A statistical model for detecting hidden liquidity

Nikolaus Hautsch and Ruihong Huang
Quantitative Products Laboratory
Humboldt-Universität zu Berlin

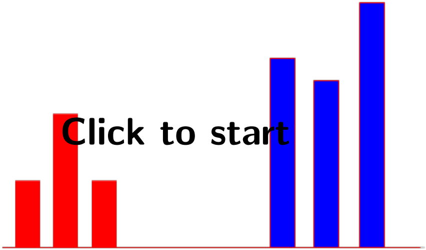
Modern electronic limit order markets

Electronic limit order books are central to financial markets:

- Modern Markets
  e.g. Euronext, XETRA, Spanish Stock Exchange, Toronto Stock Exchange, etc.
- NYSE and NASDAQ enhanced their trading protocols by:
  - NASDAQ – Single Book (previously INET and Brut)
  - NYSE – ARCA (previously Archipelago Exchange)

  Competition for liquidity motivates the undisclosed orders in these markets.

Hidden type orders



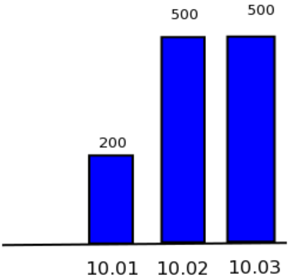
- Reserve order (or "Iceberg" order): visible part is replenished after (fully) executed
- Non-display orders: no visible part

The invisible part loses its time priority.

Business perspective

The cost of buying 1000 shares:

- without hidden liquidity
  - Price: \$10.021
  - Immediate price impact: 2 ticks
- with hidden liquidity
  - Price: \$10.009
  - Immediate price impact: 1 tick



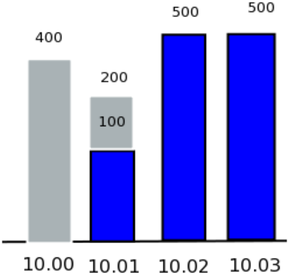
By submitting the market order in the “right” time, you can:

- save transaction cost and
- reduce market impact.

Business perspective

The cost of buying 1000 shares:

- without hidden liquidity
  - Price: \$10.021
  - Immediate price impact: 2 ticks
- with hidden liquidity
  - Price: \$10.009
  - Immediate price impact: 1 tick



By submitting the market order in the “right” time, you can:

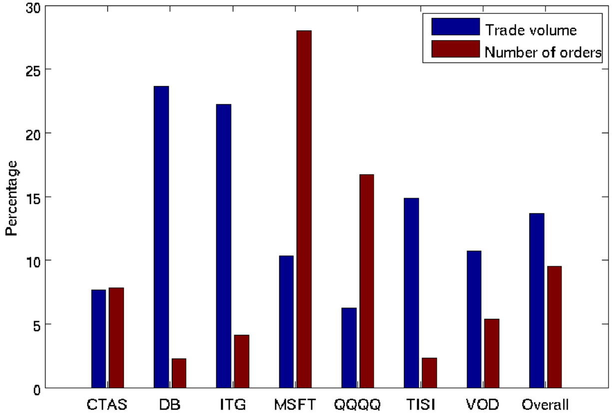
- save transaction cost and
- reduce market impact.

How much liquidity is hidden?

- Approximately 28% of the volume on Australian Stock Exchange was hidden in 1993.(Aitken, Berkman and Mak,2001)
- Hidden orders account for more than 12% of all orders executed on Island.(Hasbrouck and Saar,2002)
- Hidden depth on Euronext Paris accounts for 45% of the total depth available at the best five quotes.(D'Hondt, De Winne and Francois-Heude,2004):
- Hidden liquidity represents 20% of the inside depth in the Nasdaq 100 stocks in 2002(Tuttle, 2005)
- In Xetra 9% of non-marketable orders are "iceberg"s. (Frey and Sandas, 2009)

Hidden liquidity in 7 stocks in NASDAQ

Period: Jan 2 – 31, 2009



Formulate of research problem

How can we discover hidden liquidity?

