

# 目录

数据处理.....	2
目标.....	2
数据预处理 .....	2
数据划分 .....	2
Feature 测试、结果展示说明 .....	2
Feature 1. Quantity Trend Aligned with Quote Price .....	3
1. Idea.....	3
2. Indicator 的计算 .....	3
3. 测试结果 .....	4
Feature 2. Total Quote Quantity Trend .....	8
1. Idea.....	8
2. Indicator 的计算 .....	8
3. 测试结果 .....	8
Feature 3. Quote Quantity Weighted Prcie .....	11
1. Idea.....	11
2. Indicator 的计算 .....	11
3. 测试结果 .....	11
Feature 4. Quote Pressure Index .....	15
1. Idea.....	15
2. Indicator 的计算 .....	15
3. 测试结果 .....	15
Coefficient of Correlation : .....	19
Feature 5. Use vol/openInt as Intensity Multiplier .....	20

# 数据处理

## 目标

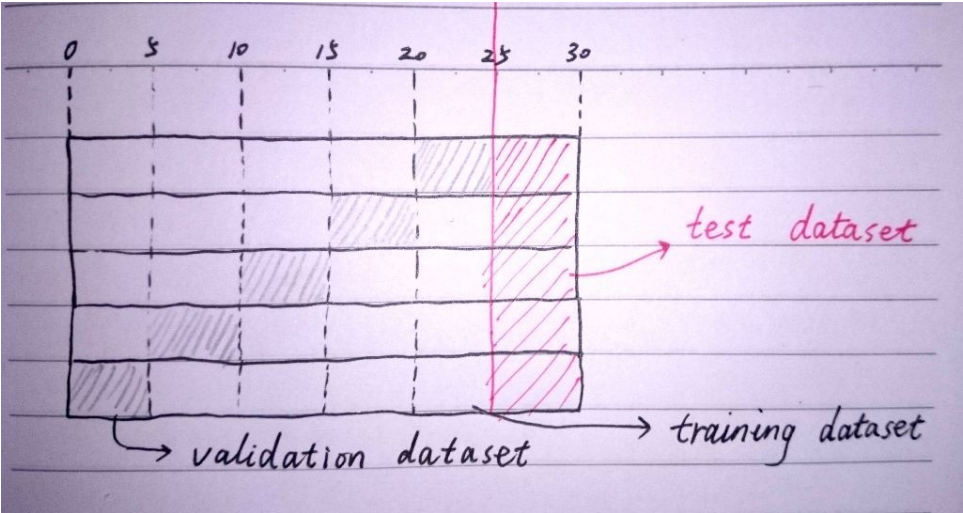
预测 30 - 60tick 后的 price move  
( 测试 feature 时，均默认使用 44 ticks )

## 数据预处理

	预处理	目的
1	drop 每天的尾部 44 tick	否则求得的 price move 会有跨天的
2	Drop 每天 9 点和 21 点开盘后的 50 秒	这段时间 price move 的 variance 较高，会污染样本
3	删去涨跌停的数据	涨跌停后 quote price 有一边为 0
4	删去涨跌停的这部分数据前方 44 tick 和后方 100 tick	前方的数据会影响 price move 的计算； 后方的数据会影响 indicator 的计算（如果 indicator 需要使用一段时间的数据了话）

## 数据划分

0-25 天的数据作为 accessible dataset, 25-30 天的数据作为 test dataset  
再将 0-25 天的 accessible dataset 分为 training dataset 和 validation dataset. ( 见图表 1 )



图表 1 数据集划分方式

## Feature 测试、结果展示说明

- 1 使用 training data 作为输入数据，OLS 回归得到估计的斜率与截距。
- 2 若斜率正负号符合设计逻辑，继续利用得到的模型预测 validation data 的 price move，与真实数据比较。
- 3 最后使用 test data 作为测试。

**结果展示：**下面的每个 Feature，首先列出在 training data, validation data 上的 Rsquare 均值，及 test data 上的 Rsquare。接着给出几个具体的回归结果。

# Feature 1. Quantity Trend Aligned with Quote Price

## 1. Idea

- 1 观察到，在价格变动前（以上涨为例），askQty\_0 不断减小。  
考虑用差分描述这一趋势
- 2 如果只看某一档的差分，当一档对应的价格改变时，得到的 Qty 的差分就是两个不同价格上的 Qty 之差，没有意义。  
考虑将原始数据中，按档位对齐的五档买卖数据，处理成按价格的数值对齐。
- 3 时常出现的一个情况是（以上涨为例），ask0 的量被消耗完，接着这个价位变成 bid0，这种情况下直接差分也无法描述。  
考虑将 askQty 都乘以-1，bidQty 乘以+1（不变），这样之后再做 diff 就解决了问题。同时这样也减少了一个有限数与 NaN 相加减的情况，减少了信息损失。

