Statistical Arbitrage Using Limit-Order Book Imbalance

Anton Rubisov
University of Toronto

17 September 2015

Background Information

Exploratory Data Analysis

Modelling Imbalance

Incorporating Price Change

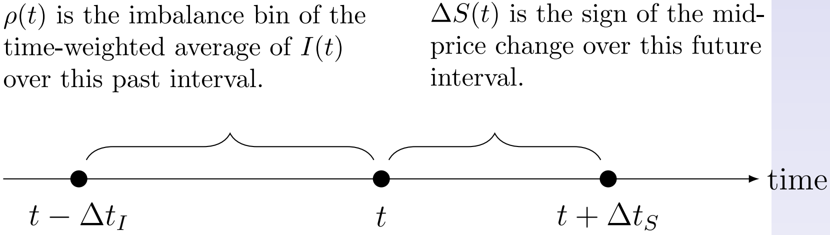
Next, consider a two-dimensional CTMC $Z(t)$ that jointly models imbalance bin $\rho(t)$ and price change $\Delta S(t)$, where

$$\rho(t) \in \{1, 2, \ldots, \#_{bins}\}$$

is the bin corresponding to imbalance averaged over the interval $[t - \Delta t_I, t]$, and

$$\Delta S(t) = \operatorname{sgn}(S(t + \Delta t_S) - S(t)) \in \{-1, 0, 1\}$$

is the *sign* of the change in midprice of the *future* time interval Δt_S.



Transition Matrix

Using MLE, we obtain a generator matrix $\mathbf{G}$ for the CTMC. The transition matrix over a step of size Δt_I is given by

$$\mathbf{P}(\Delta t_I) = e^{\mathbf{G}\Delta t_I}$$

called our *one-step transition probability matrix*. Matrix entries $p_{ij}(\Delta t_I)$ give the probability of transition from one (imbalance, price change) pair to another over the time interval Δt_I. This can be written semantically as

$$P_{ij} = \mathbb{P}\left[\varphi(\rho_{\text{curr}}, \Delta S_{\text{future}}) = j \mid \varphi(\rho_{\text{prev}}, \Delta S_{\text{curr}}) = i\right]$$

Predicting Future Price Change

Using Bayes' Rule, we can transform the **P** matrix to

$$\mathbb{P}\left[\Delta S_{\text{future}} = j \mid B, \rho_{\text{curr}} = i\right] = \frac{\mathbb{P}\left[\rho_{\text{curr}} = i, \Delta S_{\text{future}} = j \mid B\right]}{\mathbb{P}\left[\rho_{\text{curr}} = i \mid B\right]}$$

This allows us to predict future price moves.
We'll call the collection of these probabilities the **Q** matrix.

Predicting Future Price Change

	$\Delta S_{\text{curr}} < 0$			$\Delta S_{\text{curr}} = 0$			$\Delta S_{\text{curr}} >$	
$\rho_{curr} = 1$		2	3	1	2	3	1	2
$\Delta S_{\text{future}} < 0$								
$\rho_{\text{prev}} = 1$	**0.53**	0.15	0.12	0.05	0.10	0.14	0.08	0.13
$\rho_{\text{prev}} = 2$	0.10	**0.58**	0.14	0.07	0.04	0.10	0.13	0.06
$\rho_{\text{prev}} = 3$	0.08	0.12	**0.52**	0.09	0.06	0.03	0.11	0.10
$\Delta S_{\text{future}} = 0$								
$\rho_{\text{prev}} = 1$	0.41	0.75	0.78	**0.91**	0.84	0.79	0.42	0.79
$\rho_{\text{prev}} = 2$	0.79	0.36	0.71	0.83	**0.92**	0.82	0.75	0.37
$\rho_{\text{prev}} = 3$	0.79	0.74	0.40	0.81	0.83	**0.91**	0.70	0.76
$\Delta S_{\text{future}} > 0$								
$\rho_{\text{prev}} = 1$	0.06	0.10	0.09	0.04	0.06	0.07	**0.50**	0.09

Trading Strategies Informed by the **Q** Matrix

Naive Use market orders to buy (sell) if price is predicted to move up (down).

Naive+ Post at-the-touch limit orders when zero price change is predicted.

Naive++ Post a limit order to buy (sell) is price is predicted to move up (down).

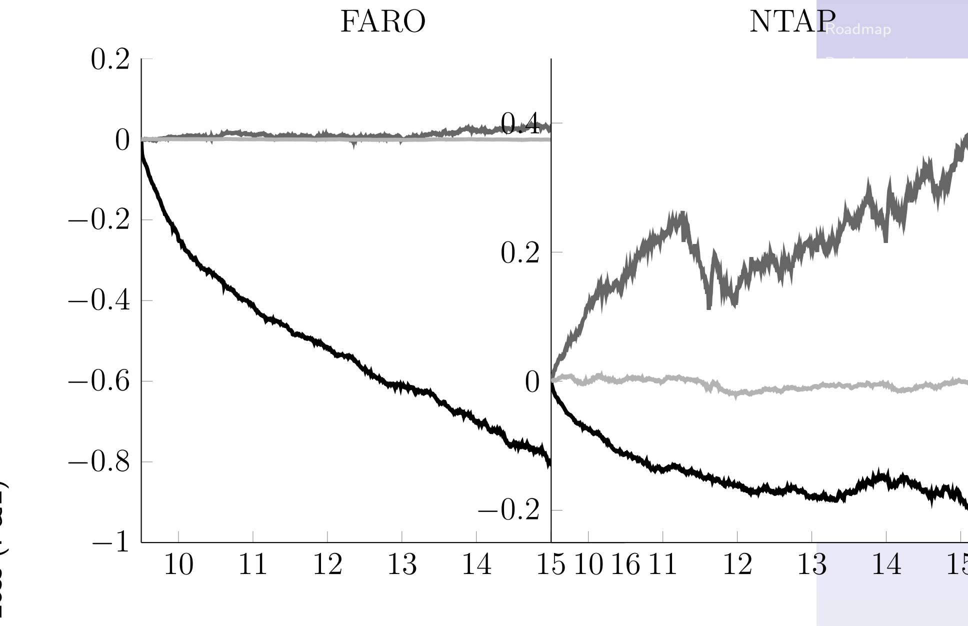
Calibration of Naive Trading Strategies

Need to select:

- price change observation period Δt_S
- imbalance averaging period Δt_I
- number of imbalance bins $\#_{bins}$

Calibration done on the first day of the trading year, same parameters used for all days.
Brute-force search of parameter space, using max Sharpe ratio criterion, found that $\Delta t_S = \Delta t_I = 1sec$, and $\#_{bins} = 4$

Results of Naive Trading Strategies



Conclusions from Naive Trading Strategies

Why is the Naive strategy producing, on average, normalized losses?

- Backtest is out-of-sample; evidence to reject time-homogeneity
- Calibration is done on first trading day; likely nonrepresentative of trading activity
- Price change probability matrix $\mathbf{Q}$ obtained using midprices, ignoring bid-ask spread; $\operatorname{sgn}(\Delta S)$ may be insufficient for create profit, especially on `FARO`

Conclusions from Naive Trading Strategies

Why do the Naive+ and Naive++ strategies outperform the Naive strategy?

- LOs vs MOs means different transaction price is being used (only MO loses value)
- Naive only executes when predicting non-zero price change
  - Only sign, not magnitude
  - Only *if one was already seen*