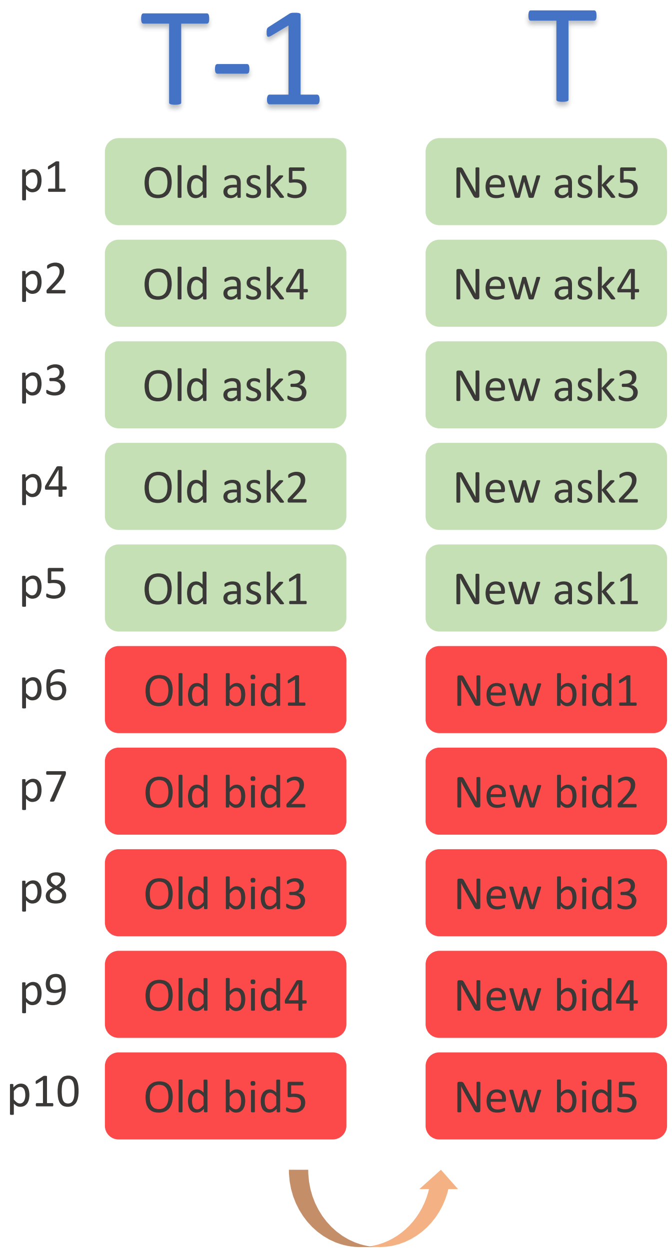
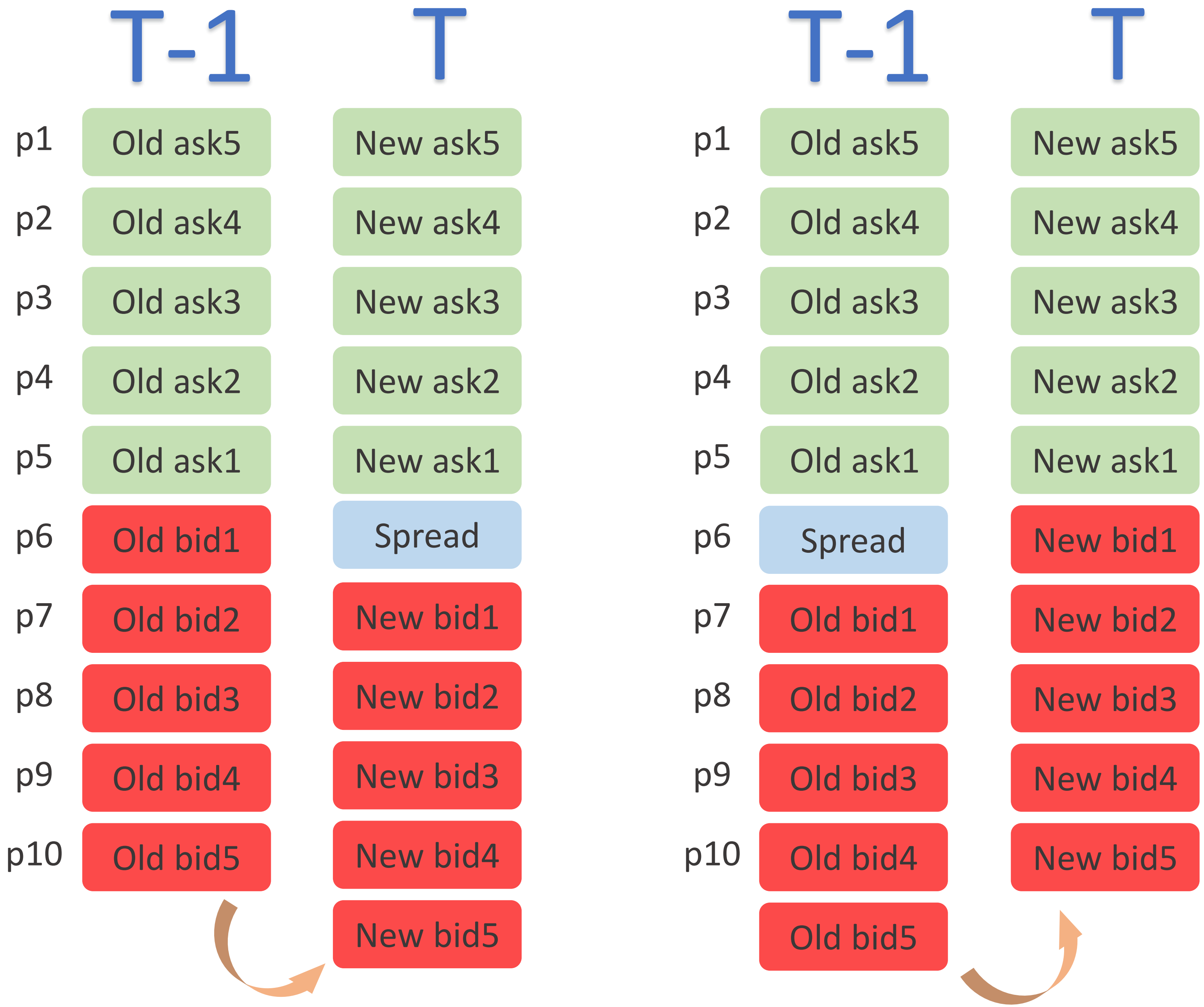
情景汇总