Currently,

- "Iceberg" detection algorithm (Frey and Sandas, 2009)
- "Pinging" the pools by "fleeting" orders (Industry and Hasbrouck and Saar, 2009)

Our approach: A statistical model!

Formulate of research problem

How can we discover hidden liquidity?

Currently,

- "Iceberg" detection algorithm (Frey and Sandas, 2009)
- "Pinging" the pools by "fleeting" orders (Industry and Hasbrouck and Saar, 2009)

Our approach: A statistical model!

Outline

Visible order flow

The visible order flow data records all visible limit order activities (e.g. submission, cancellation) and all trades. Advantages:

- Much cheaper than hidden liquidity data;
- Public available in real time;
- Allows us to reconstruct the visible part of order book.

Disadvantages:

- Record only executions of hidden liquidity via market (marketable) orders.

We need more sophisticated algorithm and statistical model for detecting it.

Signal of hidden liquidity

Which type of orders can signal hidden liquidity?

- All limit orders placed inside of the spread
- All marketable orders when the spread is at least 2 ticks
- Marketable orders with sizes greater than visible volumes at the corresponding best price, when the spread is 1 tick

Signal of hidden liquidity

Which type of orders can signal hidden liquidity?

- All limit orders placed inside of the spread
- All marketable orders when the spread is at least 2 ticks
- Marketable orders with sizes greater than visible volumes at the corresponding best price, when the spread is 1 tick

Signal of hidden liquidity

Which type of orders can signal hidden liquidity?

- All limit orders placed inside of the spread
- All marketable orders when the spread is at least 2 ticks
- Marketable orders with sizes greater than visible volumes at the corresponding best price, when the spread is 1 tick

Logit Regression

p_i is the probability of an order (with one of previously mentioned types) executed with hidden liquidity.

$$\begin{aligned} \text{logit}\, p_i = \log \frac{p_i}{1-p_i} = x_i'\beta = \beta_0 &+ \beta_1 \begin{pmatrix} \text{Rel. location} \\ \text{to best price}_i \end{pmatrix} + \beta_2 \begin{pmatrix} \text{log} \\ \text{spread}_i \end{pmatrix} \\ &+ \beta_3 \begin{pmatrix} 1^{st}\text{-level} \\ \text{Imalance}_i \end{pmatrix} + \beta_4 \begin{pmatrix} 2^{nd}\text{-level} \\ \text{Imbalance}_i \end{pmatrix} + \beta_5 \begin{pmatrix} \text{lagged} \\ \text{volume}_i \end{pmatrix} \\ &+ \beta_6 \begin{vmatrix} \text{lagged} \\ \text{return}_i \end{vmatrix} + \beta_7 d_i^{\text{first hour}} + \beta_8 d_i^{\text{last hour}} \end{aligned}$$

Logit Regression (continue...)

Rel. Location : |order position − best price in same side|

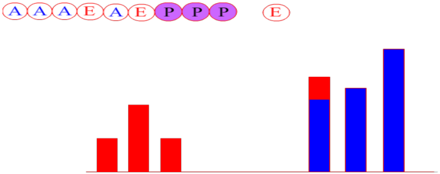
$n-$level imbalance: $\log\left(\sum_{k=1}^{n} \text{Depth at } k \text{ level in the opposite side}\right) - \log\left(\sum_{k=1}^{n} \text{Depth at } k \text{ level in the same side}\right), \; n = 1, 2$

Volume : 5 minute cummulated trade size

Return : 5 minute return

first hour : 9:35 – 10:30

last hour : 15:00 – 15:55



Bayesian Treatment

- Prior distribution.

$$\beta \sim \mathcal{N}(0, \alpha^{-1}\mathbf{I})$$

- likelihood function

$$\pi(y|x,\beta) = \prod_{i=1}^{n} \left[\Lambda(x_i'\beta)^{y_i}(1-\Lambda(x_i'\beta))^{1-y_i}\right]$$

where $y = (y_1, \ldots, y_n)'$ is our observations. $\Lambda(\cdot)$ is the logistic function

