SRG Market microstructure

Report on my research

Vsevolod Zaostrovsky

Supervisors: Anton Belyakov, Alexey Savin

Vega Institute Foundation

January 14, 2024

Data specification

Спецификация формата данных Full Orders Log (тип A):

Наименование поля	Тип	Описание
NO	*Int*	Номер записи
SECCODE	*String*	Код инструмента
BUYSELL[1]	*Char*	Признак купли/продажи
		B= Купить
		S= Продать
TIME[2]	*LongInt*	Время в формате **HHMMSSZZZXXX** с марта 2016
		Время в формате HHMMSSZZZ до марта 2016
ORDERNO	*Int*	Номер заявки
ACTION	*Byte*	Тип события:
		0=Снятие заявки
		1=Постановка заявки
		2=Сделка
PRICE[3]	*Float*	Цена заявки
VOLUME	*Int*	Объем
		Для action=1 – видимый объем поставленной заявки
		Для action=2 – объем сделки
		Для action=0 – остаток видимой части заявки
Следующие поля заполняются только для сделок (ACTION=2)		
TRADENO	*Int*	Номер сделки
TRADEPRICE	*Float*	Цена сделки

Data preparing — 1

Initially, the parser data was in the form of:

$$[[102.260, 50], [102.280, 100], [102.294, 35], [102.310, 200], [102.500, 2546], [0., 0.]]$$

I changed it to give the data in the following form:

$$10:00:01.000609985\ [[61.782, 40000], [61.870, 100000], \cdots]$$
$$\text{Price } 61.79 \text{ Vol } 40000$$
$$10:00:01.000609985\ [[61.870, 100000], \cdots]$$

I choose all the deals and for each deal the parser prints the LOB before and after the deal and the price and volume of order in between. Then using Python I collected vectors of four numbers: (Time, AskBefore, AskAfter, Volume).

Data preparing – 2

	Time	AskBefore	AskAfter	Volume
0	1.0006099849997554	61.782	61.87	40000.0
1	2.0004541379967122	61.84	61.84	48000.0
2	2.0004635579971364	61.84	61.84	48000.0
3	2.000483832001919	61.842	61.842	13000.0
4	2.000486406999698 4	61.842	61.847	87000.0
5	2.000486406999698 4	61.847	61.847	7000.0
6	2.0005590279979515	61.792	61.845	10000.0
7	2.00059898699692	61.79	61.842	100000.0
8	2.000793077997514	61.835	61.835	11000.0
9	2.0008411289963988	61.815	61.815	11000.0

Our methodology to fit parameters ρ, κ, q

We chose regression to find parameters:

$$\frac{\Delta A_{k+2}}{\Delta t_{k+2}} - \frac{\Delta A_{k+1}}{\Delta t_{k+1}} = -\rho \Delta A_{k+1} + \rho\lambda x_{k+1} + (\kappa + \lambda)(\frac{x_{k+2}}{\Delta t_{k+2}} - \frac{x_{k+1}}{\Delta t_{k+1}}).$$

Where all the information needed can be extracted from the l3 data:

- ΔA_k is an ask change after execution of the limit order with the depth x_k. So, $\Delta A_k = \text{AskAfter}(k) - \text{AskBefore}(k)$ and $x_k = \text{Volume}(k)$.
- Δt_k is a time between k and $k+1$ orders of dataset. So, $\Delta t_k = \text{Time}(k+1) - \text{Time}(k)$

Regression results – all the data.

OLS Regression Results

Dep. Variable:	SUM	R-squared:	0.043
Model:	OLS	Adj. R-squared:	0.042
Method:	Least Squares	F-statistic:	144.3
Date:	Sun, 14 Jan 2024	Prob (F-statistic):	1.74e-91
Time:	14:25:56	Log-Likelihood:	-90059.
No. Observations:	9743	AIC:	1.801e+05
Df Residuals:	9739	BIC:	1.802e+05
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
-rho	-1.503e+04	981.693	-15.314	0.000	-1.7e+04	-1.31e+04
rho lambda	0.0009	0.000	7.521	0.000	0.001	0.001
kappa + lambda	6.593e-09	5.22e-10	12.628	0.000	5.57e-09	7.62e-09
const	0.4313	27.003	0.016	0.987	-52.501	53.363

Omnibus:	5504.252	Durbin-Watson:	2.828
Prob(Omnibus):	0.000	Jarque-Bera (JB):	16410123.140
Skew:	1.015	Prob(JB):	0.00
Kurtosis:	204.045	Cond. No.	1.89e+12

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.89e+12. This might indicate that there are strong multicollinearity or other numerical problems.