情景1

- 五档Asks和Bids价格均未变
- 这种情景占81.5%
- 这其中，spread为最小价格变动单位的又占到90%
- 没有发生成交的占50%

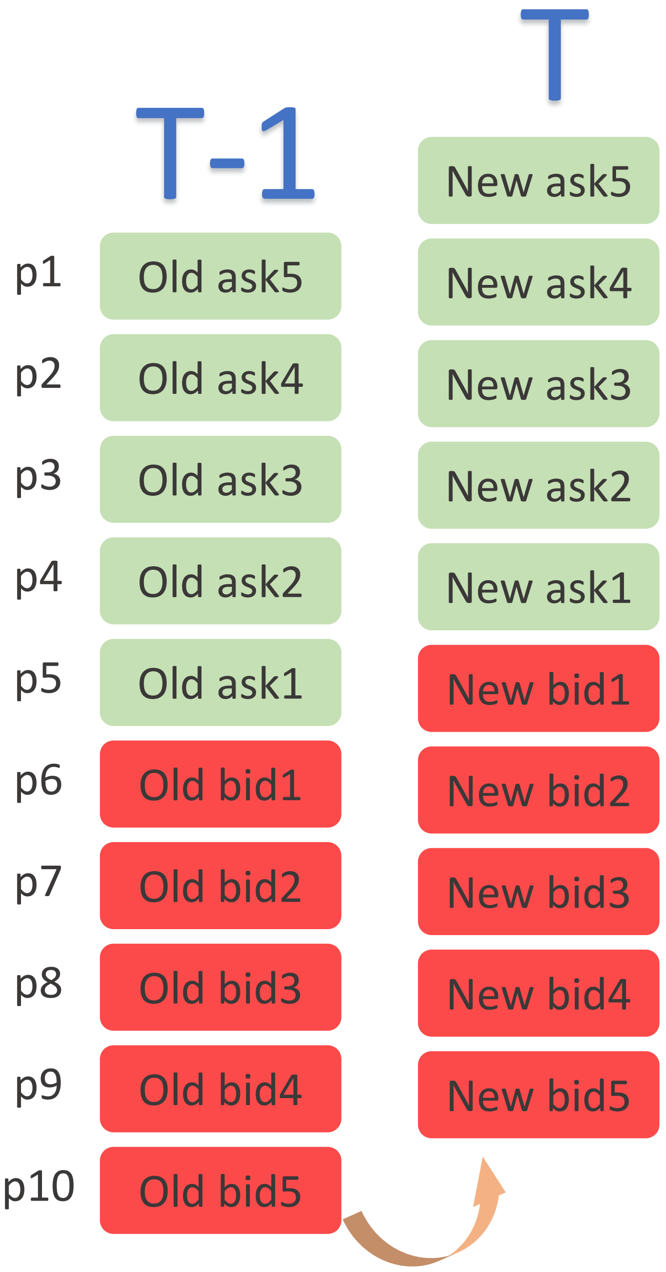
情景汇总



情景2

- 只有一边的价格发生了变化
- 只涉及到Spread拉大或缩小
- Bid变Ask 没变， 和Ask变Bid没变的，各占6%， 共计12%

情景汇总



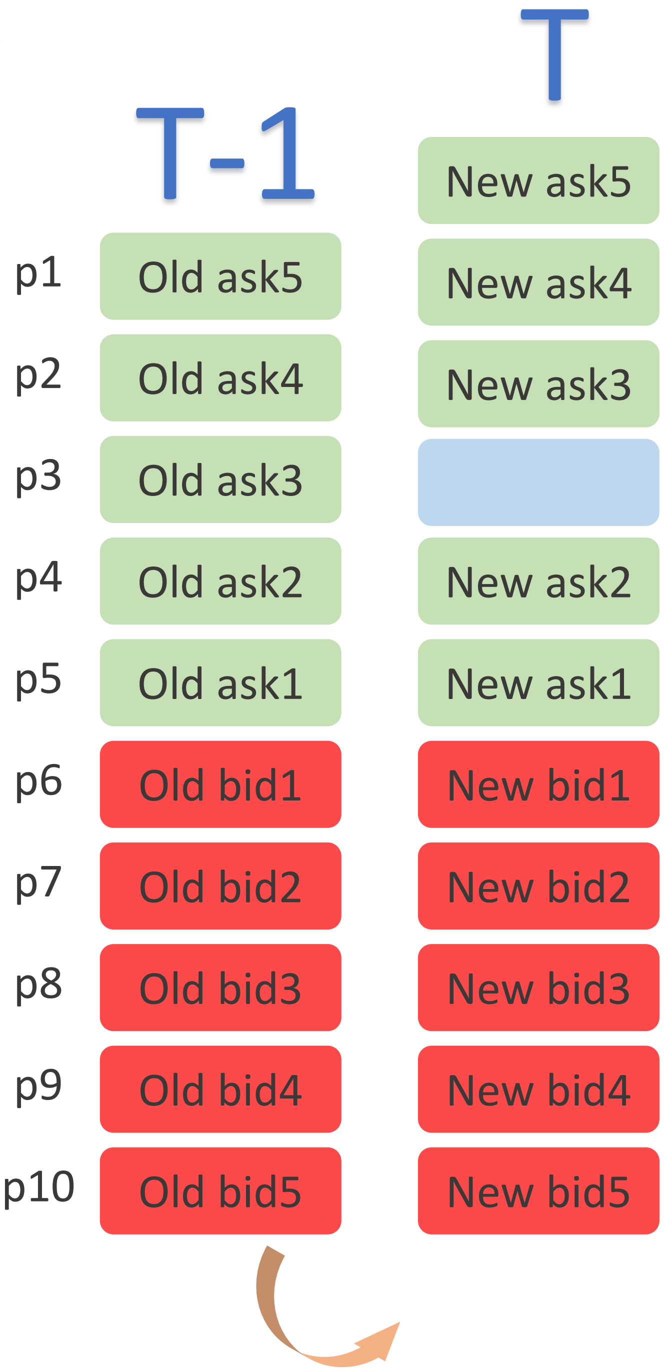
情景3

- Bid和Ask一起向一个方向平移
- 这种情景占6%

情景汇总