- posterior distribution

$$\pi(\beta|y,x) \propto \exp\left(\frac{\alpha}{2}\beta'\beta\right) \times \prod_{i=1}^{n} \left[\Lambda(x_i'\beta))^{y_i}(1-\Lambda(x_i'\beta)^{1-y_i}\right]$$

Approximation of posterior distribution

The log posterior and its second derivative are

$$\ln \pi(\beta|y,x) = \frac{\alpha}{2}\beta'\beta + \sum_{i=1}^{n}\left[y_i\Lambda(x_i'\beta) + (1-y_i)(1-\Lambda(x_i'\beta)\right] \quad (1)$$

$$\Sigma = -\nabla\nabla \ln \pi(\beta|y,x) = \alpha + \sum_{i=1}^{n}\Lambda(x_i'\beta)(1-\Lambda(x_i'\beta))x_i x_i' \quad (2)$$

By Laplace approximation, the posterior distribution can be approximated by

$$\mathcal{N}(\widehat{\beta}_{\mathrm{MAP}}, \widehat{\Sigma}_{\mathrm{MAP}}).$$

The $\widehat{\beta}_{\mathrm{MAP}}$ is the *maximum posterior estimator* which can be found by *iterative reweighted least squares (IRLS)* algorithm.

Maginal Effects

The marginal effect is

$$\gamma = \frac{\partial \operatorname{E}[y|x]}{\partial x}$$

and can be estimated by

$$\widehat{\gamma} = \Lambda(\bar{x}'\hat{\beta})[1 - \Lambda(\bar{x}'\hat{\beta})]\hat{\beta}$$

Where $\bar{x}$ is the sample mean. And

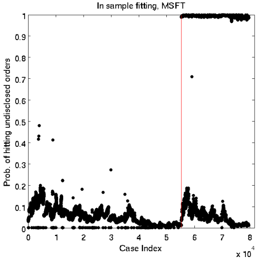
$$\text{Asy.Var}(\gamma) = [\Lambda(1-\Lambda)]^2[\mathbf{I} + (1-2\Lambda)\beta x']\Sigma[\mathbf{I} + (1-2\Lambda)x\beta']$$

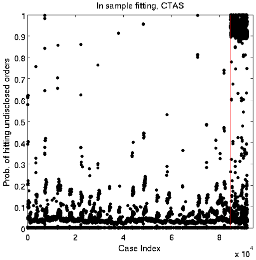
Summary Statistics

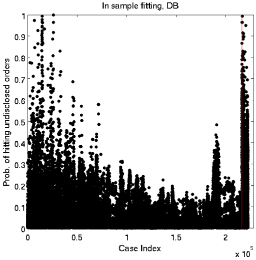
	#hidden	#order	Rel. Loc.			Spr. ($\times 10^{-3}$)	
			mean	max	min	mean	std
CTAS	6971	91660	1.03	11	1	0.94	0.33
DB	5944	222390	3.14	73	1	3.2	2.9
ITG	3142	73530	2.83	57	1	4.6	2.5
MSFT	24340	79635	1.00	4	1	0.92	0.24
QQQQ	21424	126431	1.00	4	1	0.62	0.1
TISI	1420	63253	2.57	77	1	5.6	3.1
VOD	6936	106215	1.00	4	1	1	0.15

	imbalance 1			imbalance 2		volum	abs. ret.
	mean	max	min	mean	std	median	median (%)
CTAS	-0.19	5.29	-7.54	-0.25	0.92	60254	0.389
DB	-0.04	8.89	-8.76	-0.17	1.36	19032	0.731
ITG	-0.04	5.99	-6.39	-0.03	0.87	7280	0.816
MSFT	-0.43	5.34	-8.60	-0.54	1.17	2108885	0.461
QQQQ	-0.40	4.96	-10.7	-0.50	1.38	4695257	0.374
TISI	-0.06	6.68	-6.39	-0.11	1.02	4399	0.614
VOD	-0.38	7.54	-7.50	-0.60	1.04	63138	0.383