## 2. Indicator 的计算

操作	理由
将买卖 quote 分开	否则之后 groupby 命令会混在一起无法区分
分别按照价格的数值 groupby	区分不同价格数值的 Qty
按 index 与 column 对齐合并	1 价格不会乱；2 index 能保持完整
得到一个 DataFrame，index 为 datetimeindex，column 为 quote 中出现过的价格数值，value 为这一时刻这一价格上 Qty 的量（绝大部分 value 是 NaN，因为每个 index 只有 10 个价格）。Bid 为正，ask 为负。（见图表）	
做 diff(n=1)	Qty “变化量” 的量化
对整个 DataFrame 取  np.sign(df) * np.log(np.abs(df) + 1)	不同档 Qty 的 scale 不同，之间不可直接做运算； DataFrame 中 abs 最小的数据是 0 和 1，虽然没有 0-1 之间的数据，但若直接求 log 则 0 和 1 无法区分，所以需要加上 1
沿 axis=1 取 mean	沿不同档的 quote 的一个平均，由于无法区分哪个是哪一档所以无法赋予权重。之后测试中通过使用不同档数目的 Qty 来测试这一影响
沿 axis=0 取 rolling mean	Qty 的变化是一个连续的过程，不能只考虑 1 个 tick



TA 品种 , halflife=12 , insample 取 0-20 天 , outsample 为 20-25 天

OLS Regression Results

```
=====
Dep. Variable:          midPrc      R-squared:                0.018
Model:                  OLS         Adj. R-squared:           0.018
Method:                 Least Squares   F-statistic:             3.243e+04
Date:                  Tue, 19 Jul 2016   Prob (F-statistic):       0.00
Time:                  09:34:53         Log-Likelihood:          -3.0901e+06
No. Observations:      1786394         AIC:                    6.180e+06
Df Residuals:          1786392         BIC:                    6.180e+06
Df Model:               1
=====
```

	coef	std err	t	P> t	[95.0% Conf. Int.]
const	0.0038	0.001	3.705	0.000	0.002 0.006
None	1.3327	0.007	180.079	0.000	1.318 1.347

```
=====
Omnibus:                633345.200   Durbin-Watson:           0.089
Prob(Omnibus):           0.000       Jarque-Bera (JB):        134099992.967
Skew:                   0.514        Prob(JB):                0.00
Kurtosis:               45.433       Cond. No.:               7.25
=====
```

[OutSample R\_SQUARE: 0.017340]

TA 品种 , halflife=12 , insample 取 0-20 天 , outsample 取 25-30 天

OLS Regression Results

```
=====
Dep. Variable:          midPrc      R-squared:                0.018
Model:                  OLS         Adj. R-squared:           0.018
Method:                 Least Squares   F-statistic:             3.243e+04
Date:                  Tue, 19 Jul 2016   Prob (F-statistic):       0.00
Time:                  05:34:30         Log-Likelihood:          -3.0901e+06
No. Observations:      1786394         AIC:                    6.180e+06
Df Residuals:          1786392         BIC:                    6.180e+06
Df Model:               1
=====
```

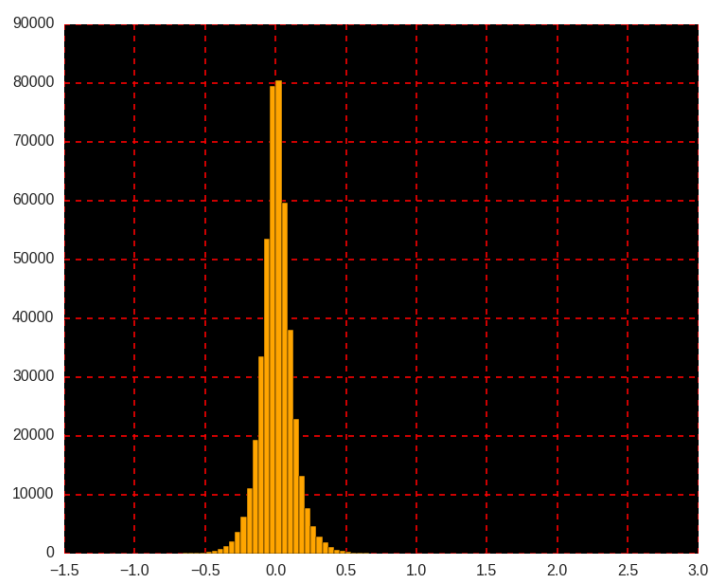
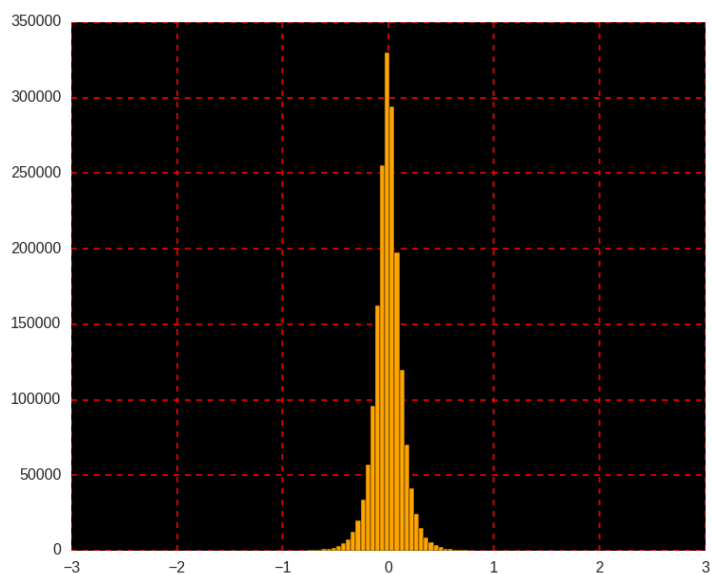
	coef	std err	t	P> t	[95.0% Conf. Int.]
const	0.0038	0.001	3.705	0.000	0.002 0.006
None	1.3327	0.007	180.079	0.000	1.318 1.347

```
=====
Omnibus:                633345.200   Durbin-Watson:           0.089
Prob(Omnibus):           0.000       Jarque-Bera (JB):        134099992.967
Skew:                   0.514        Prob(JB):                0.00
Kurtosis:               45.433       Cond. No.:               7.25
=====
```