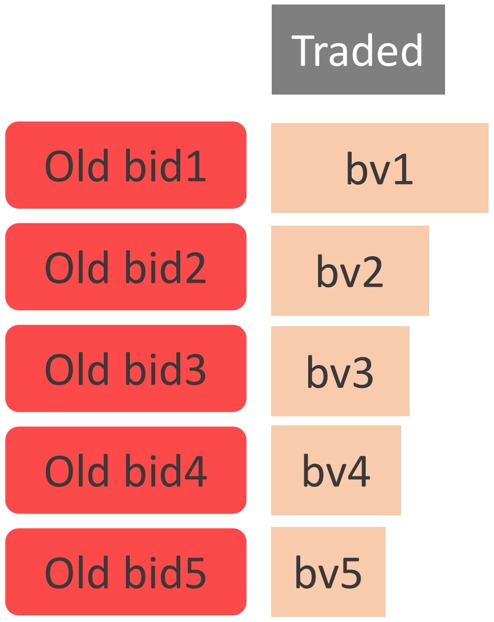
- 除了上述三种情景，还有一些情景非常少见，合计不足0.1%
- 比如Bid，Ask同时往不同方向移动，占0.02%

情景汇总



- 再比如，Ask Price1没动，但后面的某一个档位消失了，Ask被往后推了一档，这种情景占0.01%

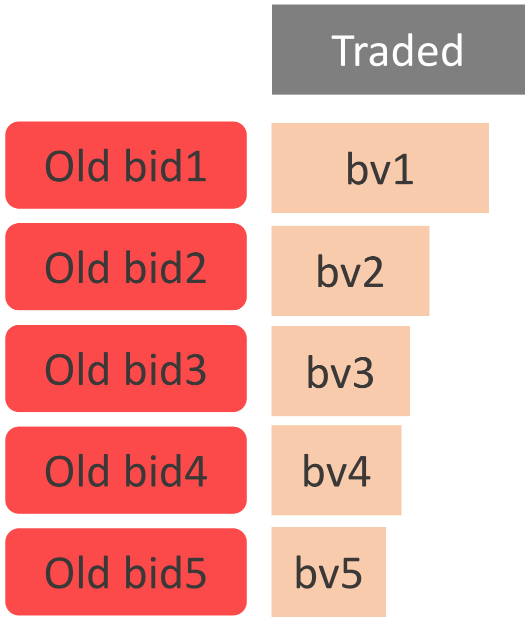
情景汇总



从成交的角度来看，

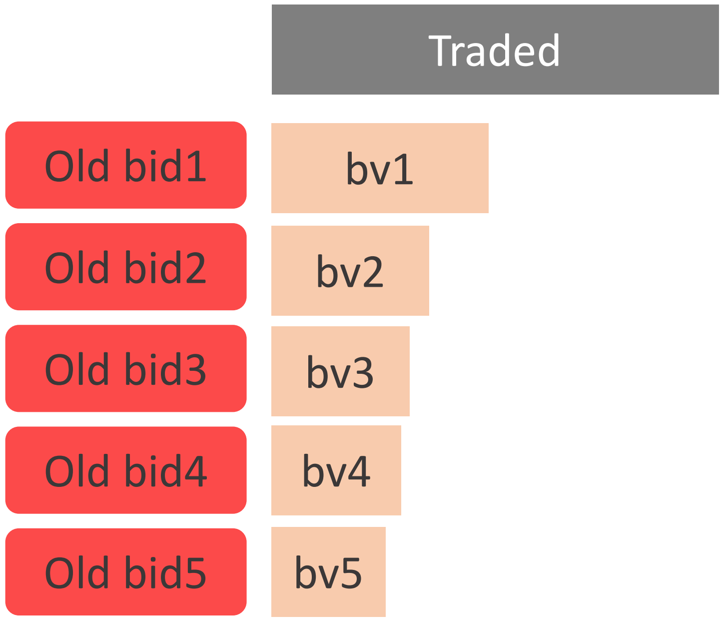
- 即使假设所有成交都是同一个方向，
- 95%的情景下，过去500ms内的成交量不够耗尽第一档的买、卖挂单

