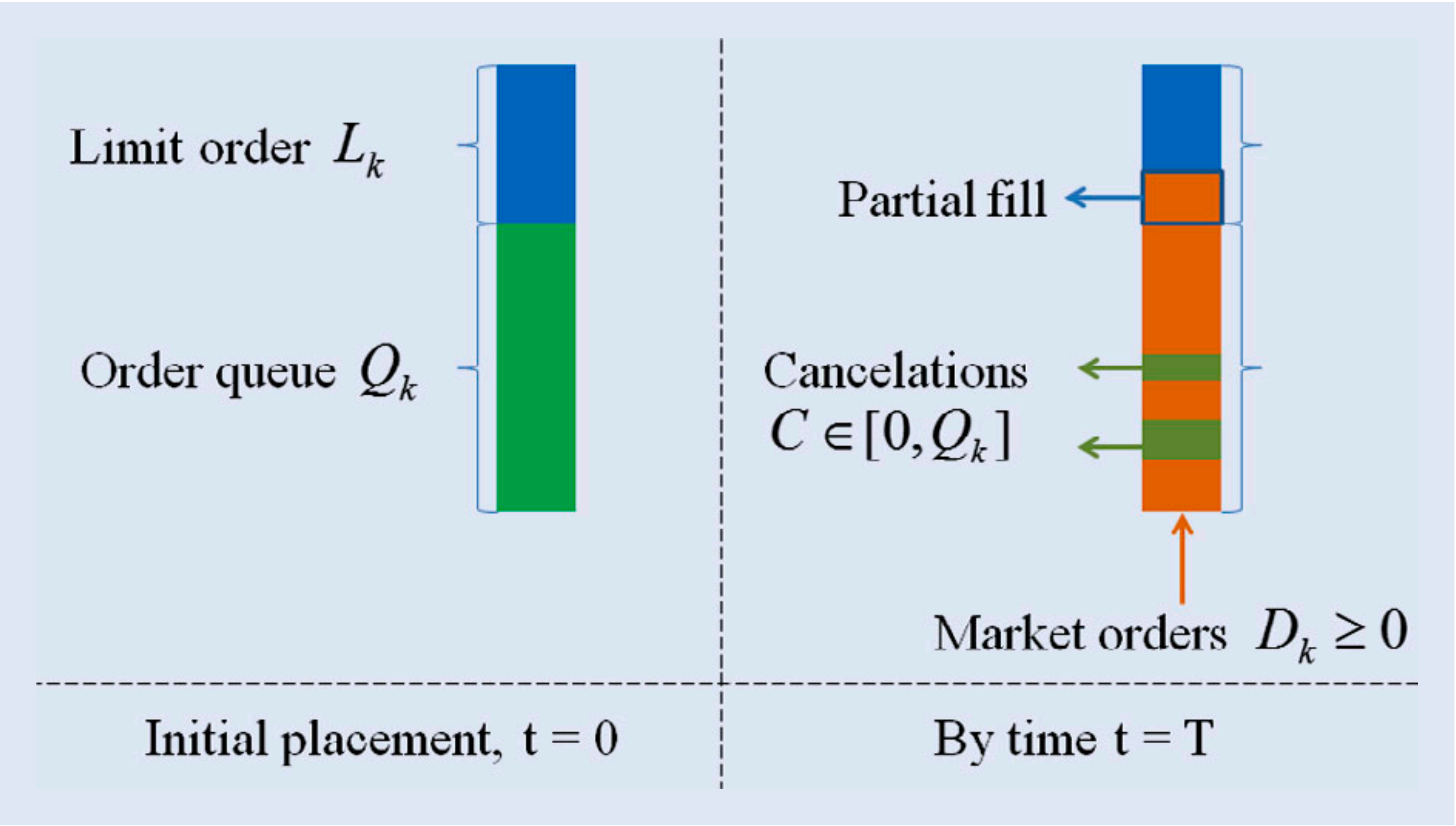


Figure 1. Limit order execution on exchange k depends on the order size L_k, the queue Q_k in front of it, total sizes of order cancellations C_k and marketable orders D_k, specifically on $\xi_k = C_k + D_k$.

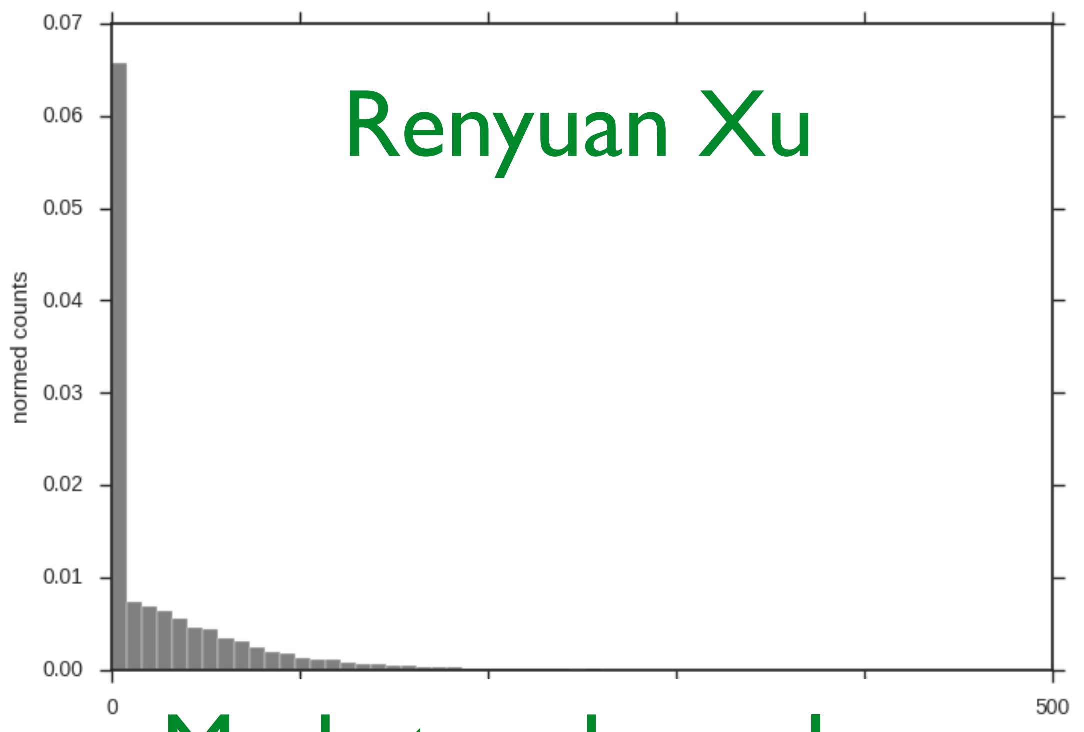
Problem 1 (Optimal order placement problem) An *optimal order placement* is a vector $X^* \in \mathbb{R}_+^{K+1}$ solution of

$$\min_{X \in \mathbb{R}_+^{K+1}} V(X) \tag{6}$$

where

$$V(X) = \mathbb{E}[v(X, \xi)] = \int_{\mathbb{R}^d} F(\mathrm{d}y) v(X, y) \tag{7}$$

is the expected execution cost for the allocation X and the expectation is taken with respect to the distribution F of order outflows $(\xi_1, \ldots, \xi_K)$ at horizon T.



Renyuan Xu

Market order volume

Need explicit model for joint distribution of order arrivals on all venues, then compute optimal strategy.
Better to do nonparametric construction directly for optimal action

Robert Almgren[1], Renyuan Xu[2]

[1] Quantitative Brokers, New York, NY 10008, USA,
ralmgren@quantitativebrokers.edu
[2] Unversity of California at Berkeley, Berkeley CA 94704, USA,
Quantitative Brokers, New York, NY 10008, USA,
renyuanxu@berkeley.edu

5.1 Basic features

$F_1 = \{P_k^{ask}, P_k^{bid}, Q_k^{ask}, Q_k^{bid}\}_{k=1}^K$

5.2 Time-insensitive set

$F_2 = \{P_k^{ask} - P_k^{bid}, \frac{P_k^{ask}+P_k^{bid}}{2}, \frac{P_k^{ask}*Q_k^{bid}+P_k^{bid}*Q_k^{ask}}{Q_k^{bid}+Q_k^{ask}}, \frac{Q_k^{bid}-Q_k^{ask}}{Q^k_{bid}+Q^k_{ask}}\}_{k=1}^K$

5.4 Time-dependent set

Denote $t = 1, 2, .., s$ as the number of look-back period with look-back window $\Delta w = 60s$, denote $F_4^t = \{f_{41}^t, f_{42}^t, f_{43}^t, f_{44}^t\}$, where

5.3 Time-sensitive set

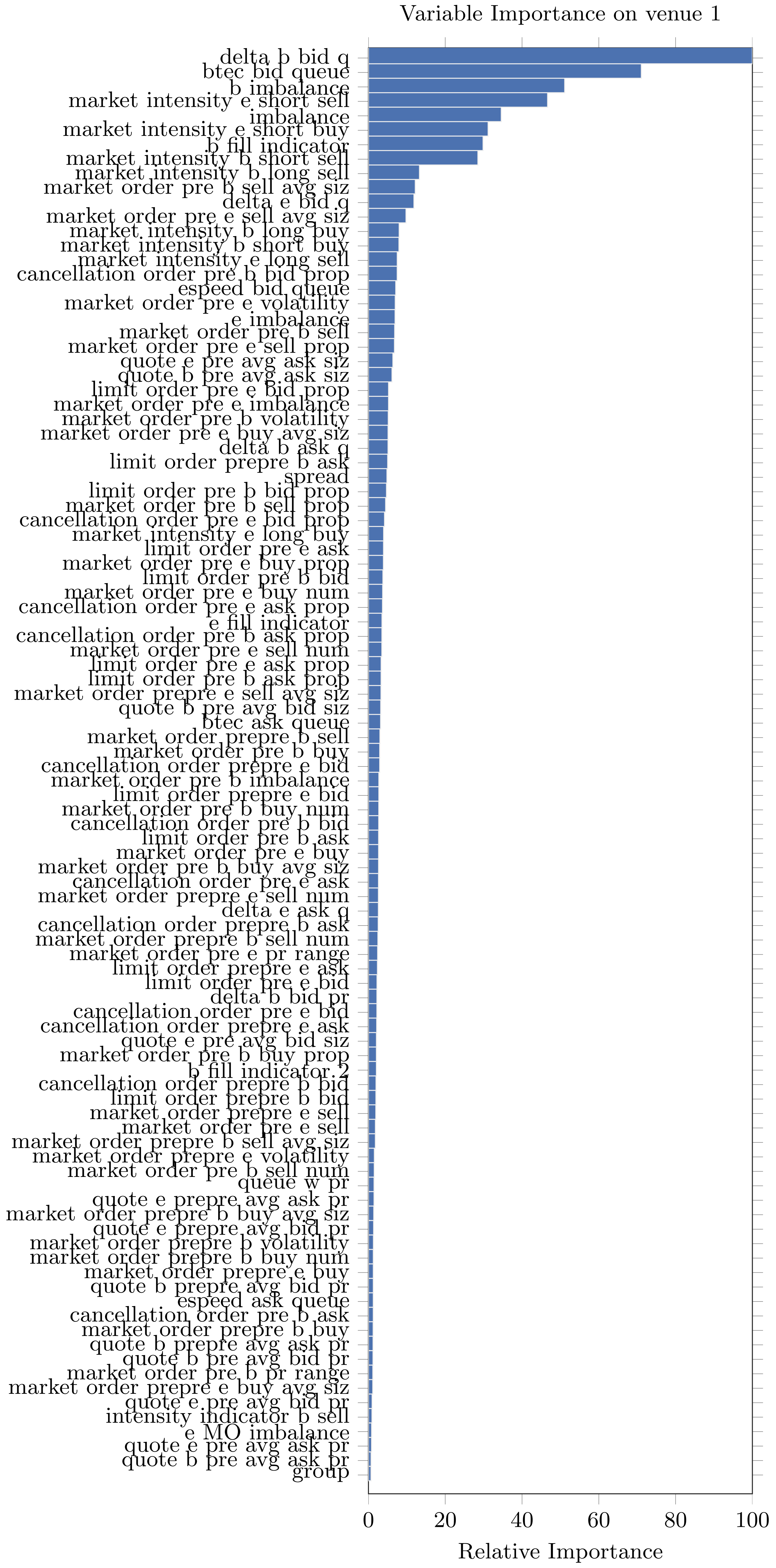
Denote $F_3 = \{f_{31}, f_{32}, f_{33}\}$, where