[OutSample R\_SQUARE: 0.006996]

计算中引入的额外参数：	对预测的影响（以 outsample Rsquare 衡量）
②沿 axis=1 取 mean 时用 <u>几档</u> ， <u>权重</u> 是多少	只用 1 档比五档好 0.2% 用 3 档和 1 档效果相近
沿 axis=0 取 rolling mean 时用的 <u>window</u> 长度	随 window 的增长，rsquare 先增后减，在 10-16 之间达到极大值
沿 axis=0 取 rolling mean 时的 <u>权重分配</u>	简单移动平均 SMA 比指数移动平均 EWMA 好 0.2%

值的分布（上方是在 training\_data 上的分布，下方是在 validation data 上的分布）：



一个问题：

当品种取 PTA, training dataset 取 0-20 天，outsample 取 25-30 天测试时，Rsquare 只有 0.6%

若反过来 training 取 10-30 天，outsample 取 0-5 天；或 insample 取 5-25 天，outsample 取 0-5 天，则是 outsample 比 insample Rsquare 高，且 insample rsquare 也有 0.9%以上。

测试发现，所用 training dataset 在时间上越靠后，线性模型斜率越小；同时 in out sample 样本时间距离越远，效果越差，推测是由于该品种价格的 variance 随时间减小较快（交易热度降低）所致。  
使用其他品种，均有类似情况。

若 training, validation, test dataset 都是从 30 天中均匀穿插随机抽取，则 feature 的预测能力在三个 dataset 上几乎没有区别。

# Feature 2. Total Quote Quantity Trend

## 1. Idea

- 1 观察到，当价格连续变动时（以上涨为例），TotalBidQty 会连续增加，TotalAskQty 会连续下降。两者之差的变化则更显著。  
考虑用差分描述这一变化。
- 2 既要能描述这一连续变化的过程，又要兼顾灵敏性。  
选择合适的窗口长度和加权法求移动平均

## 2. Indicator 的计算

操作	理由
TotalBidLot – TotalAskLot	要算的是总申买申卖量的变化量，所以可以放一起考虑
diff(n=1)	差分，量化“变化量”
np.sign(ser) * np.log(np.abs(ser) + 1)	上一步得到的量变化范围太大（outlier 太多），不适合线性模型。做一个 shrink
沿 axis=0 取 rolling mean	量化“Qty 连续变化”这一过程

## 3. 测试结果

对某个参数做控制变量测试时，其他默认参数为：  
品种 TA, 使用三档数据，沿 axis=0 算 halflife=12 的移动平均。

Cross-Validation 及不同品种平均 Rsquare：

数据集	Rsquare
Training data（平均值）	1.6%
Validation data（平均值）	1.4%
Test data Rsquare	0.6%

几个具体结果：



MA 品种 , halflife=12 , insample 取 0-20 天 , outsample 取 20-25 天

OLS Regression Results

```
=====
Dep. Variable:          midPrc      R-squared:                0.016
Model:                  OLS         Adj. R-squared:            0.016
Method:                 Least Squares   F-statistic:              2.973e+04
Date:                  Tue, 19 Jul 2016   Prob (F-statistic):       0.00
Time:                  09:38:31         Log-Likelihood:          -1.9331e+06
No. Observations:      1779031         AIC:                    3.866e+06
Df Residuals:          1779029         BIC:                    3.866e+06
Df Model:              1
=====
```

```
=====
              coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----
const          0.0012      0.001        2.150      0.032      0.000      0.002
total_indicator  0.1559      0.001     172.416      0.000      0.154      0.158
=====
```

```
=====
Omnibus:          443340.435   Durbin-Watson:           0.084
Prob(Omnibus):    0.000       Jarque-Bera (JB):        33898640.010
Skew:             -0.001       Prob(JB):                0.00
Kurtosis:         24.385       Cond. No.                1.68
=====
```

[OutSample R\_SQUARE: 0.011591]

MA 品种 , halflife=12 , insample 取 0-20 天 , outsample 取 25-30 天

OLS Regression Results

```
=====
Dep. Variable:          midPrc      R-squared:                0.016
Model:                  OLS         Adj. R-squared:            0.016
Method:                 Least Squares   F-statistic:              2.973e+04
Date:                  Tue, 19 Jul 2016   Prob (F-statistic):       0.00
Time:                  05:22:01         Log-Likelihood:          -1.9331e+06
No. Observations:      1779031         AIC:                    3.866e+06
Df Residuals:          1779029         BIC:                    3.866e+06
Df Model:              1
=====
```