In sample fitting

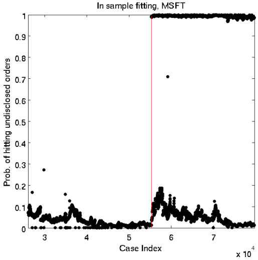


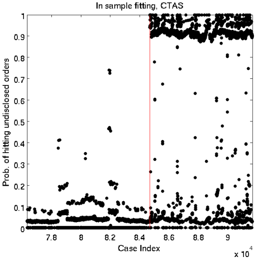


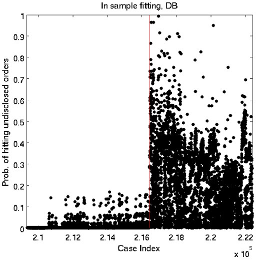


The orders hitting hiden liquidity are on the right side of the red line.

In sample fitting







The orders hitting hiden liquidity are on the right side of the red line.

Estimates

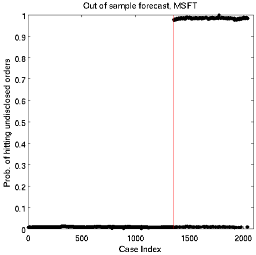
	MSFT		CTAS		DB	
	Coef.	Mag.Eff.	Coef.	Mag.Eff.	Coef.	Mag.Eff.
Constant	-0.756	–	-4.418	–	-5.97	–
Rel.Loc.	0.481	0.096	1.814	0.023	3.989	0.0087
Spr.	-3.778	-0.761	-4.618	-0.060	-7.165	-0.0156
Imb_1	-0.054	-0.011	0.026	0.00	-0.090	-0.002
Imb_2	0.034	0.007	0.032	0.001	0.054	0.0
Vlm.	0.485	0.097	0.080	0.001	0.183	0.0004
Vol.	0.043	0.009	-0.024	-0.00	0.054	0.0001
Dum_1	-0.457	-0.088	0.628	0.01	-0.563	-0.001
Dum_2	-0.302	-0.057	-0.129	-0.001	-0.021	-0.0

Red: significant at 5% level; blue: significant at 10% level.

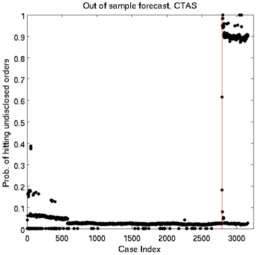
The variables except for constants and dummies are standardized to have mean 0 and deviation 1.

Out of sample forecast

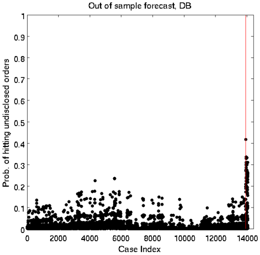
100%, 80.6%



100%, 68.3%



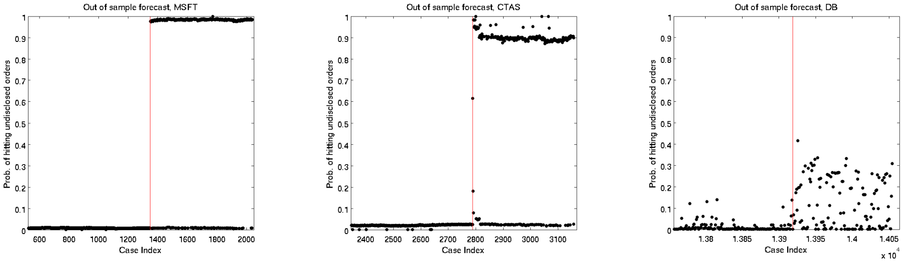
100%, 0%



The orders hitting hiden liquidity are on the right side of the red line.

Out of sample forecast

100%, 80.6% | 100%, 68.3% | 100%, 0%



The orders hitting hiden liquidity are on the right side of the red line.

Outlook

Features of the model

- It is purely based on the visible order flow data.
- It can detect the hidden liquidity.

Extension:

- Improve regression by choice-based sampling;
- Group stocks according some criteria and create a more general model;
- Predict hidden volumes.

Thank you!