- $f_{31} = \{\frac{dP_k^{ask}}{dt}, \frac{dP_k^{bid}}{dt}, \frac{dV_k^{ask}}{dt}, \frac{dV_k^{bid}}{dt}\}_{k=1}^K$
- $f_{32} = \{\lambda^{la}_{k,\Delta t}, \lambda^{lb}_{k,\Delta t}, \lambda^{ma}_{k,\Delta t}, \lambda^{mb}_{k,\Delta t}, \lambda^{ca}_{k,\Delta t}, \lambda^{cb}_{k,\Delta t}\}_{k=1}^K$
- $f_{33} = \{\lambda^{la}_{k,\Delta T}, \lambda^{lb}_{k,\Delta T}, \lambda^{ma}_{k,\Delta T}, \lambda^{mb}_{k,\Delta T}, \lambda^{ca}_{k,\Delta T}, \lambda^{cb}_{k,\Delta T}\}_{k=1}^K$
- $f_{34} = \{\mathbf{1}_{\lambda^{la}_{k,\Delta t}>\lambda^{la}_{k,\Delta T}}, \mathbf{1}_{\lambda^{lb}_{k,\Delta t}>\lambda^{lb}_{k,\Delta T}}, \mathbf{1}_{\lambda^{ma}_{k,\Delta t}>\lambda^{ma}_{k,\Delta T}}, \mathbf{1}_{\lambda^{mb}_{k,\Delta t}>\lambda^{mb}_{k,\Delta T}}, \mathbf{1}_{\lambda^{ca}_{k,\Delta t}>\lambda^{ca}_{k,\Delta T}}, \mathbf{1}_{\lambda^{cb}_{k,\Delta t}>\lambda^{cb}_{k,\Delta T}}\}_{k=1}^K$

5.4 Time-dependent set

Denote $t = 1, 2, .., s$ as the number of look-back period with look-back window $\Delta w = 60s$, denote $F_4^t = \{f_{41}^t, f_{42}^t, f_{43}^t, f_{44}^t\}$, where

- $f_{42}^t = \{\bar{P}_k^{ask}, \bar{P}_k^{bid}, \max_n(P_{k,n}^{ask}), \max_n(P_{k,n}^{bid}), max_n(|P_{k,n}^{ask}-P_{k,n}^{bid}|, Vol_k^m, Vol_k^{bid}, Vol_k^{ask})\}_{k=1}^K$
- $f_{43}^t = \{\bar{Q}_{k,t}^{ask}, \bar{Q}_{k,t}^{bid}\}$
- $f_{44}^t = \{\mathbf{1}_{TV_{k,t}^{ma}>Q_k^{ma}}, \mathbf{1}_{TV_{k,t}^{mb}>Q_k^{mb}}\}$



Smart Order Routing

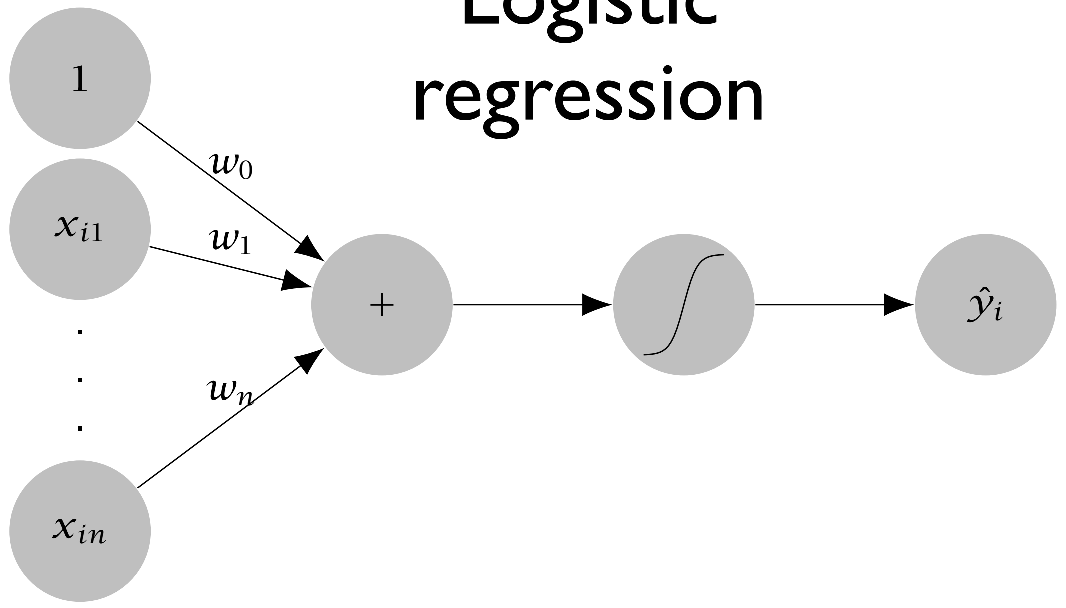
quantitativebrokers

MACHINE LEARNING FOR LIMIT-ORDER ROUTING IN CASH TREASURY MARKETS

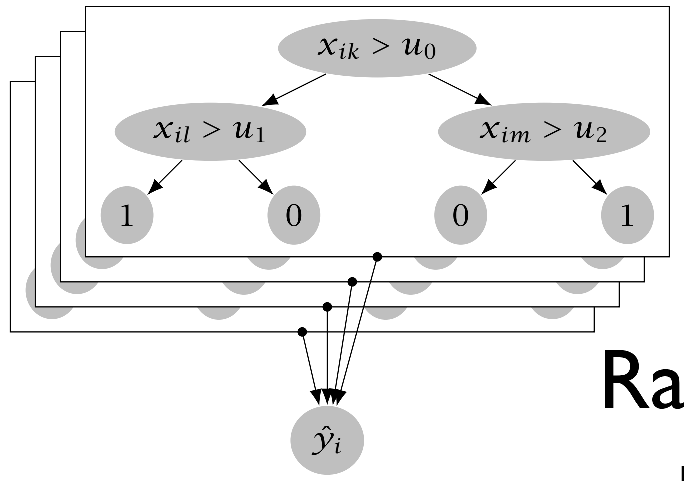
RENYUAN XU
ISAAC CARRUTHERS
APRIL 25, 2018

$$\max_{X\in\mathbb{Z}^k_+} \mathbb{E}\left[\sum_{i=1}^{k} \min\left(X_i, (\xi_i - Q_i)_+\right) \middle| M^*(X)\right]$$

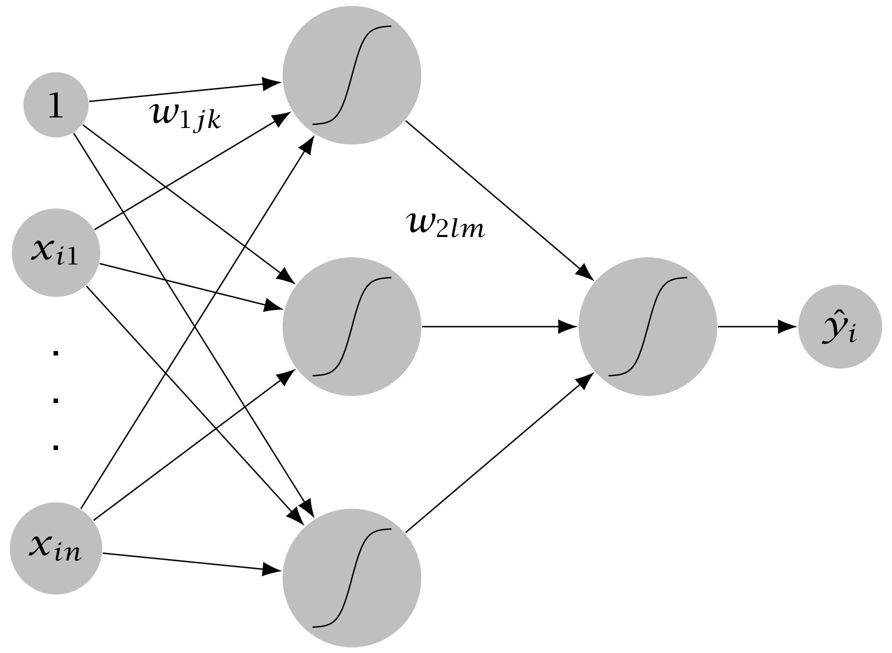
$$\text{s.t.} \sum_{i=1}^{k} X_i = S.$$



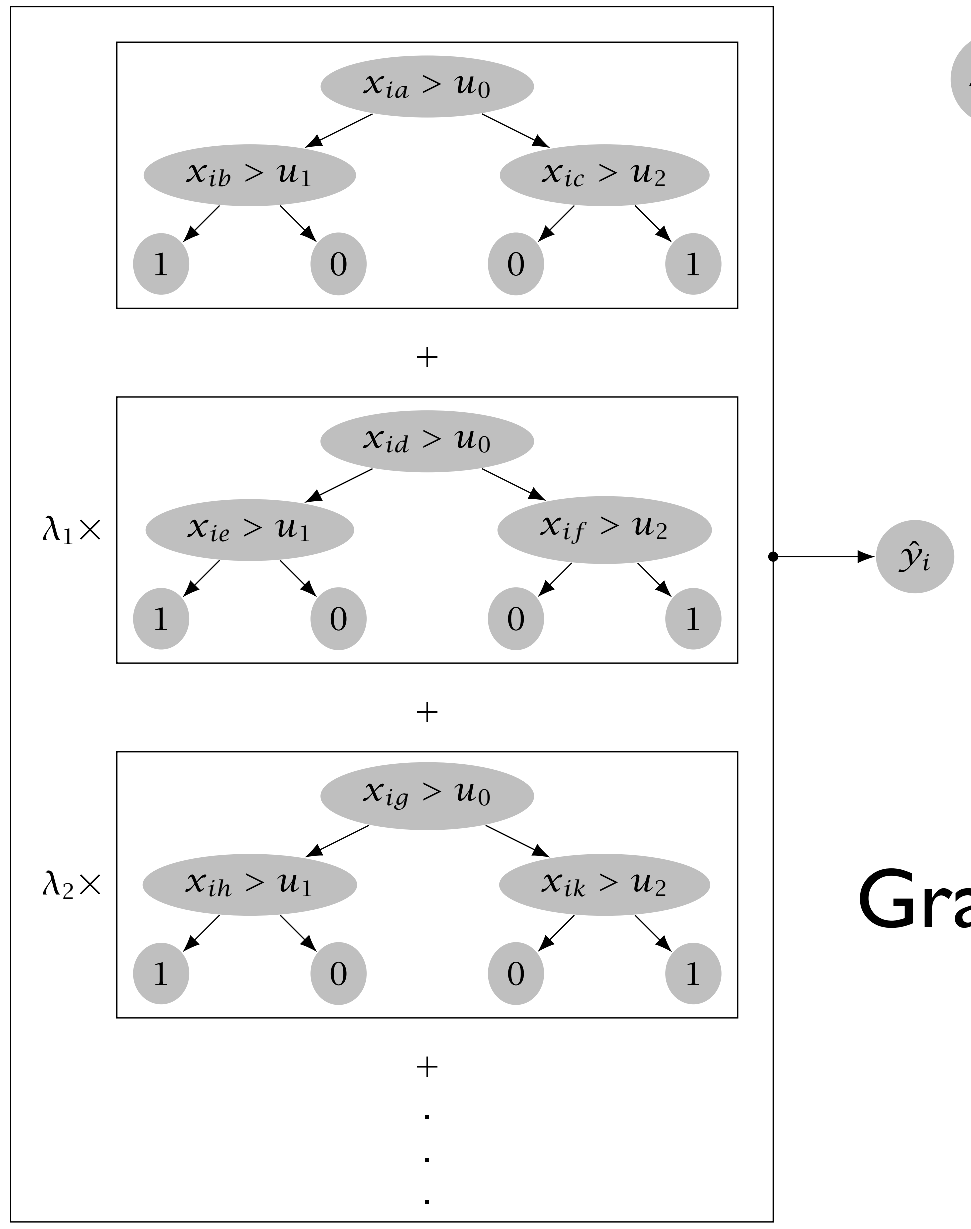
Logistic regression



Random forest regression



Multilayer perceptron



Gradient boosting regression

MARKET-DATA FEATURES

To establish a set of predictive market-data features, we designed and implemented a set of 52 different features per venue. This set contained a wide variety of calculations based on the recent history of market data, including recent price change, queue size change, signed volume, etc. From this set, we then selected a subset of 9 features per exchange, plus a single feature for aggregated quote imbalance across exchanges. We drew this subset by training an gradient boosting tree regressor on the data, and then selecting the features which provided the greatest improvement in accuracy on average.

Consensus framework

Conflicting signals

- Sweep = reversion
- Bubble = momentum

"Consensus" layer makes specific predictions to algorithm.
Also condition on market state variables.

Generic problem

$y = F(x)$ y scalar $x \in R^d$

x = signal outputs, and market state, d~10-15
y = forward return

N observations $x_1,\ldots,x_N$

how to model F?

What combination of signals gives the best prediction of future price changes, in what market conditions?