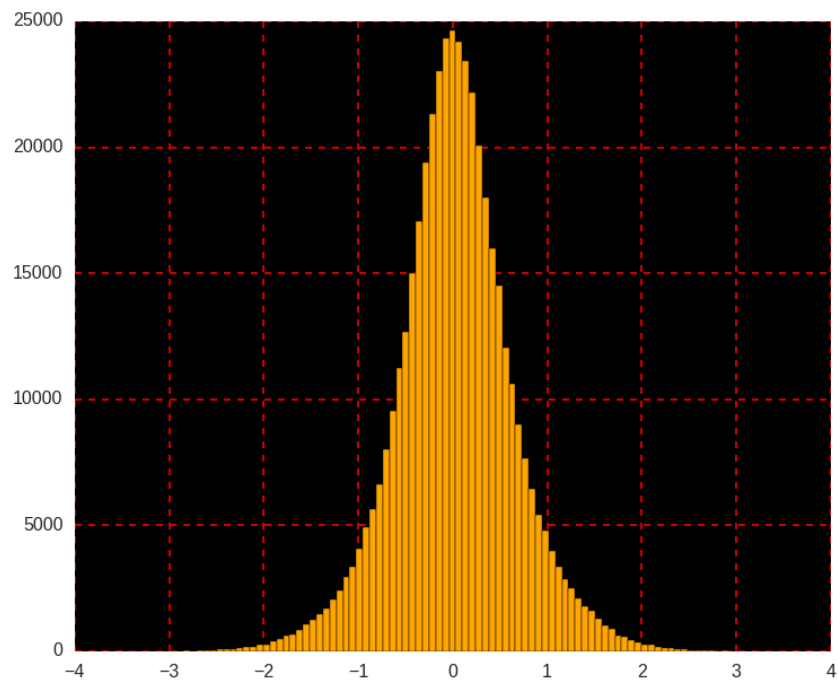
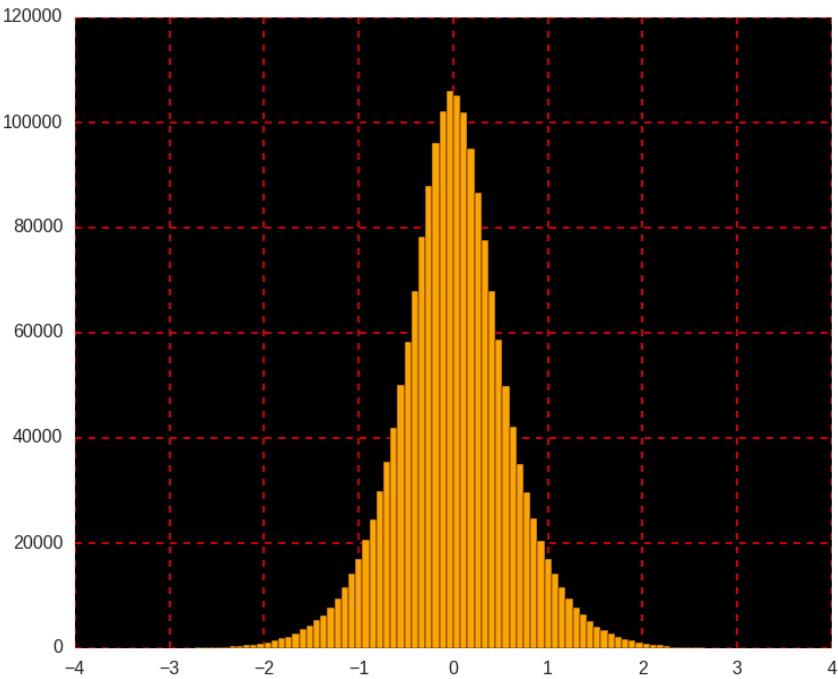
```
=====
              coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----
const          0.0012      0.001        2.150      0.032      0.000      0.002
total_indicator  0.1559      0.001     172.416      0.000      0.154      0.158
=====
```

```
=====
Omnibus:          443340.435   Durbin-Watson:           0.084
Prob(Omnibus):    0.000       Jarque-Bera (JB):        33898640.010
Skew:             -0.001       Prob(JB):                0.00
Kurtosis:         24.385       Cond. No.                1.68
=====
```

[OutSample R\_SQUARE: 0.010982]

计算中引入的额外参数：	对预测的影响（以 outsample Rsquare 衡量）
沿 axis=0 取 exponetial rolling mean 时用的 <u>halflife</u> 长度	随 halflife 的增长，Rsquare 先增后减，在 8-12 之间达到极大值
沿 axis=0 取 rolling mean 时的 <u>权重分配</u>	指数移动平均 EWMA 比简单移动平均 SMA 好 0.4%

值的分布（上方是在 training\_data 上的分布，下方是在 validation data 上的分布）：



# Feature 3. Quote Quantity Weighted Prcie

## 1. Idea

五档价格上量的相对强弱，推动了 midPrice 朝不同方向运动。  
考虑以五档 Qty 作为权重，Prc 作为值进行加权平均，得到平均价格与当前 midPrc 之差作为买卖量相对强弱的指标。

## 2. Indicator 的计算

操作	理由
给定档位（如三档 k=2，五档 k=4）求得 $lwap = \frac{\sum_{i=1}^k (bidPrc_i * bidQty_i + askPrc_i * askQty_i)}{\sum_{i=1}^k (bidQty_i + askQty_i)}$	可以看作由 Quote Qty 决定的“均衡价位”
Mid – lwap	当前价位与均衡价位之差。若假设市场将达到均衡则这个量决定了接下来价格运动的方向和大小。
做/不做 moving average	看测试结果

