Hidden liquidity

Balakaeva Maria, Gnetov Fedor, Lyappieva Anastasia

Lomonosov Moscow State University

May 7, 2024

Introduction

A growing proportion of traders on financial markets perceive a tangible benefit in concealing their trading intentions from public view. To address the rising demand, exchange operators and markets have introduced a range of order types that allow traders to hide the full extent of their standing limit orders (such as Iceberg orders or hidden orders). As a result, the proliferation of hidden liquidity has grown significantly over the past decade and nowadays accounts for a sizable proportion of overall liquidity supply in electronic equity markets.

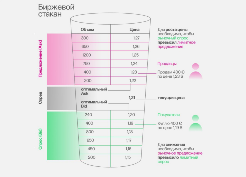
Our model

Our model allows us to draw conclusions about the driving forces affecting hidden liquidity.
The project used linear regression for panel data using dummy variables:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \sum_{j=0}^{8} \hat{\beta}_{4j} d_{4i} + \sum_{j=0}^{2} \hat{\beta}_{5j} d_{5j}$$

Data

Level 1 data includes
basic information about a transaction, such as the amount, card number, and expiration date.

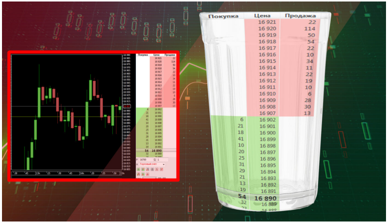


Level 2 data includes additional information, such as the tax amount, merchant's postal code, and customer code. Level 3 data includes even more information, such as line-item details of the transaction, including product descriptions, quantities, and prices. l3 data was used in the project.

Data taken from the Moscow Exchange.

Data

1) time (trading day from 10 to 19)
2) depth of the glass
3) spread
4) volatility
5) sector (IT, banks and primary sector)



More details about each sector

In the IT sector, Yandex (YNDX), Ozone (OZON), MTS (MTSS) were selected.

In banking: VTB (VTBR), rosbank (ROSB), Tinkoff (TCSG).

In the commodity sector: Gazprom (RTGZ), Lukoil (LKOH), Rosneft (ROSN).

Data

	HOUR	propotion_of_iceberg	spred	depth_sum	volatitlity	sector
0	10	0.108403	1.06000	1948.0	0.380783	1
1	11	0.031758	0.35000	1983.0	0.270379	1
2	12	0.02704	0.19002	2029.0	0.266016	1
3	13	0.020092	0.20098	2078.0	0.237938	1
4	14	0.001263	0.24100	2126.0	0.333178	1
5	15	0.029275	0.24000	2141.0	0.267890	1
6	16	0.114517	0.22900	2393.0	1.097435	1
7	17	0.037328	0.22000	2213.0	0.335445	1
8	18	0.006329	0.24098	2176.0	0.180515	1

Data

Using dummy variables transformed into the following convenient form:

	propotion_of_iceberg	spred	depth_sum	volatitlity	HOUR_10	HOUR_11	HOUR_12	HOUR_13	HOUR_14	HOUR_15	HOUR_16	HOUR_17	HOUR_18	sector_1
0	0.108403	1.06000	1948.0	0.380783	True	False	False	False	False	False	False	False	False	True
1	0.031758	0.35000	1983.0	0.270379	False	True	False	False	False	False	False	False	False	True
2	0.02704	0.19002	2029.0	0.266016	False	False	True	False	False	False	False	False	False	True
3	0.020092	0.20098	2078.0	0.237938	False	False	False	True	False	False	False	False	False	True
4	0.001263	0.24100	2126.0	0.333178	False	False	False	False	True	False	False	False	False	True
5	0.029275	0.24000	2141.0	0.267890	False	False	False	False	False	True	False	False	False	True
6	0.114517	0.22900	2393.0	1.097435	False	False	False	False	False	False	True	False	False	True
7	0.037328	0.22000	2213.0	0.335445	False	False	False	False	False	False	False	True	False	True
8	0.006329	0.24098	2176.0	0.180515	False	False	False	False	False	False	False	False	True	True

About the code

```python
import pandas as pd
import glob
import os
```

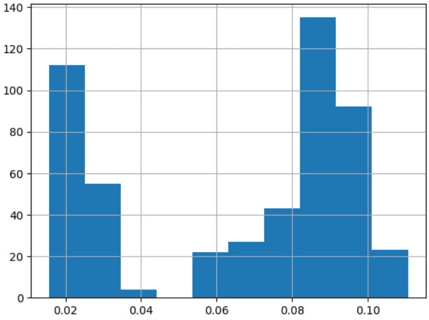
```python
from statsmodels.stats.diagnostic import het_white

# Perform White's test for heteroskedasticity
white_test = het_white(robust_results.resid, robust_results.model.exog)
```