Classic problem of supervised learning

Regression

Clustering and partition

- support vector machines
- K-means
- etc

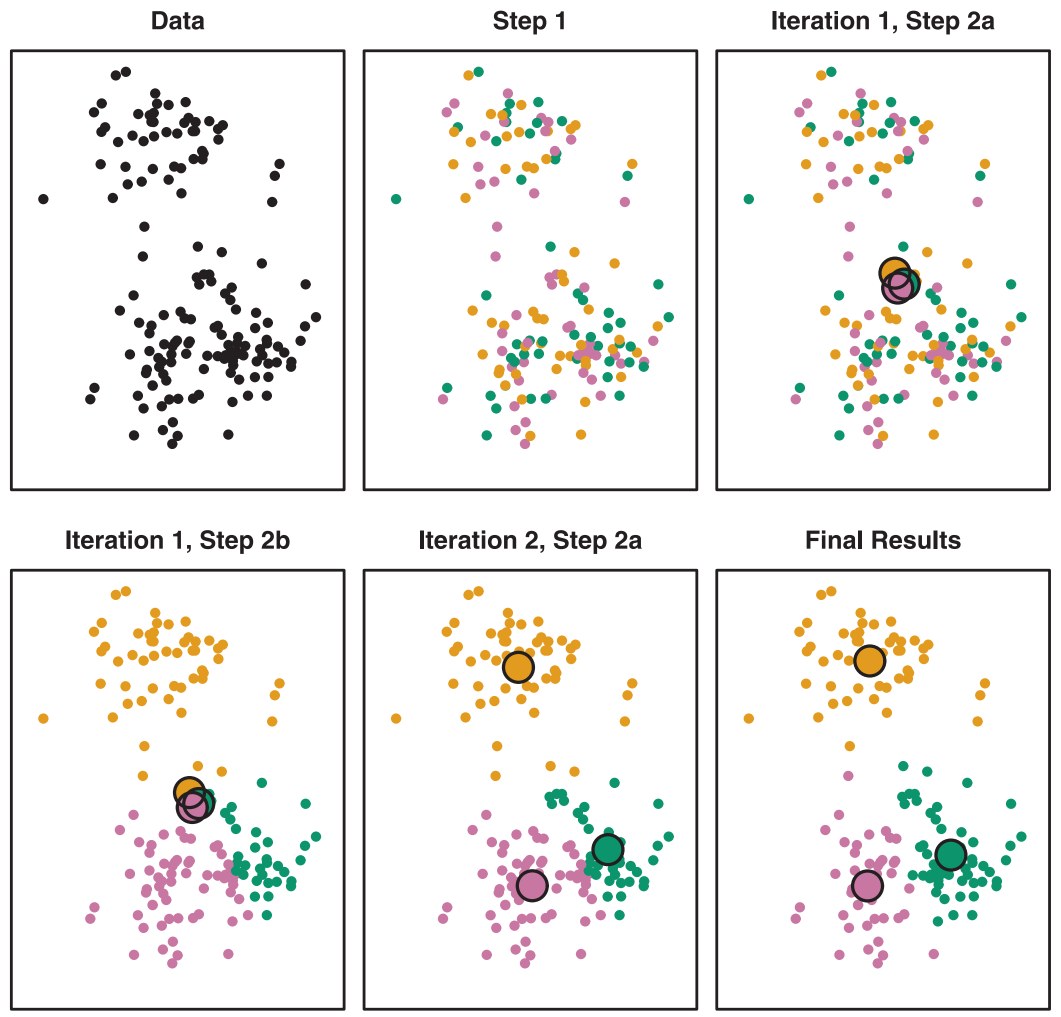
Combination methods

- random forest, etc

Gareth James • Daniela Witten • Trevor Hastie
Robert Tibshirani

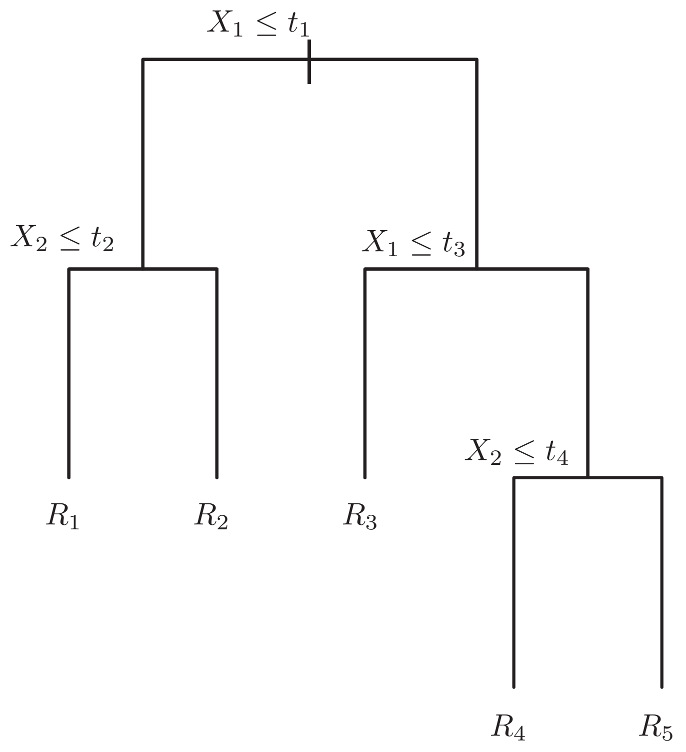
An Introduction to Statistical Learning
with Applications in R

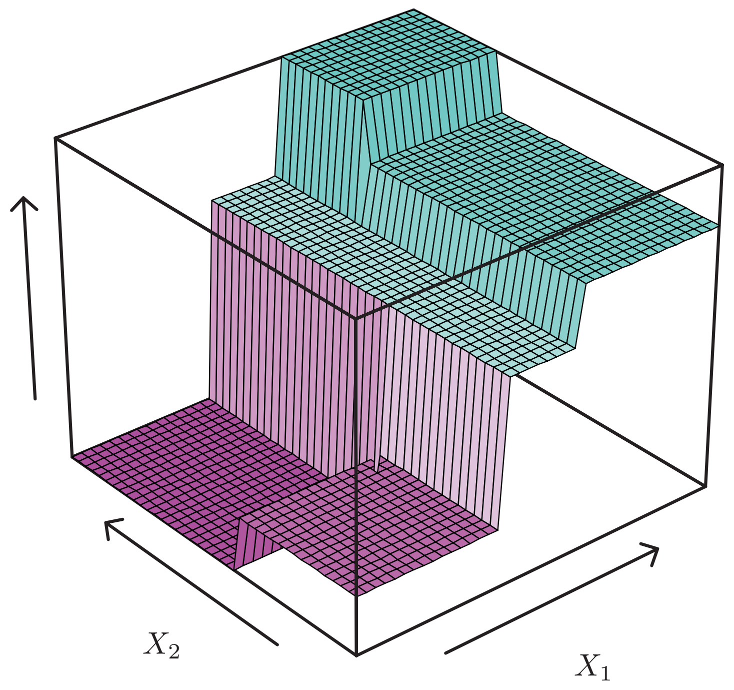
K-means



Determine clusters based on distribution of x (ignoring y)
Fit a constant function in each cluster

Hierarchical clustering
is similar





Y-means makes two innovations

Determine Voronoi clusters based on residuals in y

rather than distances in x

Use linear approximation in each cluster

rather than constant function

Resulting approximation is very accurate
and very quick to evaluate

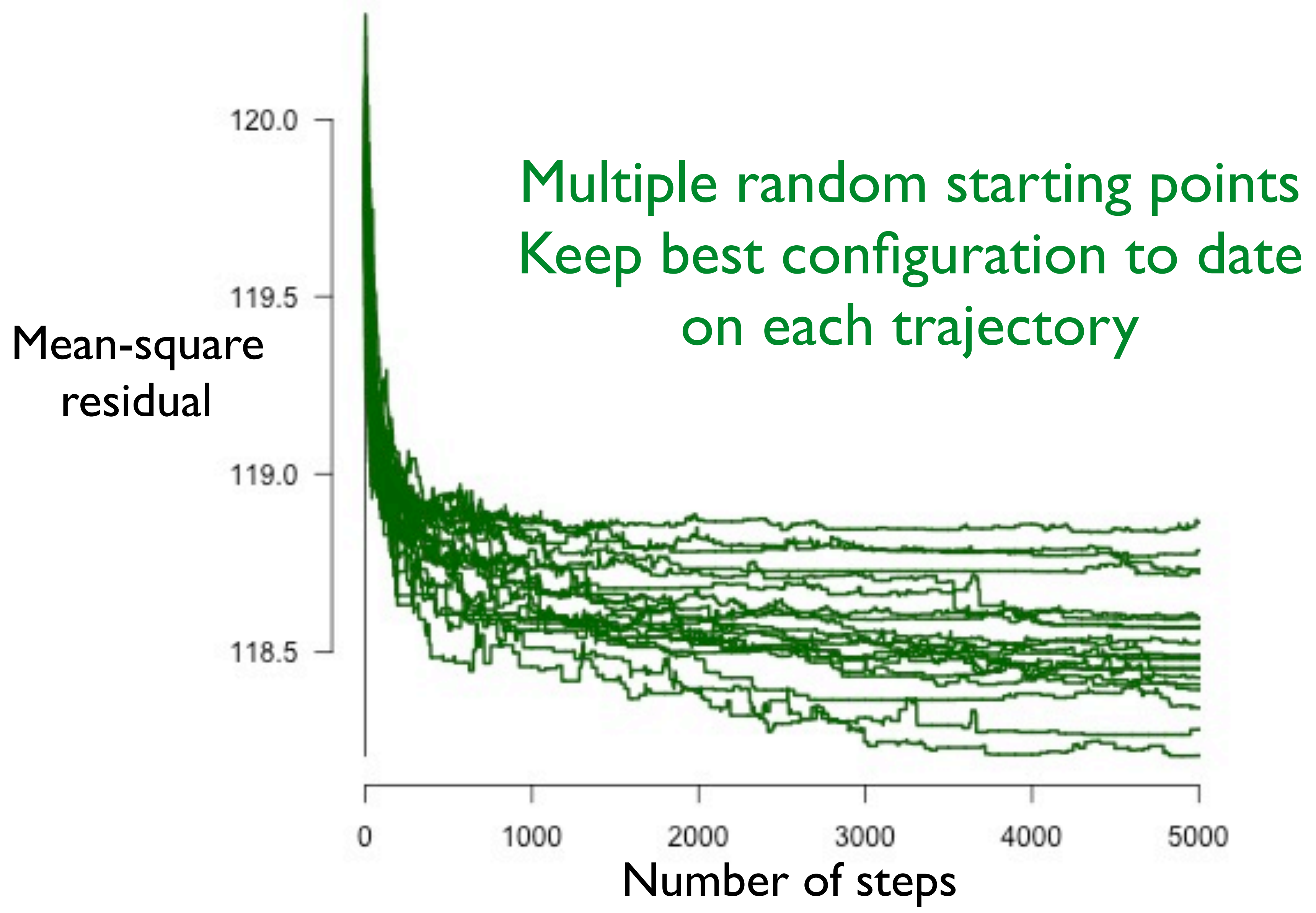
$$F_k(x) = \bar{y}_j + \beta_j'(x - \bar{x}_k), \qquad \text{for } x \in C_k$$

$$\min_{C_k,\ldots,C_K} \sum_{j=1}^{N} \frac{1}{2}\left(y_j - F_k(x_k)\right)^2$$

$$C_k, \ldots, C_K \;=\; \text{Voronoi cells}$$

Cells are parameterized
by node locations

Difficulty is optimizing node locations
Use simulated annealing:
slow and finicky, but results are good
Very fast to evaluate in real time



THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

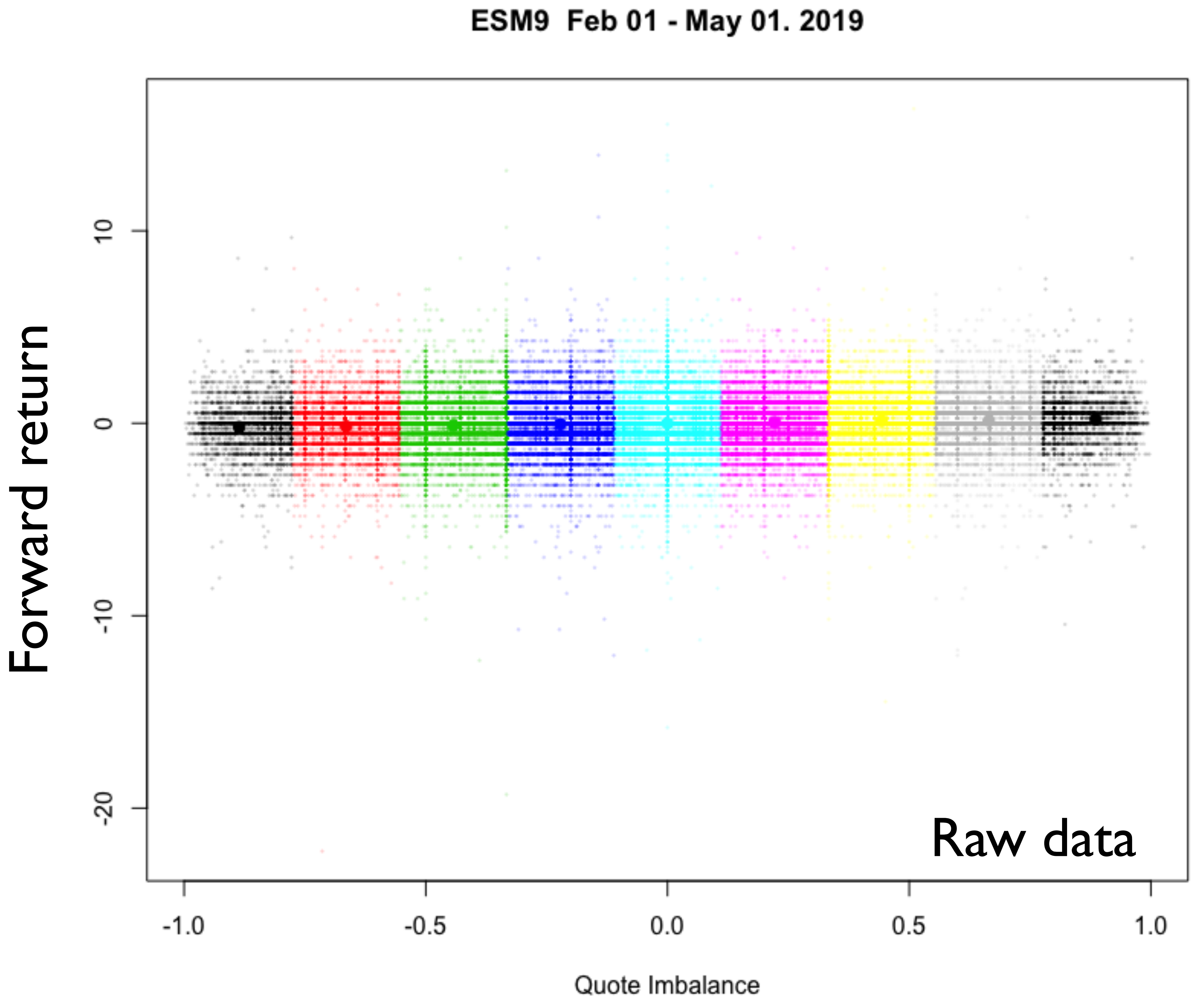
13 May 1983, Volume 220, Number 4598 **SCIENCE**

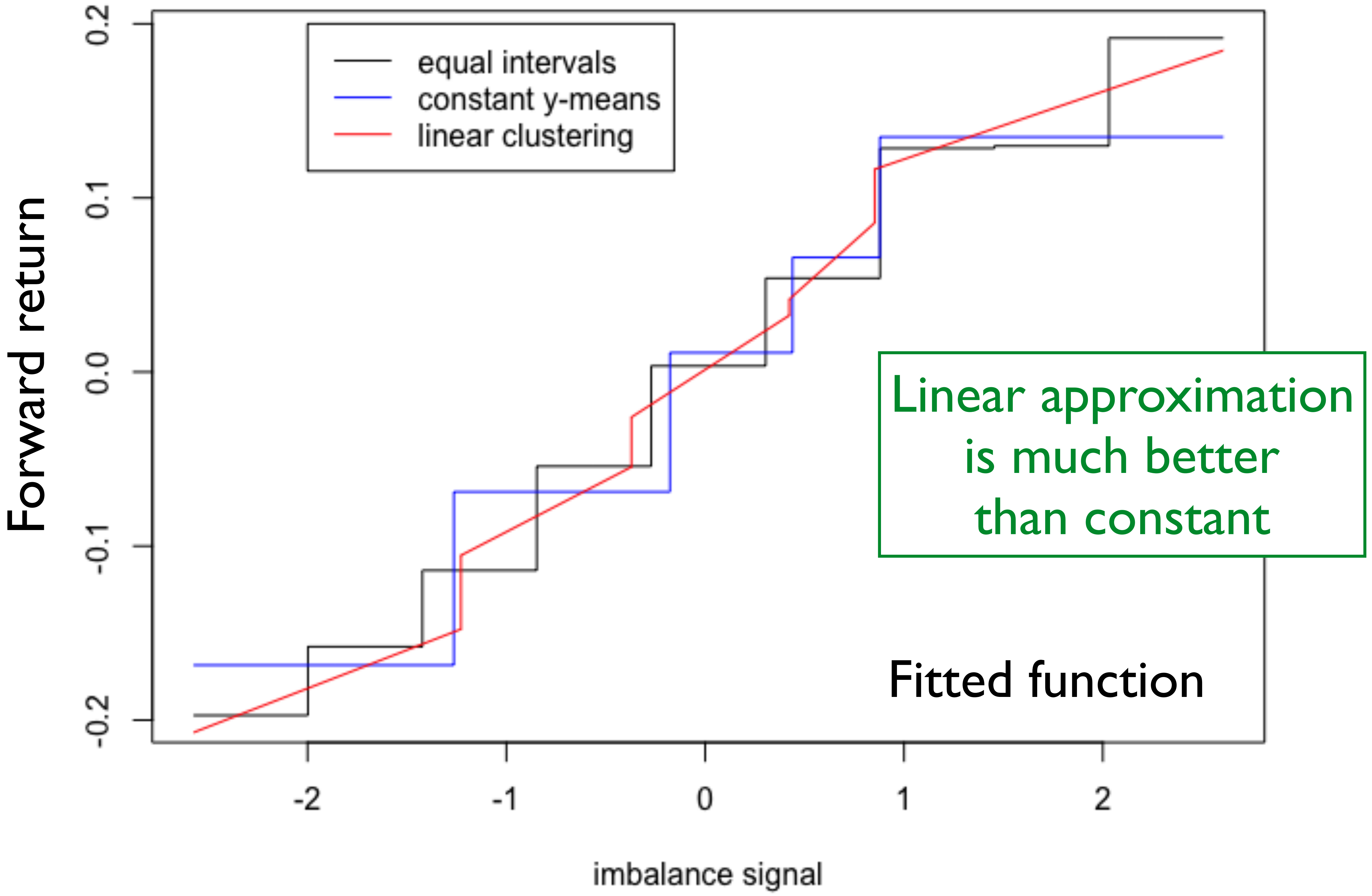
Optimization by Simulated Annealing

S. Kirkpatrick, C. D. Gelatt, Jr., M. P. Vecchi

Summary. There is a deep and useful connection between statistical mechanics (the behavior of systems with many degrees of freedom in thermal equilibrium at a finite temperature) and multivariate or combinatorial optimization (finding the minimum of a given function depending on many parameters). A detailed analogy with annealing in solids provides a framework for optimization of the properties of very large and complex systems. This connection to statistical mechanics exposes new information and provides an unfamiliar perspective on traditional optimization problems and methods.

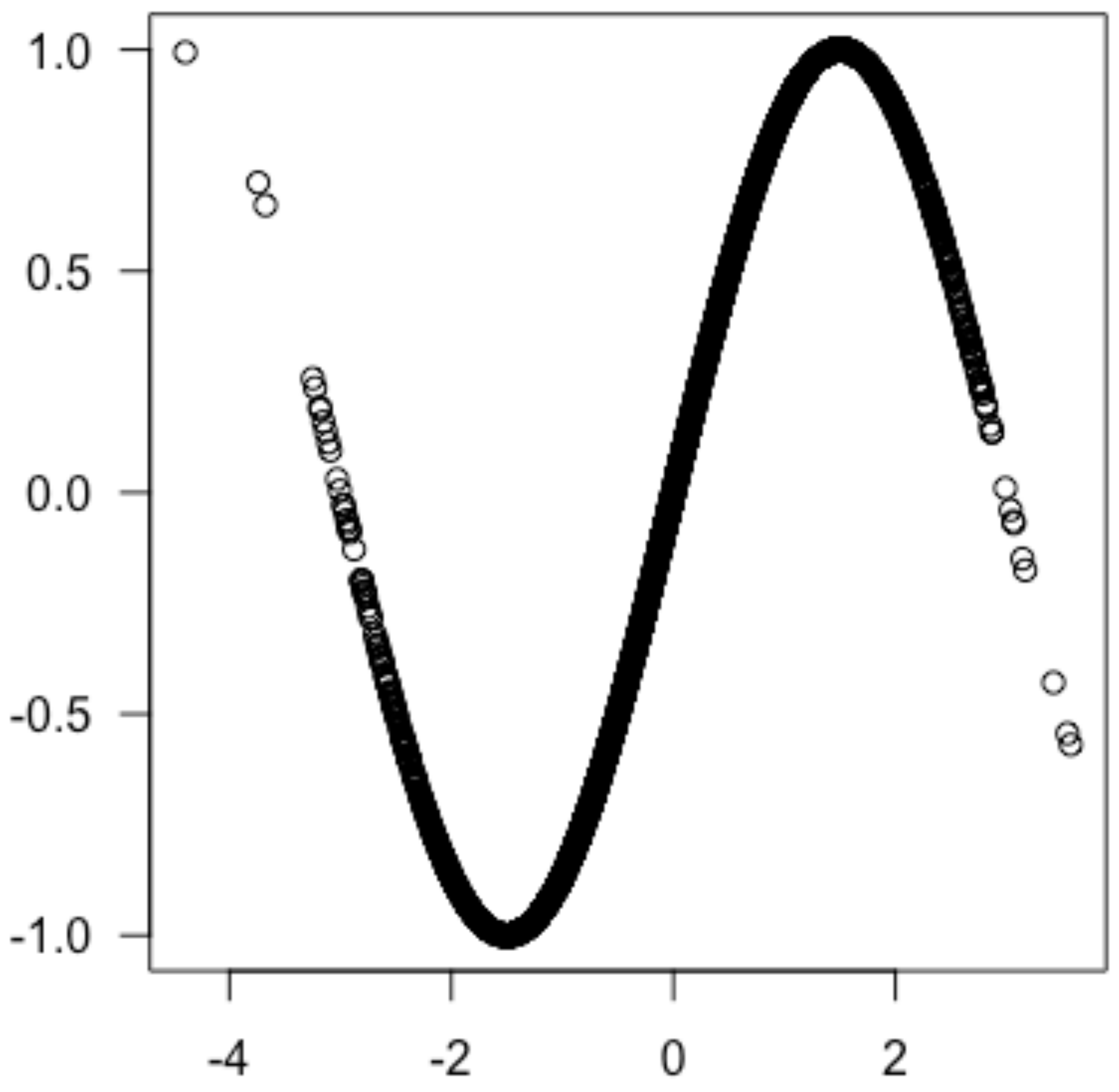
One-dimensional example: linear approximation vs constant