注：如若买方量较大，则权重较重，加权平均价更低，indicator 的值更大，因而回归出的斜率应为正数。

## 3. 测试结果

对某个参数做控制变量测试时，其他默认参数为：  
品种 TA, 使用一档数据，只使用当前 tick 的数据（不做沿时间轴的任何平均）。

Cross-Validation 及不同品种平均 Rsquare：  
insample 5.2%，outsample 4.9%

数据集	Rsquare
Training data（平均值）	5.2%
Validation data（平均值）	4.9%
Test data Rsquare	7.2%

几个具体结果：

TA 品种， insample 取 0-20 天， outsample 取 20-25 天

OLS Regression Results

```
=====
Dep. Variable:          midPrc      R-squared:                0.059
Model:                  OLS         Adj. R-squared:           0.059
Method:                 Least Squares   F-statistic:             1.125e+05
Date:                  Tue, 19 Jul 2016   Prob (F-statistic):       0.00
Time:                  05:55:10         Log-Likelihood:          -3.0506e+06
No. Observations:      1785786         AIC:                    6.101e+06
Df Residuals:          1785784         BIC:                    6.101e+06
Df Model:               1
=====
```

```
=====
              coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----
const          0.0055      0.001      5.506      0.000      0.004      0.007
indicator2      0.6128      0.002    335.447      0.000      0.609      0.616
=====
```

```
Omnibus:          686882.454    Durbin-Watson:           0.067
Prob(Omnibus):    0.000    Jarque-Bera (JB):        176137497.063
Skew:             0.635    Prob(JB):                0.00
Kurtosis:         51.637    Cond. No.:               1.83
=====
```

[OutSample R\_SQUARE: 0.074446]

TA 品种， insample 取 0-20 天， outsample 取 25-30 天

OLS Regression Results

```
=====
Dep. Variable:          midPrc      R-squared:                0.059
Model:                  OLS         Adj. R-squared:           0.059
Method:                 Least Squares   F-statistic:             1.125e+05
Date:                  Tue, 19 Jul 2016   Prob (F-statistic):       0.00
Time:                  06:01:52         Log-Likelihood:          -3.0506e+06
No. Observations:      1785786         AIC:                    6.101e+06
Df Residuals:          1785784         BIC:                    6.101e+06
Df Model:               1
=====
```

```
=====
              coef      std err          t      P>|t|      [95.0% Conf. Int.]
-----
const          0.0055      0.001      5.506      0.000      0.004      0.007
indicator2      0.6128      0.002    335.447      0.000      0.609      0.616
=====
```

```
Omnibus:          686882.454    Durbin-Watson:           0.067
Prob(Omnibus):    0.000    Jarque-Bera (JB):        176137497.063
Skew:             0.635    Prob(JB):                0.00
Kurtosis:         51.637    Cond. No.:               1.83
=====
```

[OutSample R\_SQUARE: 0.071767]

TA 品种 , insample 取 0-15 , 20-25 天 , outsample 取 15-20 天

OLS Regression Results

```
=====
Dep. Variable:          midPrc      R-squared:                0.060
Model:                  OLS          Adj. R-squared:           0.060
Method:                 Least Squares  F-statistic:             1.149e+05
Date:                   Tue, 19 Jul 2016  Prob (F-statistic):       0.00
Time:                   05:51:59      Log-Likelihood:          -3.0126e+06
No. Observations:      1785786      AIC:                    6.025e+06
Df Residuals:          1785784      BIC:                    6.025e+06
Df Model:               1
=====
```

	coef	std err	t	P> t	[95.0% Conf. Int.]
const	0.0020	0.001	2.053	0.040	9.07e-05 0.004
indicator2	0.6086	0.002	338.948	0.000	0.605 0.612

```
=====
Omnibus:                664856.354  Durbin-Watson:           0.067
Prob(Omnibus):          0.000      Jarque-Bera (JB):        168894635.810
Skew:                   0.565      Prob(JB):                0.00
Kurtosis:               50.630      Cond. No.:               1.84
=====
```

[OutSample R\_SQUARE: 0.064212]

RM 品种 , insample 取 0-20 天 , outsample 取 25-30 天

OLS Regression Results

```
=====
Dep. Variable:          midPrc      R-squared:                0.080
Model:                  OLS          Adj. R-squared:           0.080
Method:                 Least Squares  F-statistic:             1.543e+05
Date:                   Tue, 19 Jul 2016  Prob (F-statistic):       0.00
Time:                   05:57:31      Log-Likelihood:          -1.4797e+06
No. Observations:      1782194      AIC:                    2.959e+06
Df Residuals:          1782192      BIC:                    2.959e+06
Df Model:               1
=====
```