情景汇总



- 4%的情境下，成交量只够耗尽第一档买、卖，不够耗尽第二档

情景汇总



- 以此类推，成交量能耗尽前N档的占比为：
- 前零档：95%
- 前一档：4%
- 前两档：0.3%
- 前三档：0.062%
- 前四档：0.032%
- 前五档：0.023%

价格未发生变化的情景（大约80%的场景）

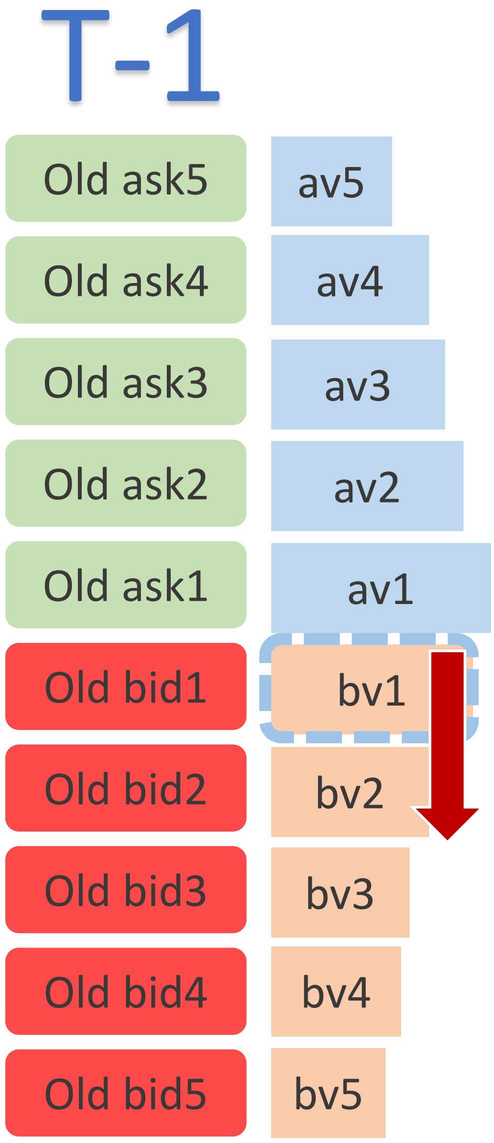
- 当价格未发生变化时，我们知道每一个价格档位的Old挂单量和New挂单量

- 如果假设只在Bid1和Ask1这两个价格发生成交, 两个方程两个未知数，有唯一解

$$\begin{cases} TradedOnBid1 + TradedOnAsk1 = Volume \\ TradedOnBid1 * Bid1 + TradedOnAsk1 * Ask1 = Turnover \end{cases}$$

- 当Bid1, Ask1上的挂单量大于500ms内的Volume， 那么假设成交单全部发生在Bid1, Ask1是合理的； 当500ms内的成交量Volume大于Old bid1, Old ask1上的挂单量，那么交易有可能不止发生在Ask1， Bid1