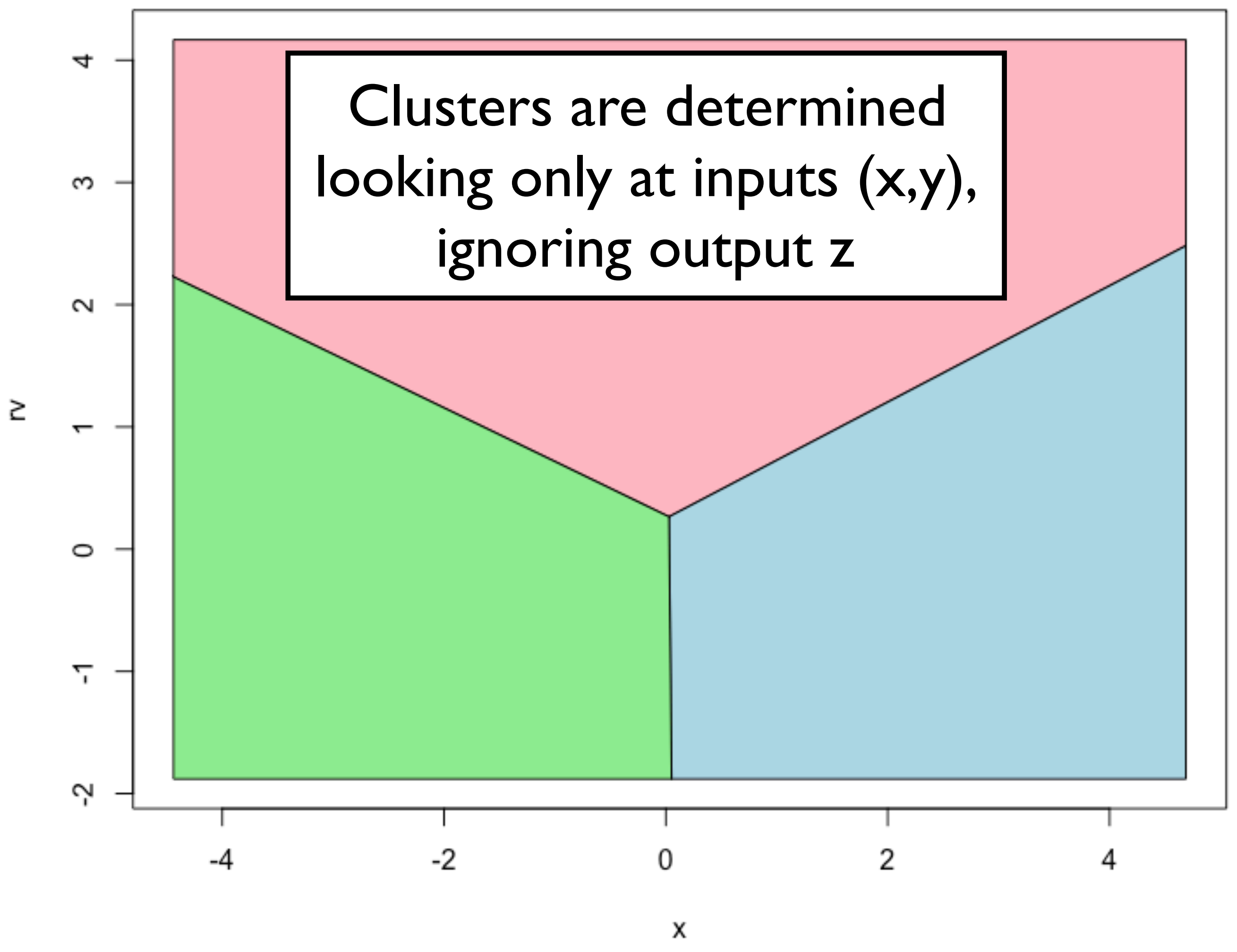


Quote imbalance

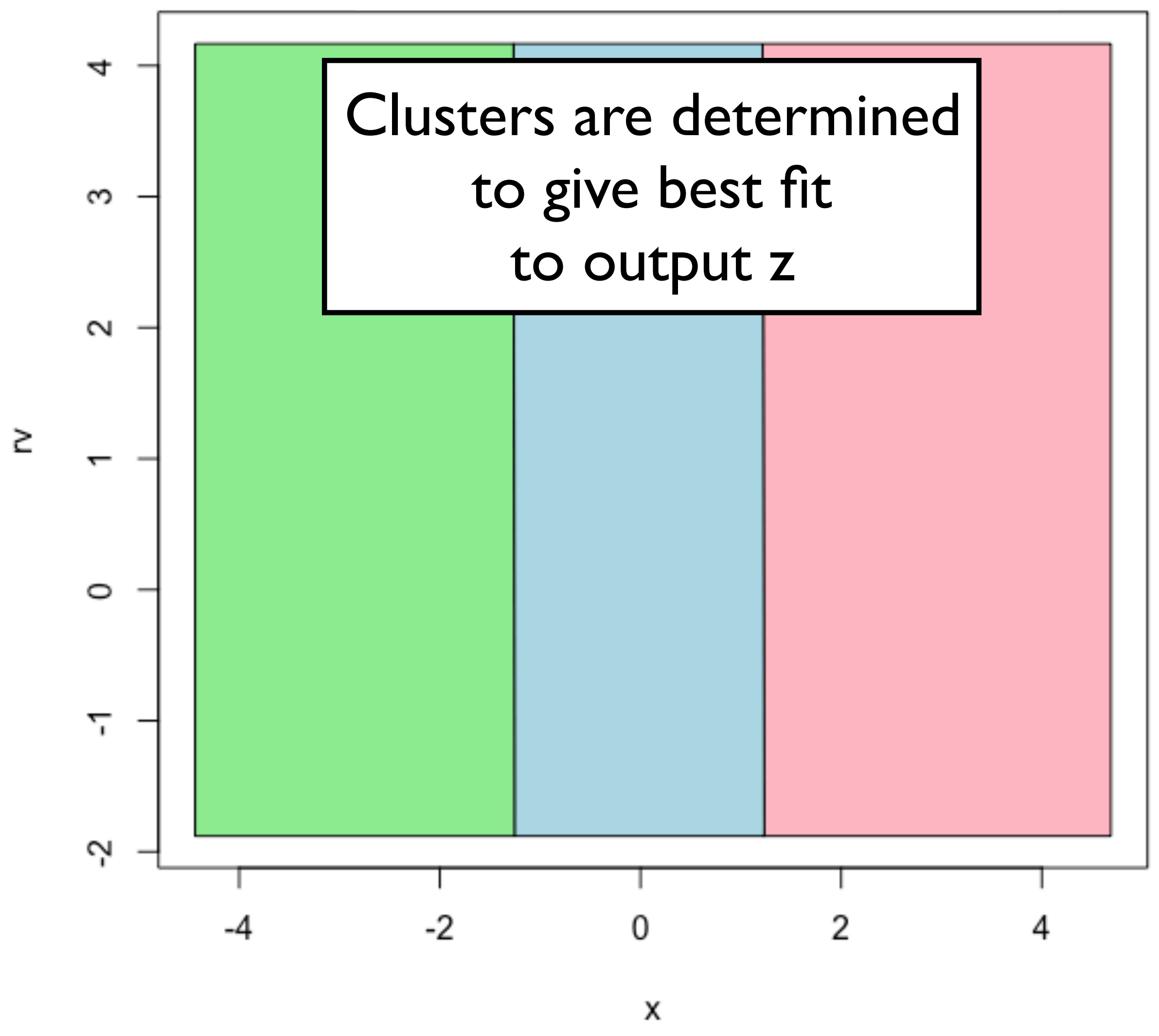
Test problem

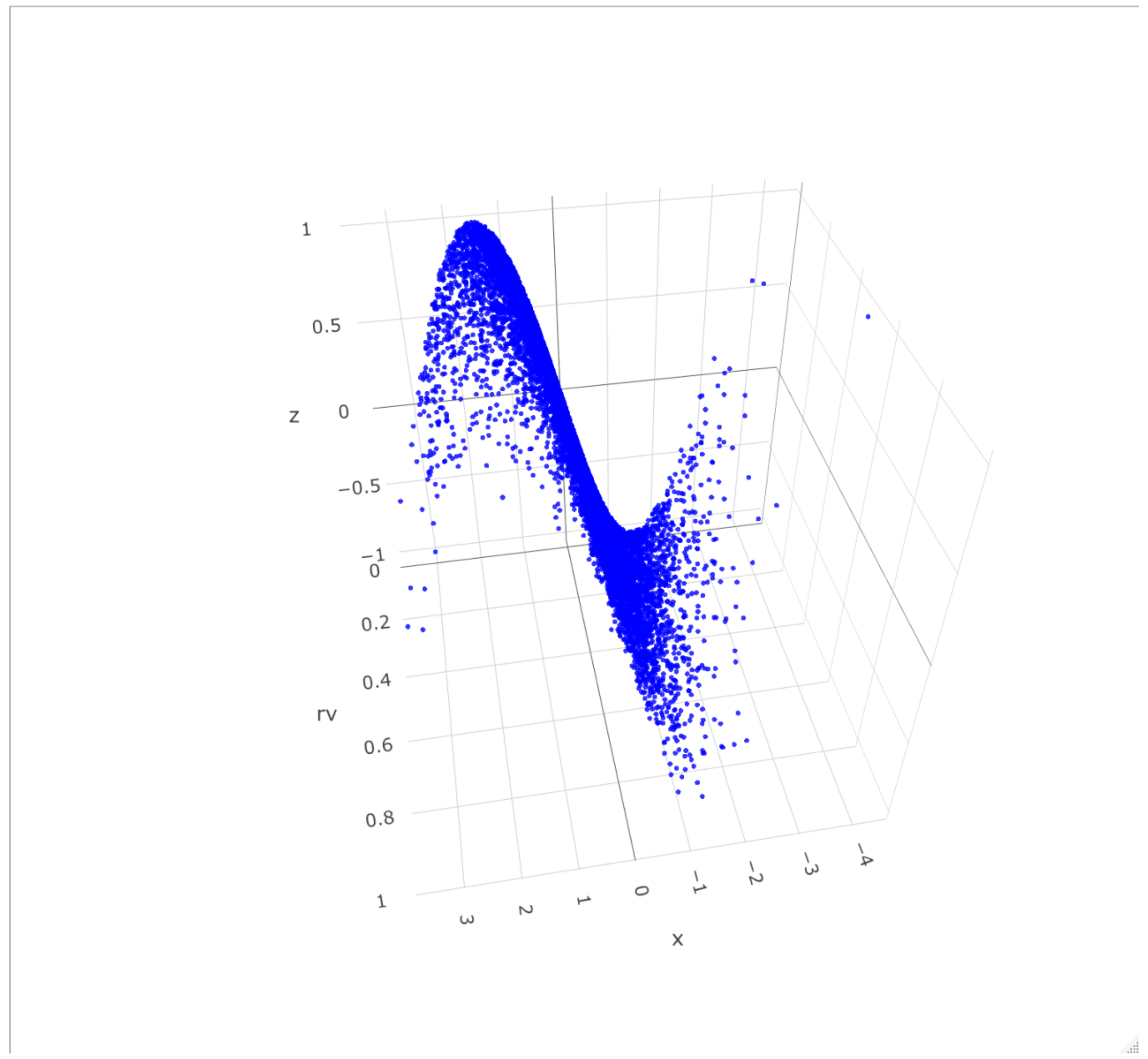


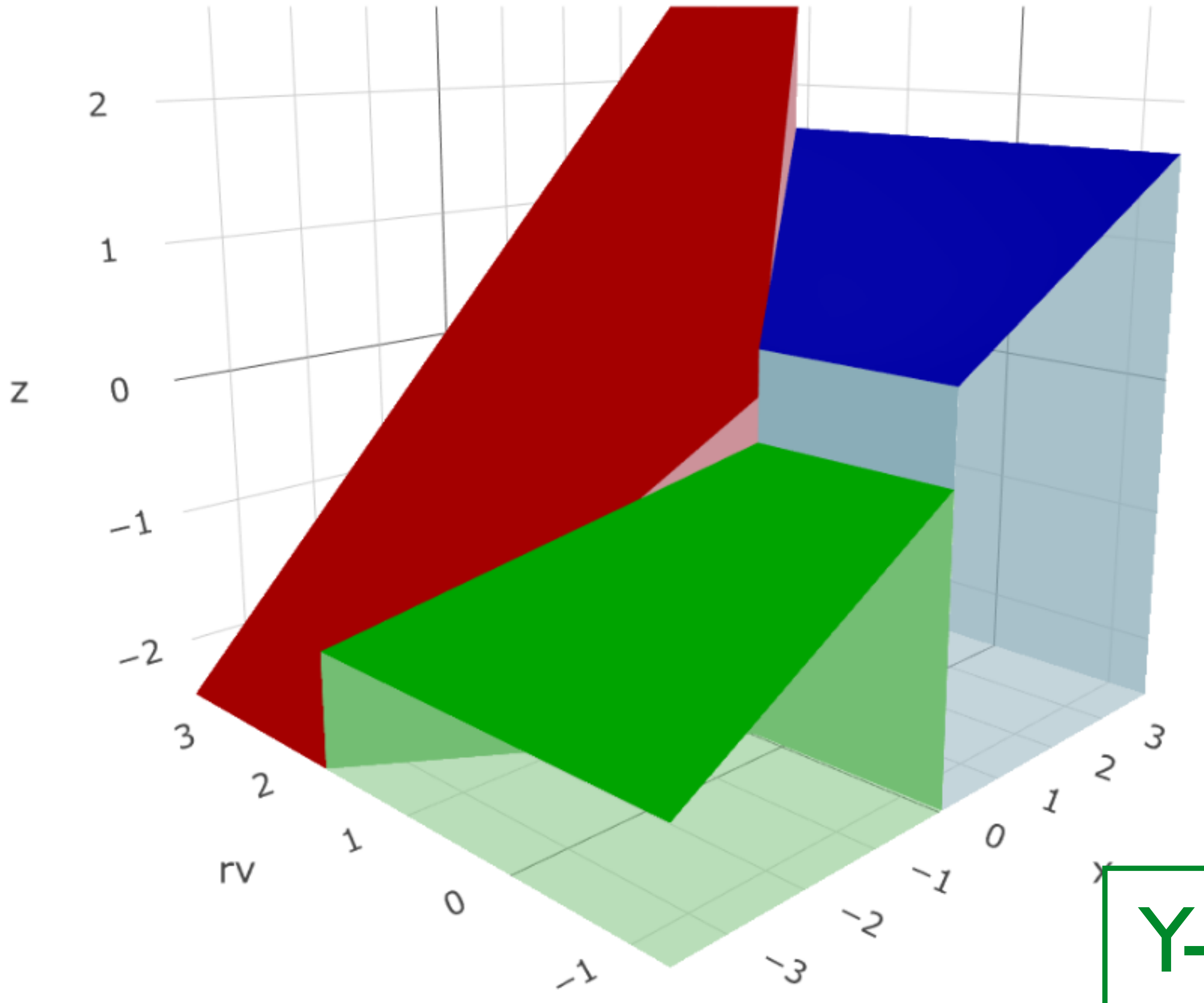
k-means

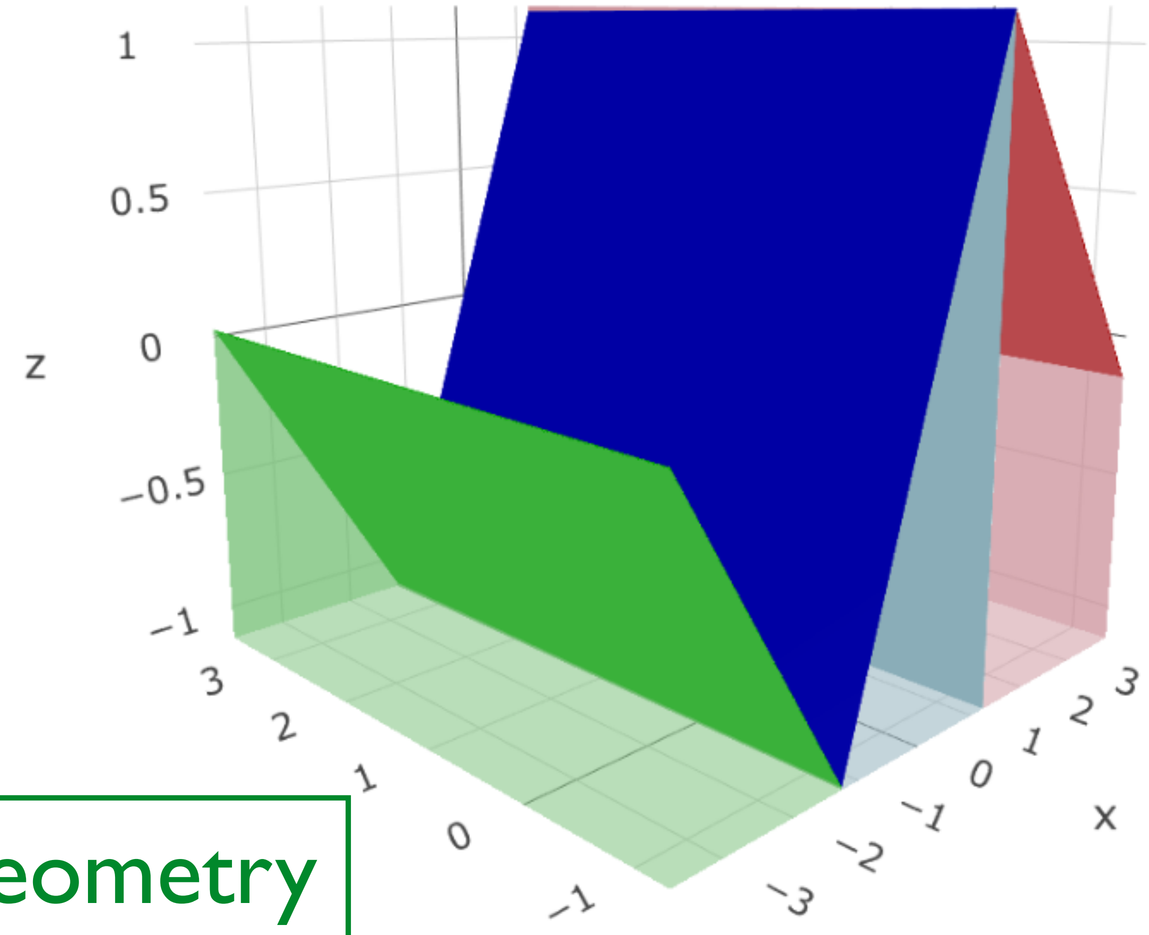


y-means



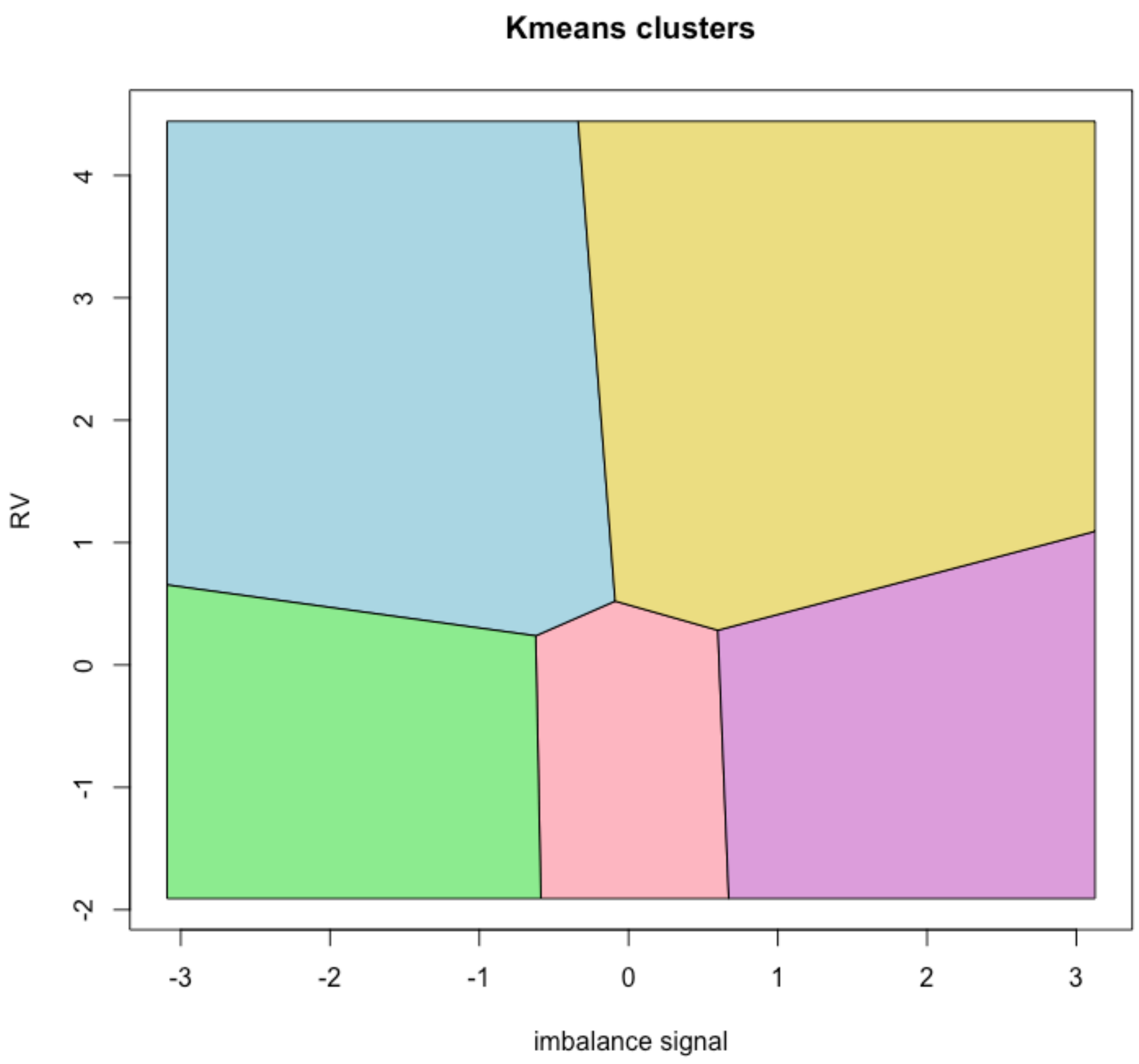