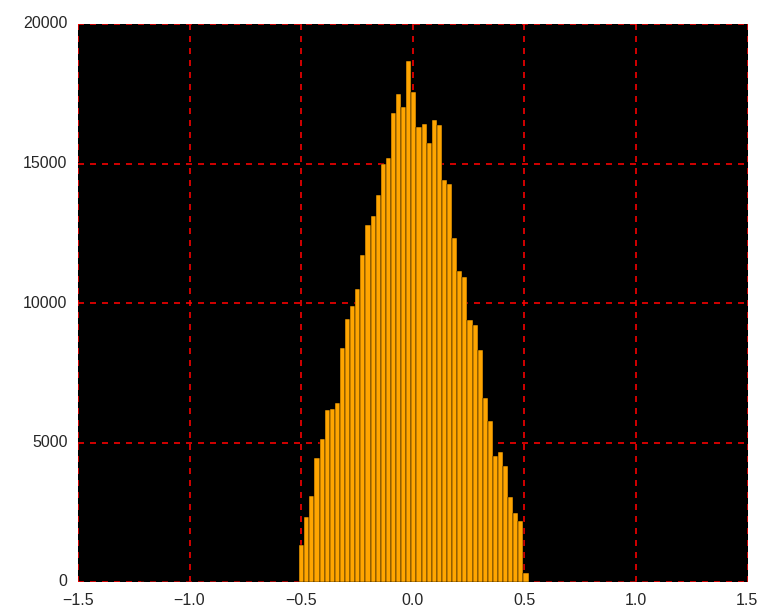
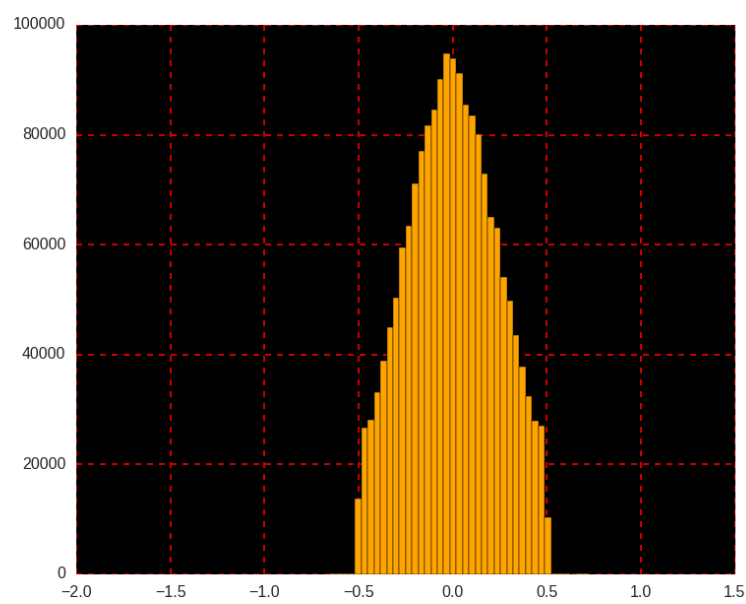
	coef	std err	t	P> t	[95.0% Conf. Int.]
const	-0.0001	0.000	-0.322	0.748	-0.001 0.001
indicator2	0.6789	0.002	392.782	0.000	0.676 0.682

```
=====
Omnibus:                621451.540  Durbin-Watson:           0.060
Prob(Omnibus):          0.000      Jarque-Bera (JB):        266577743.410
Skew:                   -0.119      Prob(JB):                0.00
Kurtosis:               62.915      Cond. No.:               4.16
=====
```

[OutSample R\_SQUARE: 0.057966]

计算中引入的额外参数：	对预测的影响（以 outsample Rsquare 衡量）
所用的档位深度	从用 1 档到用 5 档，分别为： 7.4%， 2.2%， 1.0%， 0.7%， 0.5%
不同档位间权重	调整使得权重随档位增加而递减，发现权重递减越快，效果越好。 这与上一条共同说明主要起作用的是一档，其他档位反而有负作用
是否在时间轴上做移动平均	无论 EWMA 还是 SMA，都只有负作用。 说明对于该 feature，当前 tick 的买卖双方力量情况才是关键

值的分布（上方是在 training\_data 上的分布，下方是在 validation data 上的分布）：



# Feature 4. Quote Pressure Index

## 1. Idea

- 1 Quote 的 Qty 反映了买卖方的相对强弱
- 2 对于不同的价格，相同 Qty 的订单价值 ( Prc \* Qty ) 不同。  
考虑以 Prc 与 Qty 的乘积作为衡量力量强弱的指标。Bid 方求和，ask 方求和，两者之差即为买卖方力量之差，再除以两者之和，即得到一个[-1, 1]之间取值的 indicator

## 2. Indicator 的计算

操作	理由
对于 ( 无论 bid、ask ) 每一档，求出 Prc * Qty	
Bid 各档相加得到 bid_force ； ask 各档相加得到 ask_force	考虑各档的影响 ( 具体几档好看测试结果 )
(Bid_force - ask_force) / (bid_force + ask_force)	Normalization

## 3. 测试结果

对某个参数做控制变量测试时，其他默认参数为：  
品种 TA, 只使用一档数据，只使用当前 tick 的数据 ( 不沿时间轴做平均 )。

Cross-Validation 及不同品种平均 Rsquare：

数据集	Rsquare
Training data ( 平均值 )	5.4%
Validation data ( 平均值 )	5.3%
Test data Rsquare	7.2%