посмотрим как выглядят depth sum

```python
import matplotlib.pyplot as plt

big_df['depth_sum'] /= 10000
big_df['depth_sum'].hist()
plt.show()
```



Data processing

The peculiarity of the two variables Spread and Volatility needs to be normalized, because some companies have shares worth thousands, while others have pennies, this does not indicate the importance of the company, just someone divided everything into 10 shares, and someone into 100,000 Therefore, we initially normalized them and only then calculated spred and volatility: (x-min)/(max-min) Then everything will lie in the interval [0,1]

Combating multicollinearity and heteroscedasticity

When working with the data, heteroscedasticity (White's test) and multicollinearity were discovered. The first problem was solved by introducing robust errors. To combat multicolinearity, was removed 1 hour and one of the sectors: banking (with which a dependence was found).

```
{'Test Statistic': 38.05332338905303, 'Test Statistic p-value': 0.034239120022604555, 'F-Statistic': 1.6077156826287116, 'F-Test p-value': 0.03263007725862992}
```

'Test Statistic': 38.05332338905303, 'Test Statistic p-value': 0.034239120022604555, 'F-Statistic': 1.6077156826287116, 'F-Test p-value': 0.03263007725862992

Results

Robust linear Model Regression Results

Dep. Variable:	propotion_of_iceberg	No. Observations:	513
Model:	RLM	Df Residuals:	499
Method:	IRLS	Df Model:	13
Norm:	HuberT		
Scale Est.:	mad		
Cov Type:	H1		
Date:	Mon, 06 May 2024		
Time:	22:51:51		
No. Iterations:	21		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0343	0.007	5.160	0.000	0.021	0.047
spred	-0.0303	0.033	-0.917	0.359	-0.095	0.034
depth_sum	0.4884	0.217	2.252	0.024	0.063	0.913
volatitlity	0.0167	0.024	0.700	0.484	-0.030	0.064
sector_1	-0.0342	0.015	-2.211	0.027	-0.064	-0.004
sector_2	-0.0150	0.014	-1.106	0.269	-0.041	0.012
time_11	0.0045	0.007	0.650	0.516	-0.009	0.018
time_12	-0.0143	0.007	-2.040	0.041	-0.028	-0.001
time_13	-0.0124	0.007	-1.757	0.079	-0.026	0.001
time_14	-0.0112	0.007	-1.571	0.116	-0.025	0.003
time_15	-0.0164	0.007	-2.276	0.023	-0.031	-0.002
time_16	-0.0131	0.007	-1.790	0.073	-0.027	0.001
time_17	-0.0040	0.007	-0.534	0.593	-0.019	0.011
time_18	-0.0010	0.007	-0.140	0.889	-0.015	0.013

Results

Robust linear Model Regression Results

Dep. Variable:	propotion_of_iceberg	No. Observations:	513
Model:	RLM	Df Residuals:	499
Method:	IRLS	Df Model:	13
Norm:	HuberT		
Scale Est.:	mad		
Cov Type:	H1		
Date:	Mon, 06 May 2024		
Time:	23:09:42		
No. Iterations:	21		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0162	0.021	-0.762	0.446	-0.058	0.026
spred	-0.0303	0.033	-0.917	0.359	-0.095	0.034
depth_sum	0.4884	0.217	2.252	0.024	0.063	0.913
volatitlity	0.0167	0.024	0.700	0.484	-0.030	0.064
sector_0	0.0342	0.015	2.211	0.027	0.004	0.064
sector_2	0.0192	0.008	2.387	0.017	0.003	0.035
time_10	0.0164	0.007	2.276	0.023	0.002	0.031
time_11	0.0209	0.007	3.021	0.003	0.007	0.034
time_12	0.0021	0.007	0.311	0.756	-0.011	0.016
time_13	0.0040	0.007	0.580	0.562	-0.009	0.017
time_14	0.0052	0.007	0.762	0.446	-0.008	0.019
time_16	0.0033	0.007	0.481	0.630	-0.010	0.017
time_17	0.0124	0.007	1.820	0.069	-0.001	0.026
time_18	0.0154	0.007	2.233	0.026	0.002	0.029

Results

The results turned out
to be quite unobvious at first glance: volatility and spread turned out to be insignificant.
But time (10,11 and 18), depth
of the glass, banking and oil sectors turned out to be very influential on hidden liquidity.