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

SELL 40 GCZ7 BOLT

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)
Strobe: average price on interval

For Strobe, execution approximately follows volume curve, but also opportunistic when can improve performance.
Settlement price algorithm

BUY 181 ESU8 CLOSER

Exec = 2802.73  Cost to settle = -1.08 tick = -$13.54 per lot

Settlement price interval

Trades before window
Legger: multi-asset strike price

BUY 112 FGBL
SELL 129 FBTP

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?"

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

- HFT: execution algorithms
- QB: execution algorithms
- Buy-side

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
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.

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

Market data

"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

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 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

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
Price forecast for each Treasury futures

Threshold= 0.75 ticksize

Price forecasts across 10 minutes
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.

Strongest signal is when VRP is high or low but quote size is not large.
Sweep (reversion) signal

BUY 23 GCQ4 BOLT

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

Sharp motion up followed by forecast of reversion down to specific level until specific time
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:

  \[ y_t = \rho y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_p \Delta y_{t-p} + \zeta_t \]

  - \( M_0 : \rho = 1 \)
  - \( M_1 : \rho > 1 \)

- When \( \rho > 1 \) 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

Implied volatility

Implied volatility

Strike price (moneyness)

Strike price (moneyness)
Prices from option models
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

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

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

Calendar spreads are 1:1, so prices just add and subtract: prices are always on grid.
User Defined Spreads of OZNX9 Put Options on 2019−10−03
Implied price compared with direct

Out-of-the-money: relatively thick book

Jobless Claims

CDT on Thu 01 Aug 2019
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.
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.

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.

Order is filled when queue depletes

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

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

where

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

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 $T$. 
5.1 Basic features
\[ F_1 = \{ P_{\text{ask}}^k - P_{\text{bid}}^k, Q_{\text{ask}}^k, Q_{\text{bid}}^k \}_{k=1}^K \]

5.2 Time-insensitive set
\[ F_2 = \{ P_{\text{ask}}^k - P_{\text{bid}}^k, \frac{P_{\text{ask}}^k + P_{\text{bid}}^k}{2}, \frac{P_{\text{ask}}^k + Q_{\text{ask}}^k}{Q_{\text{bid}}^k + Q_{\text{ask}}^k}, Q_{\text{bid}}^k, Q_{\text{ask}}^k \}_{k=1}^K \]

5.3 Time-sensitive set
Denote \( t = 1, 2, \ldots, s \) as the number of look-back window
\[ \Delta w = 60s, \text{denote } F_3 = \{ f_31, f_32, f_33, f_34 \}, \] where

5.4 Time-dependent set
Denote \( t = 1, 2, \ldots, s \) as the number of look-back period with look-back window
\[ \Delta w = 60s, \text{denote } F_4 = \{ f_41, f_42, f_43, f_44 \}, \] where
\[ f_41 = \{ P_{\text{ask}}^k - P_{\text{bid}}^k, \max_{l \in \text{levels}} (P_{\text{ask}}^k), \max_{l \in \text{levels}} (P_{\text{bid}}^k), V_{\text{ask}}^k, V_{\text{bid}}^k, V_{\text{ask}}^k \}_{k=1}^K \]
\[ f_42 = \{ Q_{\text{ask}}^k, Q_{\text{bid}}^k \}_{k=1}^K \]
\[ f_43 = \{ 1 \cdot \text{TV}_{\text{ask}}^k, 1 \cdot \text{TV}_{\text{bid}}^k \}_{k=1}^K \]
\[ f_44 = \{ \delta b, \delta m \}_{k=1}^K \]
Smart Order Routing

**MACHINE LEARNING FOR LIMIT-ORDER ROUTING IN CASH TREASURY MARKETS**

RENYUAN XU  
ISAAC CARRUTHERS  
APRIL 25, 2018

**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, \( x \in \mathbb{R}^d \)
\( x \) = signal outputs, and market state, \( d \sim 10-15 \)
\( y \) = 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
K-means

Determine clusters based on distribution of \( x \) (ignoring \( y \))
Fit a constant function in each cluster
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
\[ F_k(x) = \hat{y}_j + \beta'_j(x - \bar{x}_k), \quad \text{for } x \in C_k \]

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

\[ C_k, \ldots, C_K = \text{Voronoi cells} \]

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

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.

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 modified Monte Carlo method has been used for the present study to compute properties of a two-dimensional rigid-sphere system. The method is similar to that described by Metropolis et al. (1953) and has been found to be more efficient in practice. The present calculations have been carried out on the Los Alamos MANIAC computer, and the results are presented here.

Multiple random starting points
Keep best configuration to date on each trajectory

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

Figure 14: 1D clustering and fitting results. The black line is the benchmark where the clusters are defined as equally spaced intervals of x axis; The blue line is the result with mean value as prediction in the objective function; The red line is the result with linear regression value in each cluster as prediction in the objective function. (a) use 'oi' as x. (b) use argsiz. (c) The bigger points are center of the equally spaced bins as in (a) and small points are the data points colored by their corresponding bin index. This is just an illustration that the predictive trend is too small compared to the variance of data, which may explain why we see the value of objective function does not improve much whatever methods we try.

Linear approximation is much better than constant.
Test problem

- **k-means**: Clusters are determined looking only at inputs \((x,y)\), ignoring output \(z\).
- **y-means**: Clusters are determined to give best fit to output \(z\).

Y-means geometry is much better than k-means.
2-d example

Kmeans clusters

Forward return

Realized Volatility
Treasury roll forecasting

Price of Active and Deferred 5-Yr Contracts During Roll Periods

Price Difference between Active and Deferred 5-Yr Contracts During Roll Periods

Spread = Difference front - deferred

This difference is actively traded

Front-month and deferred contract prices

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
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 \( Y = a \times exp(b \times X) \)

- Features of Two Variables
  - Correlation
  - Difference in Z scores

Technique: iterative regression
When is Reversion Stronger

In other problems involving the prediction of asset price movements, volume is seen as important confirmation of whether a trend is real or not. In general, many signals are said to be more trustworthy when volume is high, as opposed to when volume is low, as relative to volume’s historical averages (Investopedia LLC, 2018). The volume of treasury futures contracts is also

Reversion over Different Time Series

The first half of this paper has brought up the idea of reversion being a significant predictor of changes in treasury futures spread, and the last section solidified its presence with respect to the 10 year U.S. treasury futures contract, while throwing it out for other contracts. Up until now, we have been considering reversion as applied to the change in the spread in the 5 days prior to the start of the roll period. This section will observe other possible time frames.

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 \( P_1 \) in the multivariate model to forecast \( S_{-10,0} \) is a reversion signal.

The second predictor \( P_2 \) 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:

\[ \text{net imbalance} = (\text{long open interest} - \text{short open interest}) / \text{total open interest}. \]

The third predictor \( P_3 \) is \( (\rho - 1) \) where \( \rho \) 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 16th, which was the tenth trading day prior to the first intention day of August 30th (10). 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