几个具体结果：

TA 品种，insample 取 0-20 天，outsample 取 20-25 天

OLS Regression Results						
=====						
Dep. Variable:	midPrc	R-squared:	0.060			
Model:	OLS	Adj. R-squared:	0.060			
Method:	Least Squares	F-statistic:	1.146e+05			
Date:	Tue, 19 Jul 2016	Prob (F-statistic):	0.00			
Time:	07:16:31	Log-Likelihood:	-3.0496e+06			
No. Observations:	1785786	AIC:	6.099e+06			
Df Residuals:	1785784	BIC:	6.099e+06			
Df Model:	1					
=====						
	coef	std err	t	P> t	[95.0% Conf. Int.]	
-----						
const	0.0056	0.001	5.650	0.000	0.004	0.008
ForceIndexLevel_1	0.6288	0.002	338.491	0.000	0.625	0.632
=====						
Omnibus:	687475.038	Durbin-Watson:	0.065			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	176608744.806			
Skew:	0.637	Prob(JB):	0.00			
Kurtosis:	51.702	Cond. No.	1.86			
=====						

[OutSample R\_SQUARE: 0.075231]

TA 品种，insample 取 0-20 天，outsample 取 25-30 天

OLS Regression Results						
=====						
Dep. Variable:	midPrc	R-squared:	0.060			
Model:	OLS	Adj. R-squared:	0.060			
Method:	Least Squares	F-statistic:	1.146e+05			
Date:	Tue, 19 Jul 2016	Prob (F-statistic):	0.00			
Time:	07:15:48	Log-Likelihood:	-3.0496e+06			
No. Observations:	1785786	AIC:	6.099e+06			
Df Residuals:	1785784	BIC:	6.099e+06			
Df Model:	1					
=====						
	coef	std err	t	P> t	[95.0% Conf. Int.]	
-----						
const	0.0056	0.001	5.650	0.000	0.004	0.008
ForceIndexLevel_1	0.6288	0.002	338.491	0.000	0.625	0.632
=====						
Omnibus:	687475.038	Durbin-Watson:	0.065			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	176608744.806			
Skew:	0.637	Prob(JB):	0.00			
Kurtosis:	51.702	Cond. No.	1.86			
=====						

[OutSample R\_SQUARE: 0.072024]



SR 品种， insample 取 0-20 天， outsample 取 25-30 天

## OLS Regression Results

Dep. Variable:	midPrc	R-squared:	0.030
Model:	OLS	Adj. R-squared:	0.030
Method:	Least Squares	F-statistic:	5.490e+04
Date:	Tue, 19 Jul 2016	Prob (F-statistic):	0.00
Time:	07:18:28	Log-Likelihood:	-2.6953e+06
No. Observations:	1791180	AIC:	5.391e+06
Df Residuals:	1791178	BIC:	5.391e+06
Df Model:	1		

	coef	std err	t	P> t	[95.0% Conf. Int.]
const	0.0071	0.001	8.709	0.000	0.005 0.009
ForceIndexLevel 1	0.3806	0.002	234.307	0.000	0.377 0.384

Omnibus:	685085.645	Durbin-Watson:	0.052
Prob (Omnibus):	0.000	Jarque-Bera (JB):	493050689.592
Skew:	0.251	Prob (JB):	0.00
Kurtosis:	84.278	Cond. No.	2.00

```
[OutSample R_SQUARE: 0.056198]
```

MA 品种, insample 取 0-20 天, outsample 取 25-30 天

## OLS Regression Results

Dep. Variable:	midPrc	R-squared:	0.048
Model:	OLS	Adj. R-squared:	0.048
Method:	Least Squares	F-statistic:	8.935e+04
Date:	Tue, 19 Jul 2016	Prob (F-statistic):	0.00
Time:	07:20:52	Log-Likelihood:	-1.9043e+06
No. Observations:	1779031	AIC:	3.809e+06
Df Residuals:	1779029	BIC:	3.809e+06
Df Model:	1		

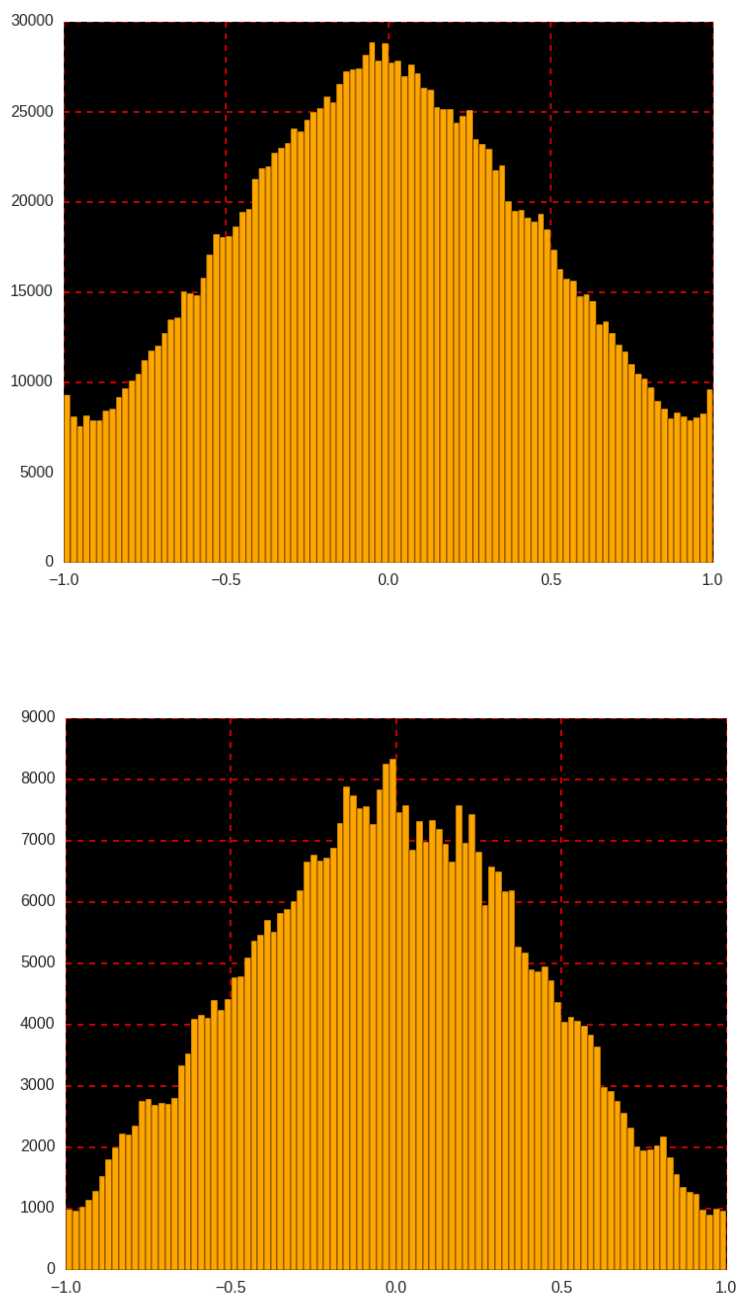
	coef	std err	t	P> t	[95.0% Conf. Int.]
const	0.0026	0.001	4.963	0.000	0.002 0.004
ForceIndexLevel 1	0.3352	0.001	298.911	0.000	0.333 0.337

