

# 最小大单数统计情况

2020 年 8 月 4 日

## 1 整体概述

这里我们一共统计了自 2020 年 1 月 2 日至 2020 年 7 月 28 日七个月期间所有交易日内的 6112 次涨停破板情况，我们按照最小大单数 0-9 分为 10 类，生成 10 个子表，其中我们统计到每一类的涨停破板情况次数如下表 1 所示：

表 1: 不同最小大单数对应的情况次数

最小大单数	情况次数	最小大单数	情况次数
0	1170	5	519
1	877	6	463
2	747	7	419
3	650	8	366
4	572	9	329

## 2 分类概述

接下来我们将从一下四个方面，分别分析不同的最小大单数对应的情况：

- 破板的概率，上午/下午破板的概率
- 破板/不破版的平均收益率
- fake\_time 与 uplimit\_time 差的统计分析
- fake\_time 与 uplimit\_time 差与收益率之间的关系

2.1 破板概率

2.1.1 全天破板概率

我们选取时间节点 14 : 57 : 00.000 作为真假破板的时间节点，将破板时间（break\_time）在 14 : 57 : 00 之前的破板情况归为真破板，反之则为假破板。经统计我们得到不同最小大单数破板时间在 14 : 57 : 00 之后的情况数分别为：

表 2: 不同最小大单数对应的假破板次数			
最小大单数	假破板次数	最小大单数	假破板次数
0	366	5	132
1	266	6	112
2	222	7	97
3	181	8	80
4	153	9	70

则相应的破板时间在 14 : 57 : 00 之前的情况数分别为：

表 3: 不同最小大单数对应的真破板次数			
最小大单数	真破板次数	最小大单数	真破板次数
0	804	5	387
1	611	6	351
2	525	7	322
3	469	8	286
4	419	9	259

基于真破板的情况，我们可以计算得到不同最小大单数的破板概率分别为：

表 4: 不同最小大单数真破板概率			
最小大单数	真破板概率	最小大单数	真破板概率
0	68.72%	5	74.57%
1	69.67%	6	75.81%
2	70.28%	7	76.85%
3	72.15%	8	78.14%
4	73.25%	9	78.72%

2.1.2 上午/下午破板概率

接下来我们在分别对于以 12 : 00 : 00.000 区分上午与下午，来分别统计上午和下午的破板情况。

表 5: 不同最小大单数上午下午真破板数统计					
最小大单数	上午真破板数	下午真破板数	最小大单数	上午真破板数	下午真破板数
0	509	295	5	284	103
1	416	195	6	258	93
2	359	166	7	243	79
3	328	141	8	218	68
4	306	113	9	203	56

表 6: 不同最小大单数上午下午假破板数统计					
最小大单数	上午假破板数	下午假破板数	最小大单数	上午假破板数	下午假破板数
0	173	193	5	90	42
1	139	127	6	76	36
2	130	92	7	70	27
3	112	69	8	60	20
4	101	52	9	57	13

由此计算得到上午下午分别的破板概率为如表 7所示，从表中数据我们可以发现，涨停时间在上午的破板概率普遍高于涨停时间在下午的破板概率，其中最小大单数为 8、9 除外。

表 7: 不同最小大单数上午下午假破板数统计					
最小大单数	上午破板概率	下午破板概率	最小大单数	上午破板概率	下午破板概率
0	74.63%	60.45%	5	75.94%	71.03%
1	74.95%	60.56%	6	77.25%	72.09%
2	73.42%	64.34%	7	77.64%	74.53%
3	74.55%	67.14%	8	78.42%	77.27%
4	75.18%	68.48%	9	78.08%	81.16%

3 平均收益率

由第二节我们知道每一类最小大单数破板的情况，接下来我们分别分析在每一中 最小大单数的情况下的破板与不破板的算术平均收益率的情况，在这里我们选用的收益率均是复权后收益率。从数据中我们可以看出，当最小大单数较小（0、1）时破板平均收益率为正，其余情况均为负；而不破板的平均收益率为正，且平均收益率数值也大于破板的平均收益率数值。

表 8: 破板平均收益率

最小大单数	收益率	最小大单数	收益率
0	0.21%	5	-0.05%
1	0.20%	6	0.16%
2	0.09%	7	0.12%
3	0.03%	8	0.14%
4	-0.03%	9	0.25%