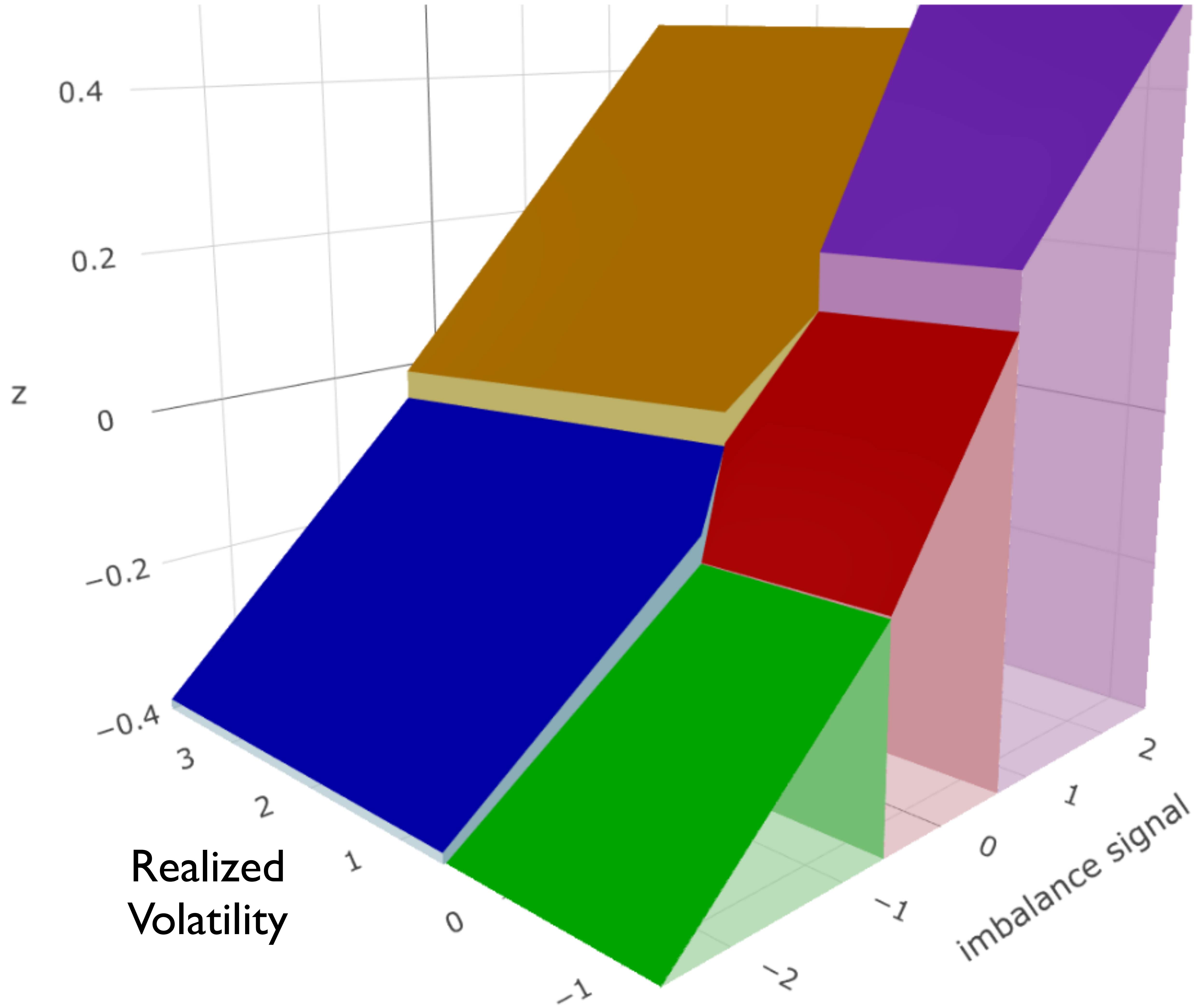


Y-means geometry
is much better
than k-means

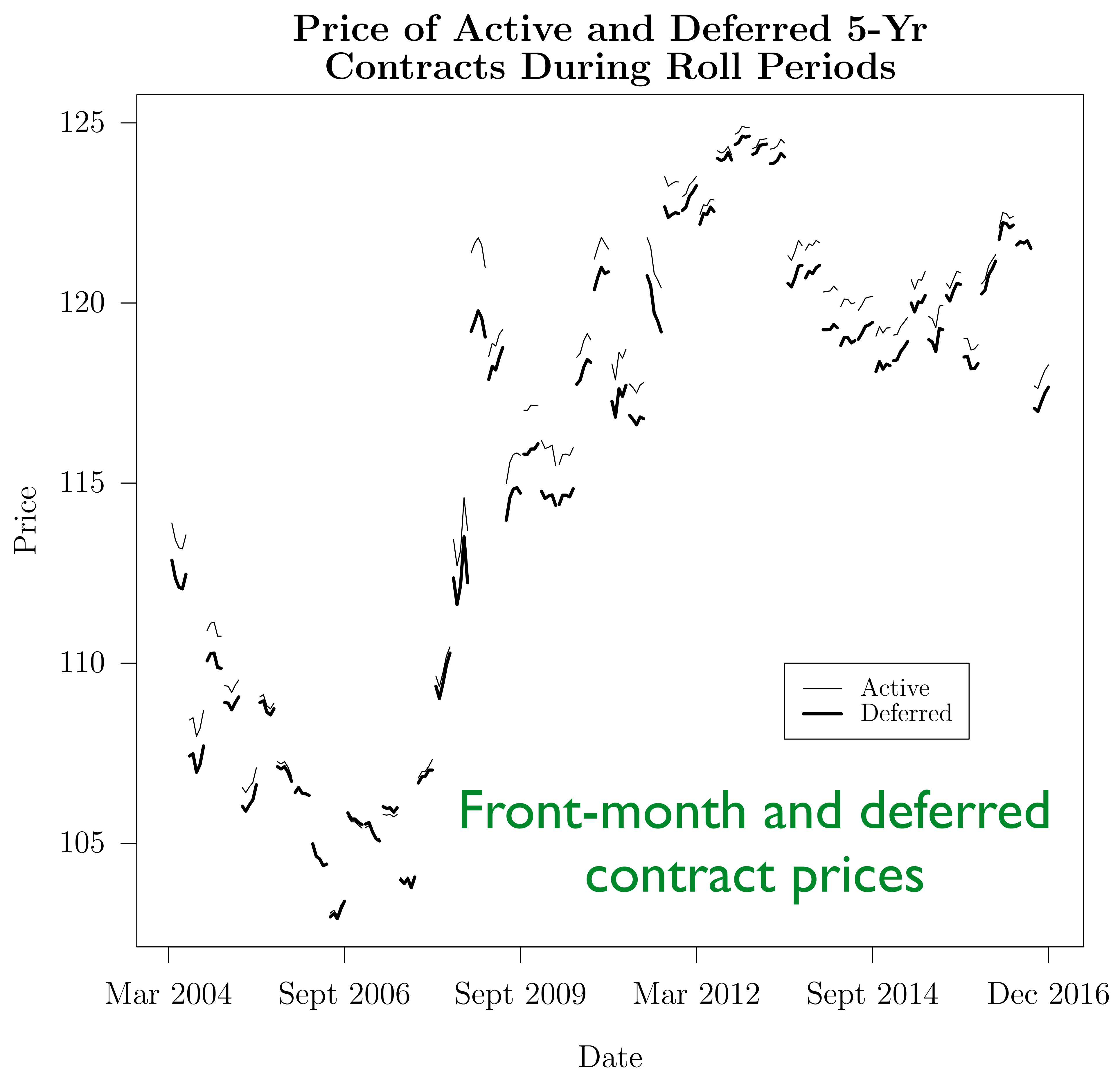
2-d example

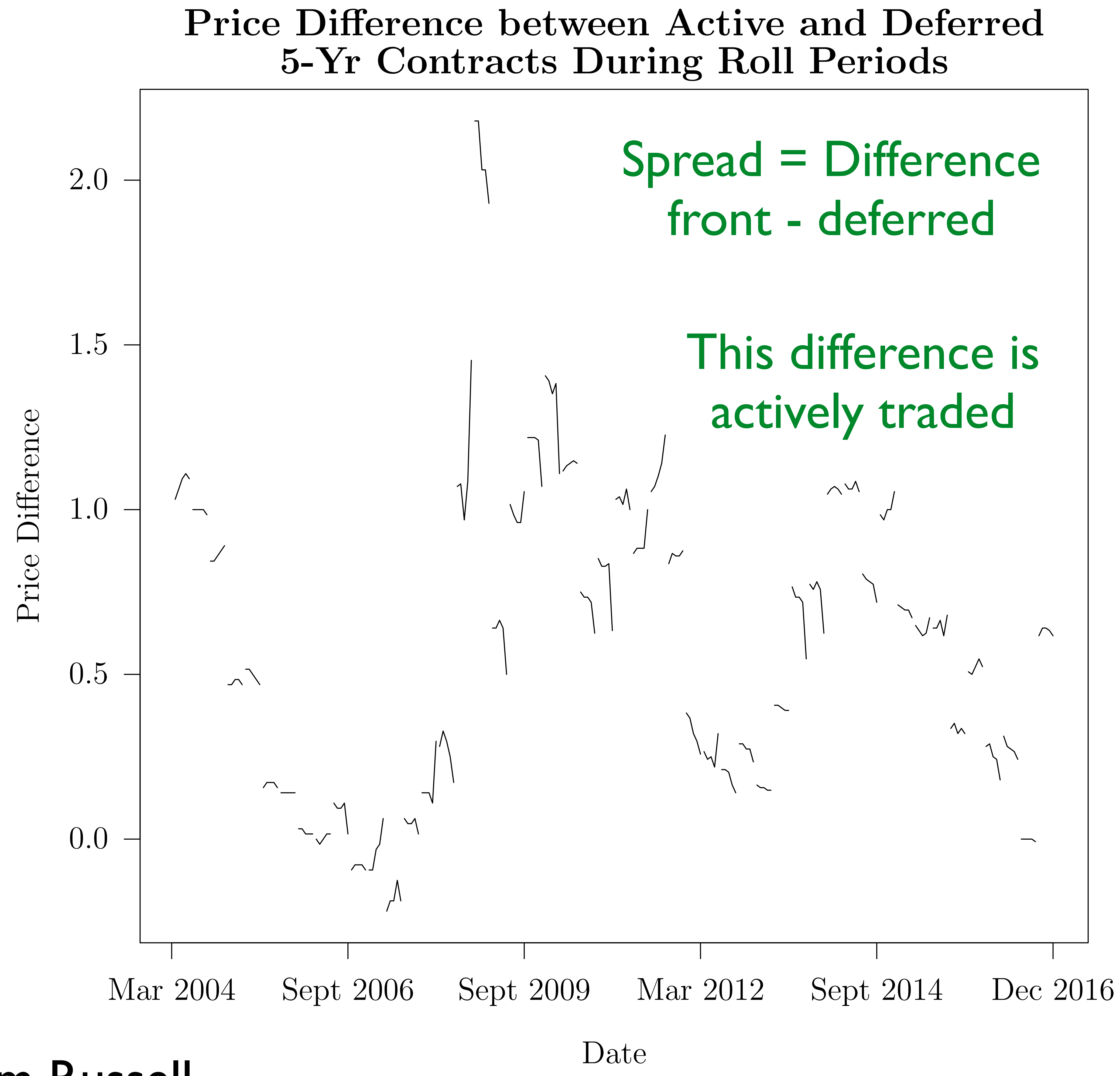


Forward return



Treasury roll forecasting





Pictures: Sam Russell