Omnibus:	460106.554	Durbin-Watson:	0.061
Prob (Omnibus):	0.000	Jarque-Bera (JB):	40976297.045
Skew:	0.013	Prob (JB):	0.00
Kurtosis:	26.511	Cond. No.	2.12

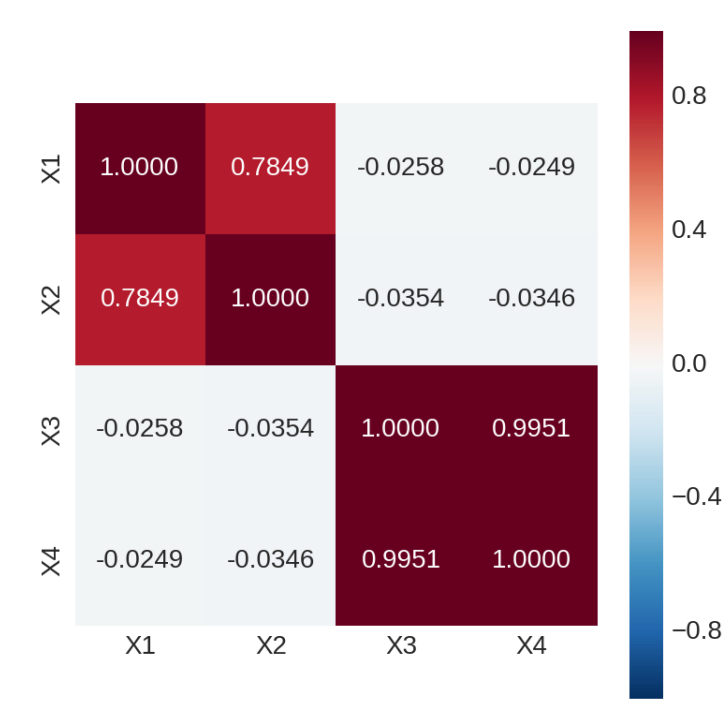
[OutSample R SQUARE: 0.086914]

计算中引入的额外参数：	对预测的影响（以 outsample Rsquare 衡量）
所用的档位深度	从用 1 档到用 5 档，分别为： 7.5%，2.3%，0.9%，0.5%，0.3%
不同档位间权重	调整使得权重随档位增加而递减，发现权重递减越快，效果越好。 这与上一条共同说明主要起作用的是一档，其他档位反而有负作用
是否在时间轴上做移动平均	无论 EWMA 还是 SMA，都只有负作用。 说明对于该 feature，当前 tick 的买卖双方力量情况才是关键（这三个的影响和 feature 3.都类似）

值的分布（上方是在 training\_data 上的分布，下方是在 validation data 上的分布）：



# Coefficient of Correlation :



X1, X2, X3, X4 分别对应四个 Feature:

X1	1-Quantity Trend Aligned with Quote Price
X2	2-Total Quote Quantity Trend
X3	3-Quote Quantity Weighted Prcie
X4	4-Quote Pressure Index

## Feature 5. Use vol/openInt as Intensity Multiplier

对于 volume 和 openInt 的使用，我的想法是将它们进行一定的变换，作为一个强度因子，乘到已有的 feature 上，用来增加准确性。

当这个强度因子比较大的时候，乘积也就大，对应较大 price move 的概率就更大。

构造方法比如：

- volume\_diff 直接 shrink：用每一 tick 间的成交量作为强度；

- volume\_diff 除以之前一段时间的 volume\_diff 的加权平均：表示这一 tick 的成交量在之前这段时间内的相对强弱。

具体做了几个测试，发现这样并不能提升已有 feature 的效果（从 rsquare 来看）。