价格未发生变化的情景



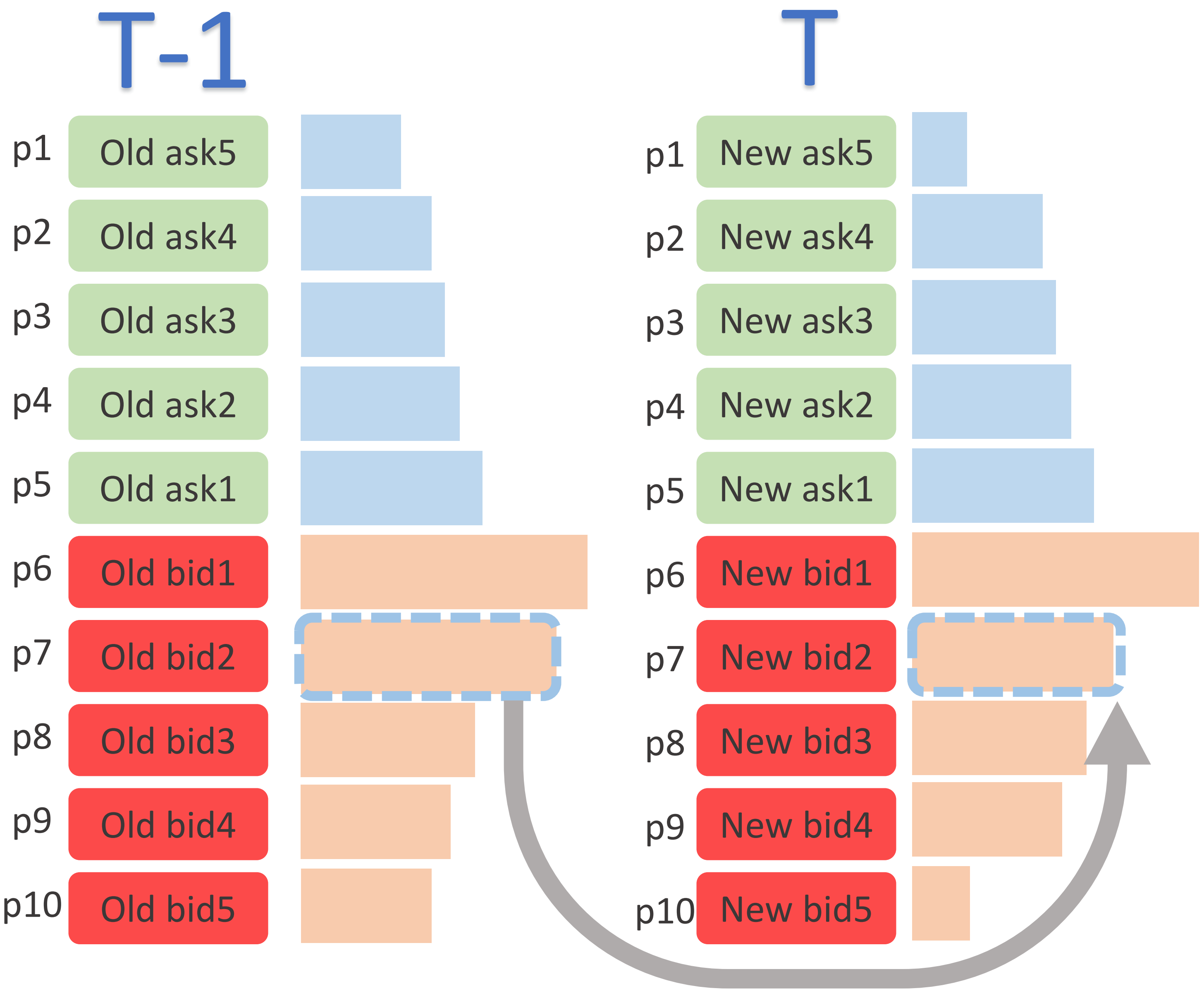
- 当500ms内的成交量Volume大于Old bid1, Old ask1上的挂单量，交易可以发生在后面的档位

- 但如果要在Old bid2发生Trade，必须先把bv1耗尽(isBid1Exhausted=True)。因此，约束条件为：要么bv1耗尽，要么Bid1之后的档位Trade量为0

- $$\begin{cases} \text{Trade@Bid1+Cancel@Bid1>OldVolume@Bid1}-M\ (1-\text{isBid1Exhausted}) \\ \text{Trade@Bid2} \leq M \cdot \text{isBid1Exhausted} \\ \text{Bid1Exhausted} \in \{0,1\} \end{cases}$$