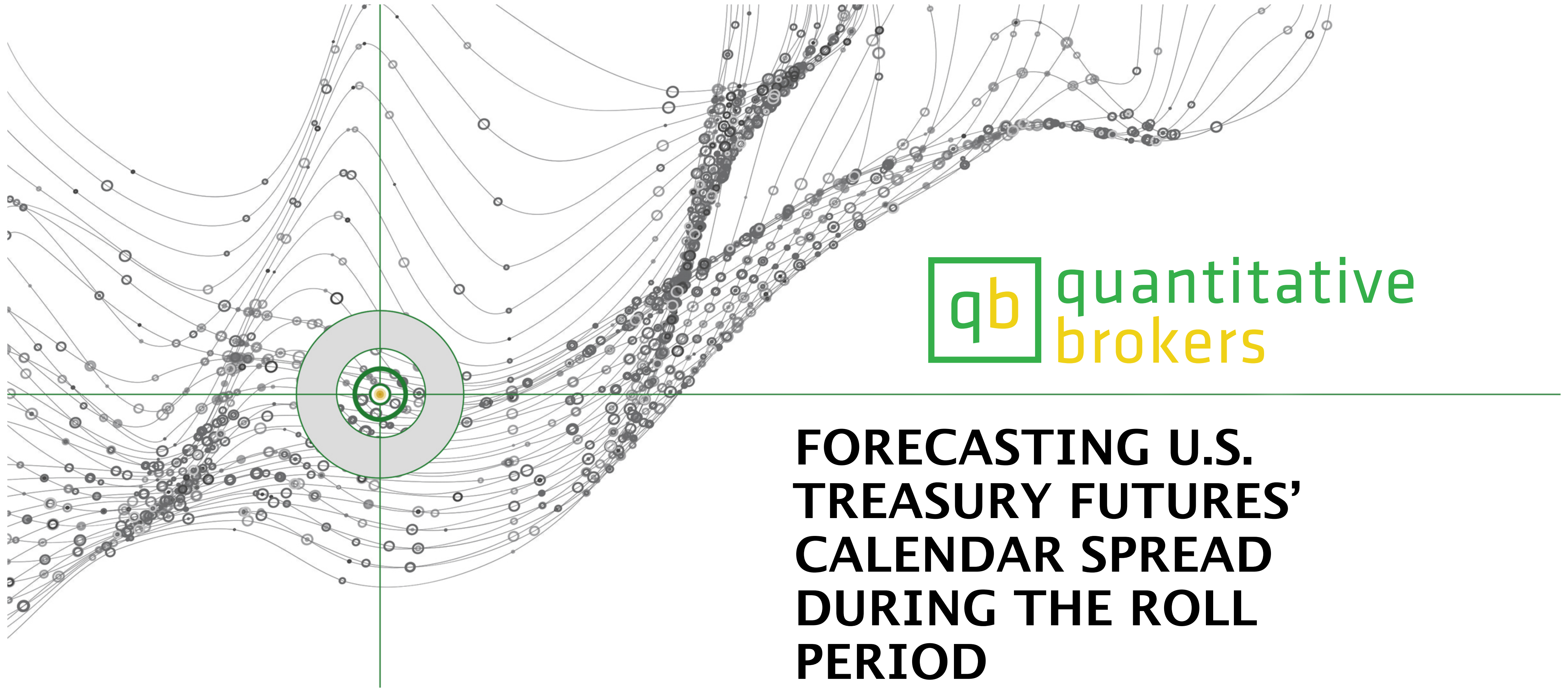
Predicting Changes in the U.S. Treasury Futures Spread During the Roll Period

Samuel Russell
Robert Almgren

June 2018

Submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Engineering Department of Operations Research and Financial Engineering