表 9: 不破板平均收益率

最小大单数	收益率	最小大单数	收益率
0	2.30%	5	1.98%
1	2.03%	6	1.86%
2	2.27%	7	2.04%
3	2.18%	8	2.14%
4	2.13%	9	2.14%

4 时间差分析

这一节我们统计分析 fake\_time 与 uplimit\_time 差值的均值，由于两个时间均发生在同一天，表中数据均忽略 0 天，具体均值如表 10所示，我们可以看出时间差值的均值在大体上随着最小大单数的增加而变小。

表 10: fake\_time 与 uplimit\_time 差值的均值

最小大单数	时间差均差	最小大单数	时间差均值
0	00 : 20 : 26.403	5	00 : 15 : 13.369
1	00 : 19 : 05.573	6	00 : 12 : 42.839
2	00 : 18 : 31.384	7	00 : 13 : 23.585
3	00 : 17 : 20.085	8	00 : 11 : 20.463
4	00 : 15 : 51.632	9	00 : 11 : 44.513

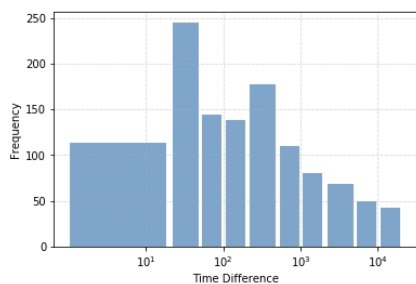
接下来我们引入其他的统计量进一步分析 fake\_time 与 uplimit\_time 的差值。如表 11、12、13、14所示，我们可以看出时间差的中位数、极差、标准差、变异系数在大体上随着最小大单数的增加均减小。

为了更加直观的反应在不同最小大单数的情况下,fake\_time 与 uplimit\_time 的时间差的分布,我们利用直方图 4的形式展示出来,分别的数据区间为: [0,20], [20,50], [50,100], [100,200], [200,500], [500,1000], [1000,2000], [2000,5000], [5000,10000], [10000,20000]

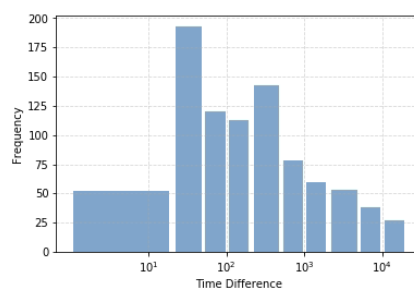
表 11: fake\_time 与 uplimit\_time 差值的中位数

最小大单数	时间差中位数	最小大单数	时间差中位数
0	00 : 02 : 38.365	5	00 : 01 : 56.190
1	00 : 02 : 40.860	6	00 : 01 : 49.840
2	00 : 02 : 26.600	7	00 : 01 : 41.480
3	00 : 02 : 18.440	8	00 : 01 : 23.255
4	00 : 02 : 02.765	9	00 : 01 : 19.050