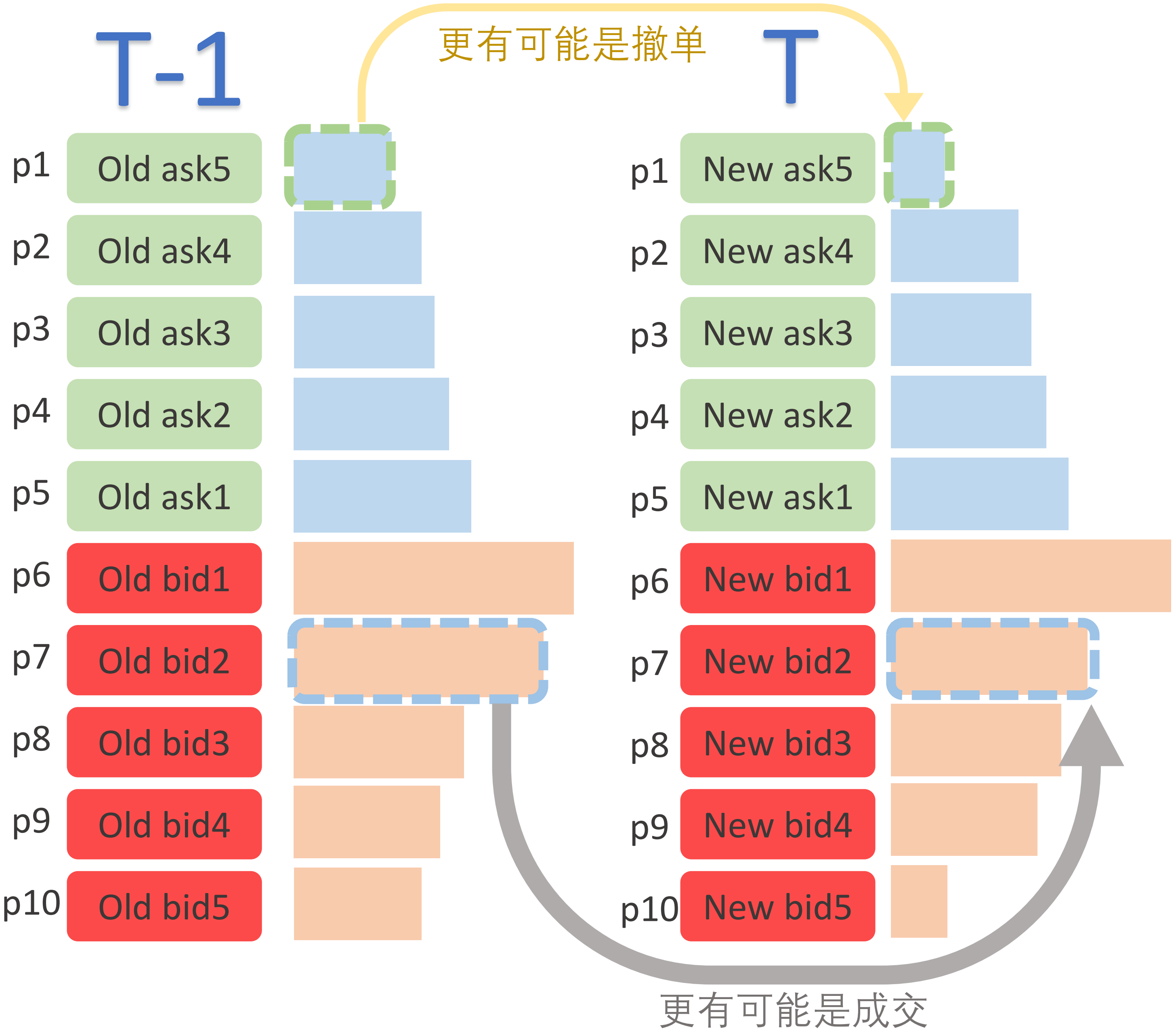
- M 是一个足够大的常数

价格未发生变化的情景



- 如果2~5档的挂单量减少，那么在此发生成交的可能性大大增加

价格未发生变化的情景



- 但是，同样是挂单量减少，Bid2上挂单量减少更倾向于是成交而非撤单，Bid5上挂单量减少则更可能是撤单

价格未发生变化的情景

- 撤单是我们无法掌控的部分，我们希望尽可能减少被归为撤单的数量。
- 我们定义加权撤单惩罚为：

$$
\begin{gathered}
\text{WeightedCanceledVolume@Bids=} \\
\text{Cancel@Bid1+} \\
0.8*\text{Cancel@Bid2+} \\
0.6*\text{Cancel@Bid3+} \\
0.4*\text{Cancel@Bid4+} \\
0.2*\text{Cancel@Bid5}
\end{gathered}
$$

$$
\begin{gathered}
\text{WeightedCanceledVolume@Asks=} \\
\text{Cancel@Ask1+} \\
0.8*\text{Cancel@Ask2+} \\
0.6*\text{Cancel@Ask3+} \\
0.4*\text{Cancel@Ask4+} \\
0.2*\text{Cancel@Ask5}
\end{gathered}
$$

价格未发生变化的情景

- 最核心的优化约束应该是:

$$
\begin{cases}
\sum_{i=1}^{5} \mathrm{Trade@}Bid_i + \sum_{i=1}^{5} \mathrm{Trade@}Ask_i = \mathrm{DeltaVolume} \\
\sum_{i=1}^{5} \mathrm{Trade@}Bid_i * Price@Bid_i + \sum_{i=1}^{5} \mathrm{Trade@}Ask_i * Price@Ask_i = \mathrm{DeltaTurnover}
\end{cases}
$$

- 变量非负约束:

$$
\mathrm{Trade@Bid_i}, \mathrm{Cancel@Bid_i} \geq 0, for\ i = 1,2,3,4,5
$$
$$
\mathrm{Trade@Ask_i}, \mathrm{Cancel@Ask_i} \geq 0, for\ i = 1,2,3,4,5
$$

- 另外新增委托也应该是非负的，由于挂单的变化量由三部分组成：新增委托-(交易+撤单)=Order变化量，因此：

$$
\mathrm{Trade@}Bid_i + \mathrm{Cancel@}Bid_i \geq \mathrm{NewVolume@}Bid_i - \mathrm{OldVolume@}Bid_i, for\ i = 1,2,3,4,5
$$
$$
\mathrm{Trade@Ask}_i + \mathrm{Cancel@Ask}_i \geq \mathrm{NewVolume@Ask}_i - \mathrm{OldVolume@Ask}_i, for\ i = 1,2,3,4,5
$$

价格未发生变化的情景

- The optimization problem：

$$\text{Minimize WeightedCanceledVolume} = \sum_{i=1}^{5} w_i * \text{Cancel}@Bid_i + \sum_{i=1}^{5} w_i * \text{Cancel}@Ask_i$$