PRINCETON UNIVERSITY



FORECASTING U.S. TREASURY FUTURES' CALENDAR SPREAD DURING THE ROLL PERIOD

SHANKAR NARAYANAN
REZA GHOLIZADEH

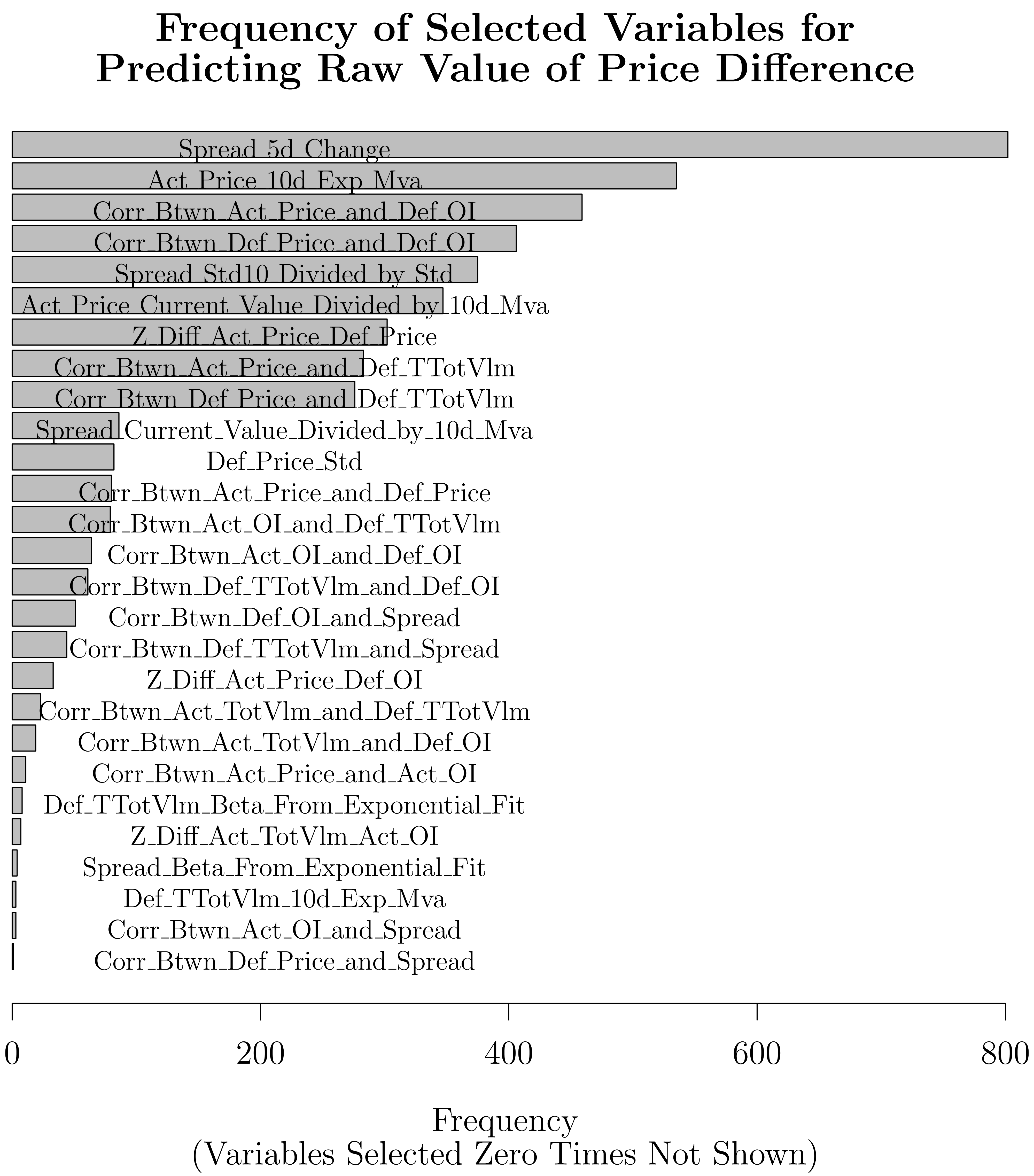
NOVEMBER 15, 2018

Sam Russell thesis

~80 features

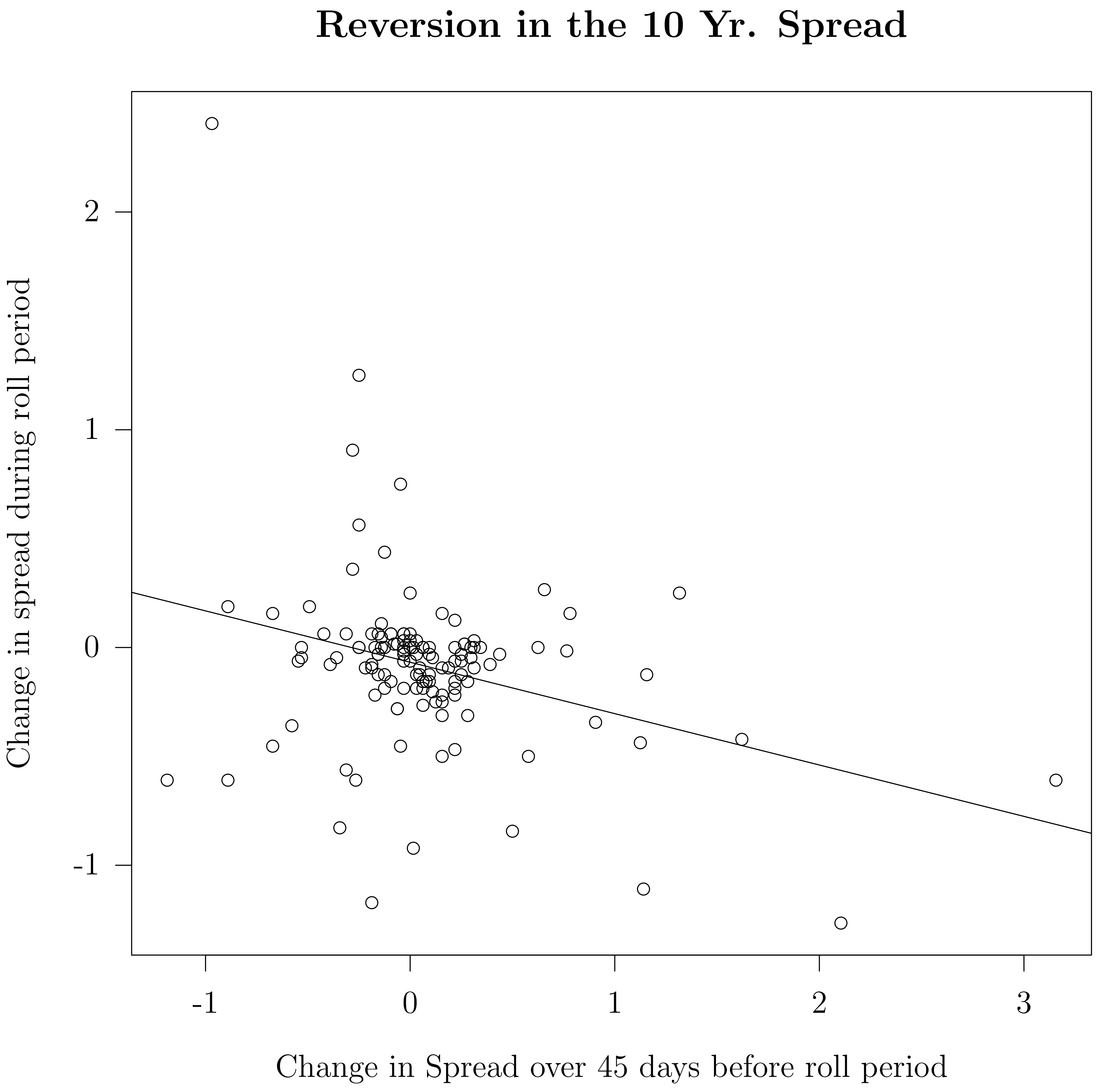
- Features of One Variable
  - Current Value
  - Standard Deviation
  - Change in value over past 5 days
  - (Standard Deviation over past 10 days) / Standard Deviation
  - Exponential moving average over past 10 days
  - (Current Value) / (Moving average over past 10 days)
  - Value of b when time series is fit to $Y = a * exp(b * X)$
- Features of Two Variables
  - Correlation
  - Difference in Z scores

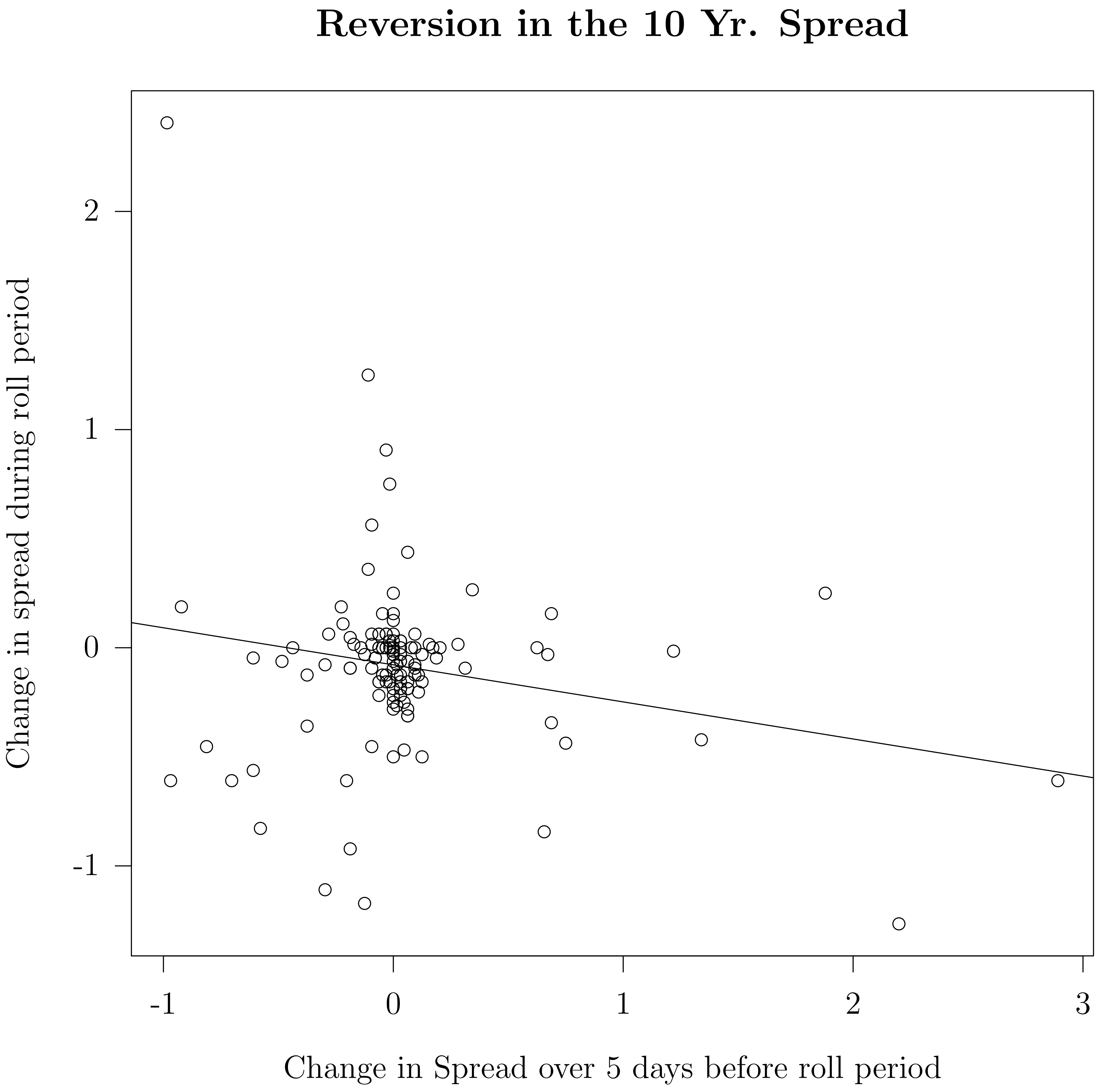
Technique: iterative regression



Sam Russell thesis

Price reversion is the single most important predictive variable





QB model

Linear predictor

$$S_{-10,0} = \alpha + \beta_1 P_1 + \beta_2 P_2 + \beta_3 P_3 + \varepsilon$$

The first predictor P1 in the multivariate model to forecast $S_{(-10,0)}$ is a reversion signal.

The second predictor P2 is obtained from the COT. The COT report is released every Friday by the CFTC and includes around 90 variables such as open interests, longs, shorts and spreads of various securities broken down by asset managers, dealers, levered funds and retail investors. (Commitments of traders)

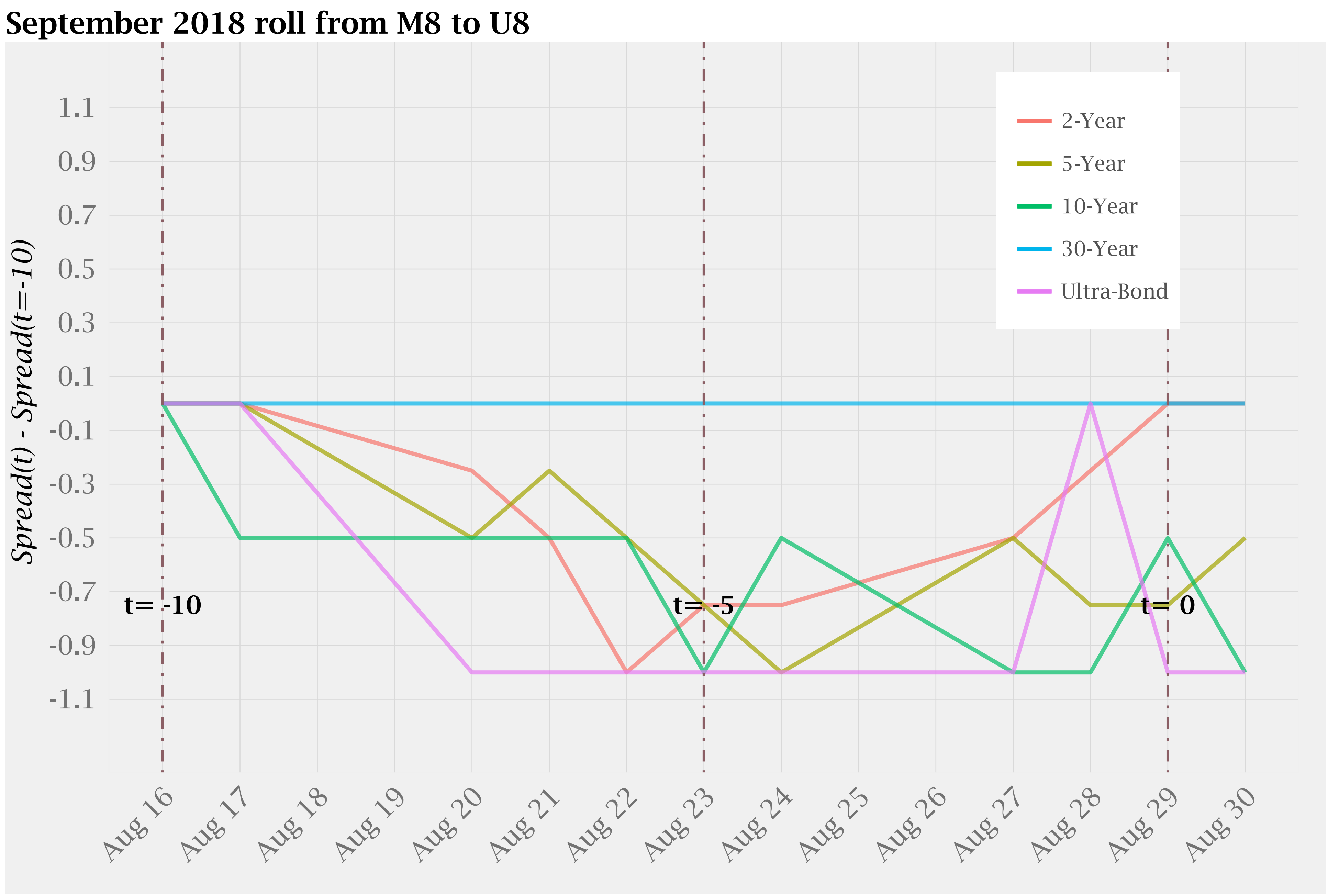
We define net position imbalance for each future as:

net imbalance = (long open interest − short open interest)/total open interest.

The third predictor P3 is $(\rho - 1)$ where ρ is the implied ratio between the near and far prices of the outrights. For illustration, Figure 3 shows the scatter plot of near vs. far

FIGURE 6

The calendar spread begins to narrow for all the futures except the 30-Year futures from around August 16^{th}, which was the tenth trading day prior to the first intention day of August 30^{th} (t=0). Our initial prediction was that the spreads would narrow. The 30-Year ended flat during the roll period but the rest ended lower from the beginning of the roll period



Conclusions

- High frequency trading is computationally demanding
- Short-term price prediction is key to performance
- Machine learning is a tool, but not automatic
- Combine ML methods with market understanding