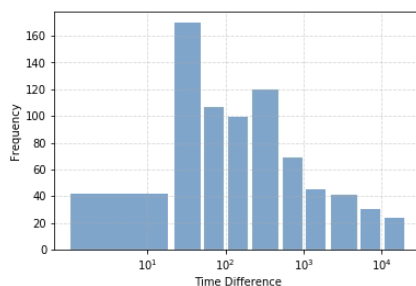
图 1: 十种最小大单数时间差分布直方图



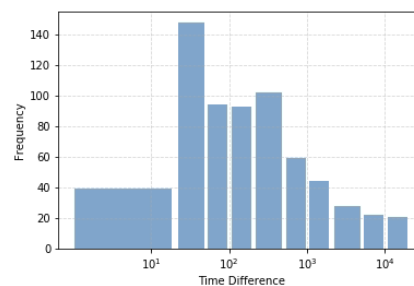
(a) min\_bigorder=0



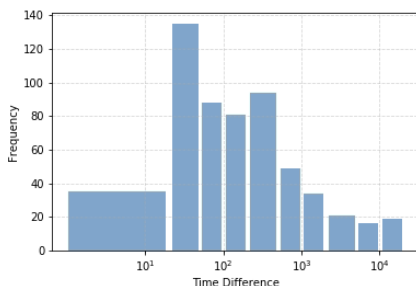
(b) min\_bigorder=1



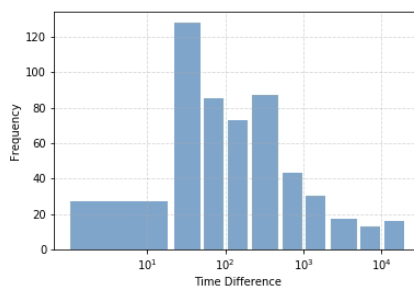
(c) min\_bigorder=2



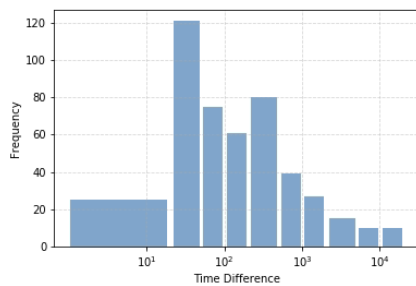
(d) min\_bigorder=3



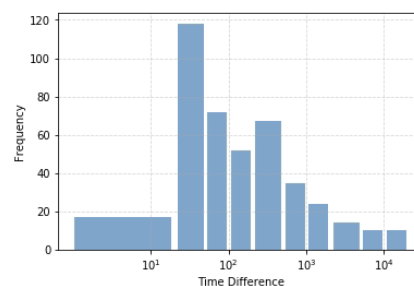
(e) min\_bigorder=4



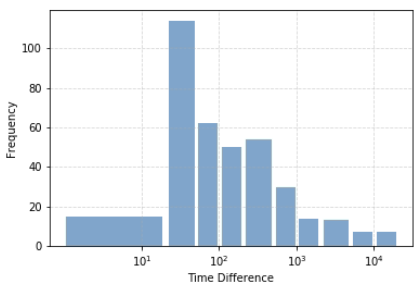
(f) min\_bigorder=5



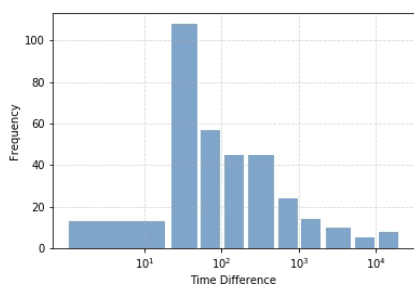
(g) min\_bigorder=6



(h) min\_bigorder=7



(i) min\_bigorder=8



(j) min\_bigorder=9

表 12: fake\_time 与 uplimit\_time 差值的极差

最小大单数	时间差极差	最小大单数	时间差极差
0	05 : 29 : 17.360	5	05 : 03 : 46.440
1	05 : 00 : 53.440	6	04 : 02 : 19.860
2	04 : 35 : 18.100	7	04 : 00 : 20.410
3	04 : 38 : 42.180	8	04 : 00 : 20.390
4	04 : 02 : 09.620	9	04 : 00 : 17.450

表 13: fake\_time 与 uplimit\_time 差值的标准差

最小大单数	时间差标准差	最小大单数	时间差标准差
0	00 : 48 : 34.924	5	00 : 41 : 47.778
1	00 : 45 : 03.899	6	00 : 35 : 16.033
2	00 : 44 : 41.013	7	00 : 36 : 44.862
3	00 : 43 : 18.201	8	00 : 33 : 32.333
4	00 : 41 : 31.948	9	00 : 35 : 54.581

表 14: fake\_time 与 uplimit\_time 差值的变异系数

最小大单数	时间差变异系数	最小大单数	时间差变异系数
0	0.421	5	0.364
1	0.424	6	0.361
2	0.415	7	0.364
3	0.400	8	0.338
4	0.382	9	0.327

4.1 时间差与收益率之间的关系

我们以秒为单位，来分析 fake\_time 与 uplimit\_time 的时间差与收益率之间的关系。我们首先计算在不同最小大单数的情况下两个变量之间的相关系数，如表 15

表 15: 时间差与收益率的相关系数			
最小大单数	相关系数	最小大单数	相关系数
0	0.052	5	0.059
1	0.028	6	0.043
2	0.046	7	0.054
3	0.062	8	0.083
4	0.044	9	0.097

接下来我们重新划分时间差区间，来计算每一个区间内的平均收益率：

表 16: 时间差区间对应的均值收益率			
时间差区间	均值收益率	时间差区间	均值收益率
[0, 0.5)	0.011	[0.5, 1)	-0.057
[1, 20)	-0.010	[20, 50)	0.001
[50, 100)	0.005	[100, 200)	0.009
[200, 500)	0.011	[500, 1000)	0.013
[1000, 2000)	0.011	[2000, 5000)	0.014
[5000, 10000)	0.016	[1000, +∞)	0.017