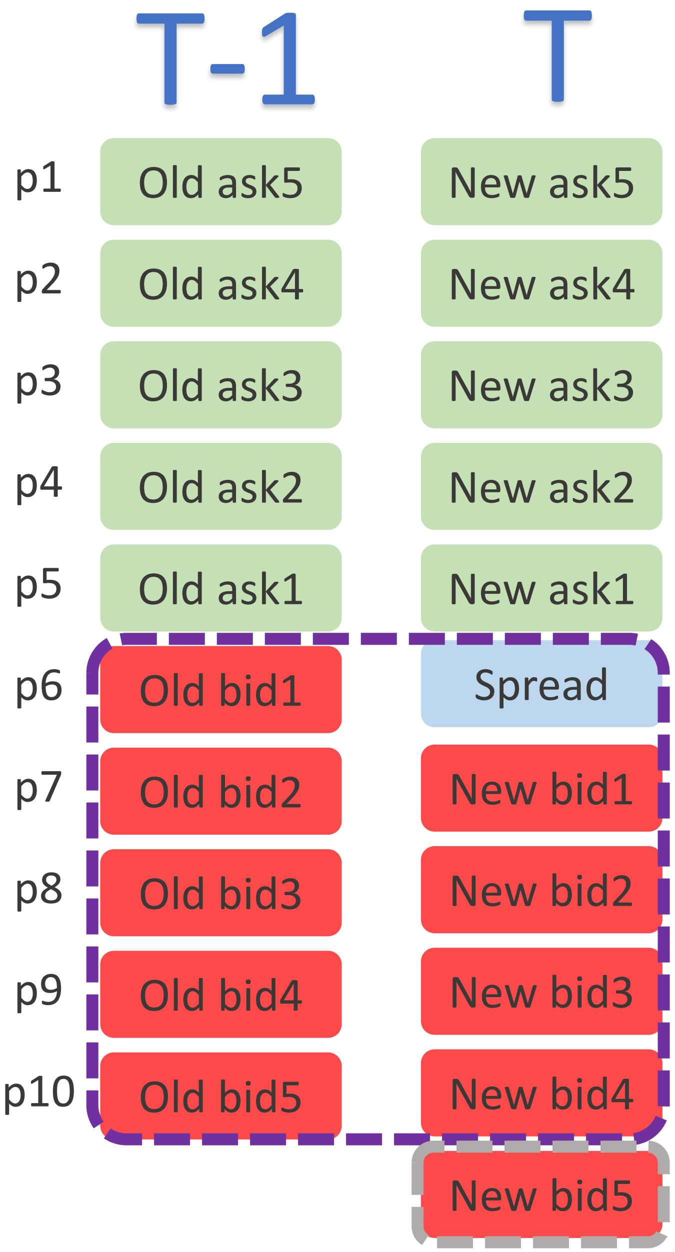
Subject to:

$$\begin{cases}
\sum_{i=1}^{5} \text{Trade@Bid}_i + \sum_{i=1}^{5} \text{Trade@Ask}_i = \text{DeltaVolume} \\
\sum_{i=1}^{5} \text{Trade@Bid}_i * \text{Price@Bid}_i + \sum_{i=1}^{5} \text{Trade@Ask}_i * \text{Price@Ask}_i = \text{DeltaTurnover} \\
\text{Trade@Bid}_i + \text{Cancel@Bid}_i + \text{NewVolume@Bid}_i - \text{OldVolume@Bid}_i \geq 0, \text{for i} = 1,2,3,4,5 \\
\text{Trade@Ask}_i + \text{Cancel@Ask}_i + \text{NewVolume@Ask}_i - \text{OldVolume@Ask}_i \geq 0, \text{for i} = 1,2,3,4,5 \\
\text{Trade@Bid}_i + \text{Cancel@Bid}_i > \text{OldVolume@Bid}_i - \text{M}\ (1 - \text{isBid}_i\text{Exhausted}), \text{for i} = 1,2,3,4 \\
\text{Trade@Bid}_i \leq \text{M} \cdot \text{isBid}_j\text{Exhausted}, \text{for i} = 1,2,3,4,5 \text{ and j} < \text{i} \\
\text{Bid}_i\text{Exhausted} \in \{0,1\}, \text{for i} = 1,2,3,4 \\
\text{Trade@Ask}_i + \text{Cancel@Ask}_i > \text{OldVolume@Ask}_i - \text{M}\ (1 - \text{isAsk}_i\text{Exhausted}), \text{for i} = 1,2,3,4 \\
\text{Trade@Ask}_i \leq \text{M} \cdot \text{isAsk}_j\text{Exhausted}, \text{for i} = 1,2,3,4,5 \text{ and j} < \text{i} \\
\text{Ask}_i\text{exhausted} \in \{0,1\}, \text{for i} = 1,2,3,4 \\
\text{Trade@Bid}_i, \text{Cancel@Bid}_i \geq 0, \text{for i} = 1,2,3,4,5 \\
\text{Trade@Ask}_i, \text{Cancel@Ask}_i \geq 0, \text{for i} = 1,2,3,4,5
\end{cases}$$

by choosing $Cancel@Bid_i, Trade@Bid_i, Cancel@Ask_i, Trade@Ask_i$

Variables in green are known or constant. Freedom variables are Trade@Bid_i and Trade@Ask_i