Regression results – more then zero.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    SUM   R-squared:                       0.029
Model:                            OLS   Adj. R-squared:                  0.028
Method:                 Least Squares   F-statistic:                     36.83
Date:                Sun, 14 Jan 2024   Prob (F-statistic):           1.88e-23
Time:                        14:28:01   Log-Likelihood:                -36363.
No. Observations:                3745   AIC:                         7.273e+04
Df Residuals:                    3741   BIC:                         7.276e+04
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
-rho         -1.255e+04   1602.043     -7.832      0.000   -1.57e+04   -9406.311
rho lambda       0.0011      0.000      3.972      0.000       0.001       0.002
kappa + lambda 8.242e-09   1.32e-09      6.239      0.000    5.65e-09    1.08e-08
const           59.1374     70.842      0.835      0.404     -79.756     198.030
==============================================================================
Omnibus:                     1636.308   Durbin-Watson:                   2.891
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1214322.502
Skew:                           0.589   Prob(JB):                         0.00
Kurtosis:                      91.208   Cond. No.                     1.23e+12
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.23e+12. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Regression results – more then 100bp.

OLS Regression Results

Dep. Variable:	SUM	R-squared:	0.439
Model:	OLS	Adj. R-squared:	0.436
Method:	Least Squares	F-statistic:	159.5
Date:	Sun, 14 Jan 2024	Prob (F-statistic):	2.17e-76
Time:	14:28:20	Log-Likelihood:	-6133.8
No. Observations:	616	AIC:	1.228e+04
Df Residuals:	612	BIC:	1.229e+04
Df Model:	3		
Covariance Type:	nonrobust		

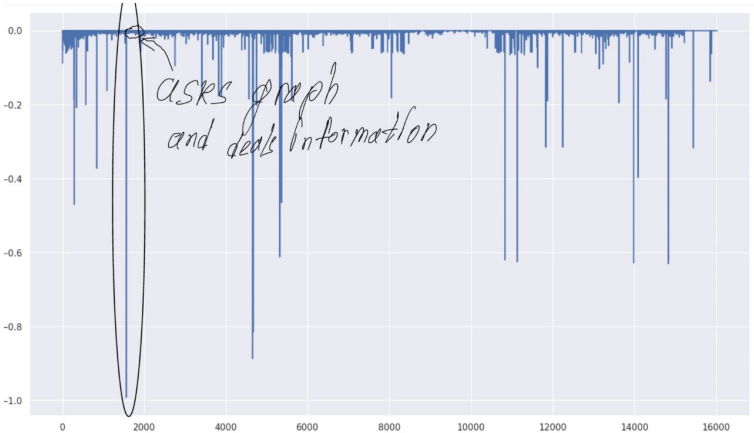
	coef	std err	t	P>|t|	[0.025	0.975]
-rho	-4099.6975	2230.535	-1.838	0.067	-8480.128	280.733
rho lambda	0.0018	0.002	1.007	0.314	-0.002	0.005
kappa + lambda	5.565e-07	2.55e-08	21.813	0.000	5.06e-07	6.07e-07
const	125.6809	243.258	0.517	0.606	-352.040	603.402

Omnibus:	265.009	Durbin-Watson:	2.600
Prob(Omnibus):	0.000	Jarque-Bera (JB):	16011.443
Skew:	1.061	Prob(JB):	0.00
Kurtosis:	27.886	Cond. No.	8.76e+10

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 8.76e+10. This might indicate that there are
strong multicollinearity or other numerical problems.

Backtest procedure.

- Find big gaps and remember next asks and deals for each.
- Consider big deal and little deals as ours and calculate metrics for them.
- Research if asks dynamics follows $A_t = \overline{p}_t + \frac{s}{2} + x_1 \kappa e^{-\rho t}$.



Important (imho) Ideas and Questions

- Backtest procedure... See the previous frame.
- It is interesting what is happening with other instruments. Also, the important idea: for some instruments OW model is useless. For which?
- It is interesting what is happening with that instrument throughout the day. We can consider structural breaks and fit the model on different sygments.
- What does mean regression results?
- Have I done everything right?
- What else can we do?
- Does my bt methodology correct?