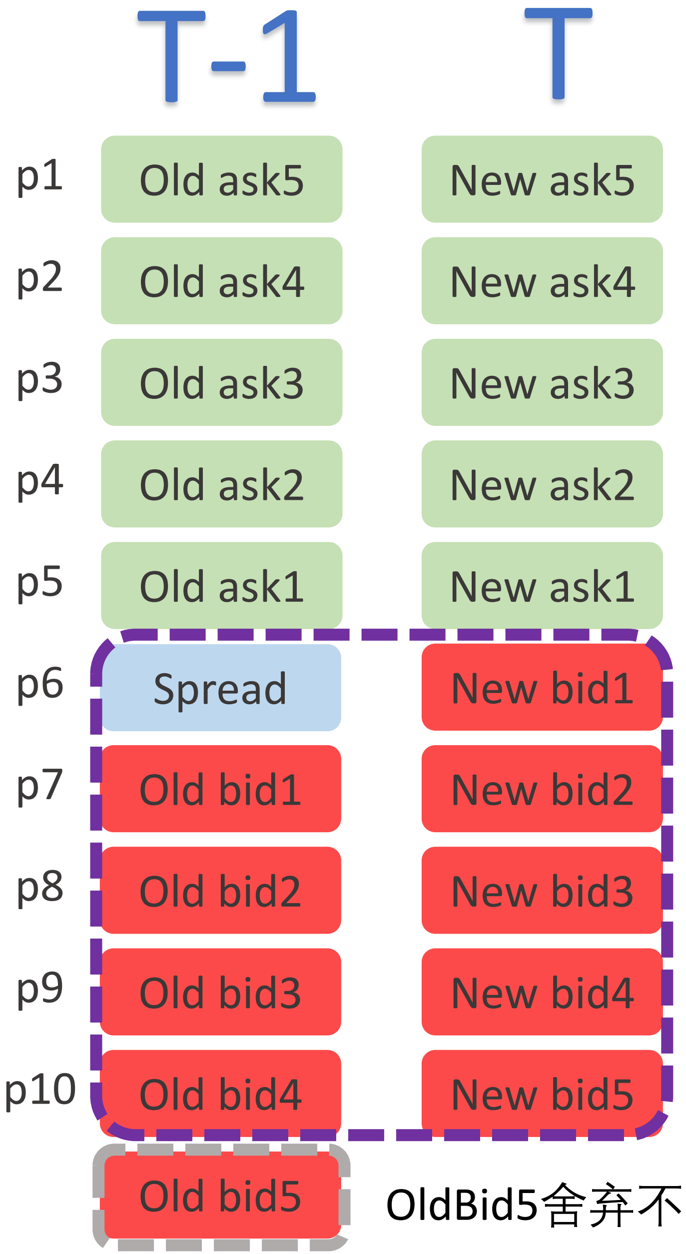
只有一边OrderBook发生变化的情景



Spread扩大的情况

Bid1按照从
OldVolumeOnBid1
变到0
后面的档位按照
Old Bid i 对应到
New Bid i-1

NewBid5舍弃不看

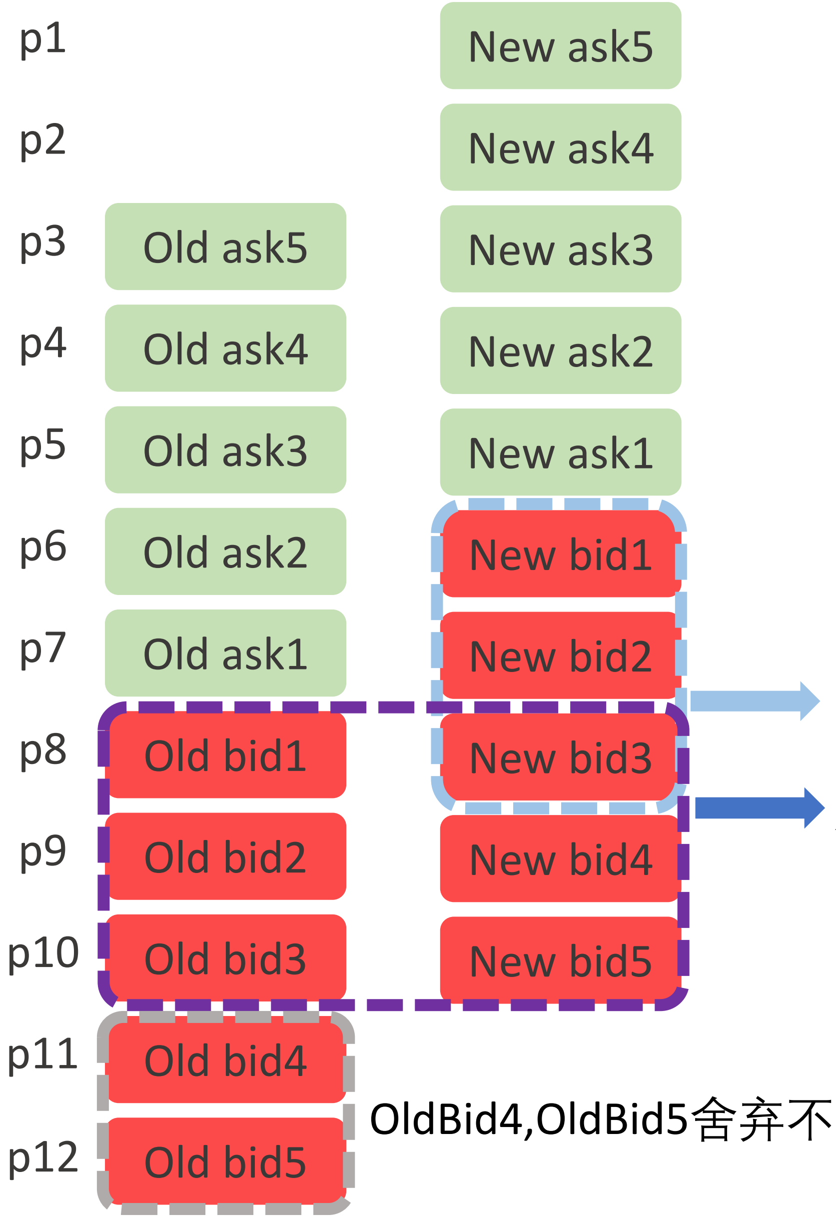


Spread收窄的情况

Bid1按照从0变到
NewVolumeOnBid1
后面的档位按照
Old Bid i 对应到
New Bid i+1

OldBid5舍弃不看

处理多个可行解

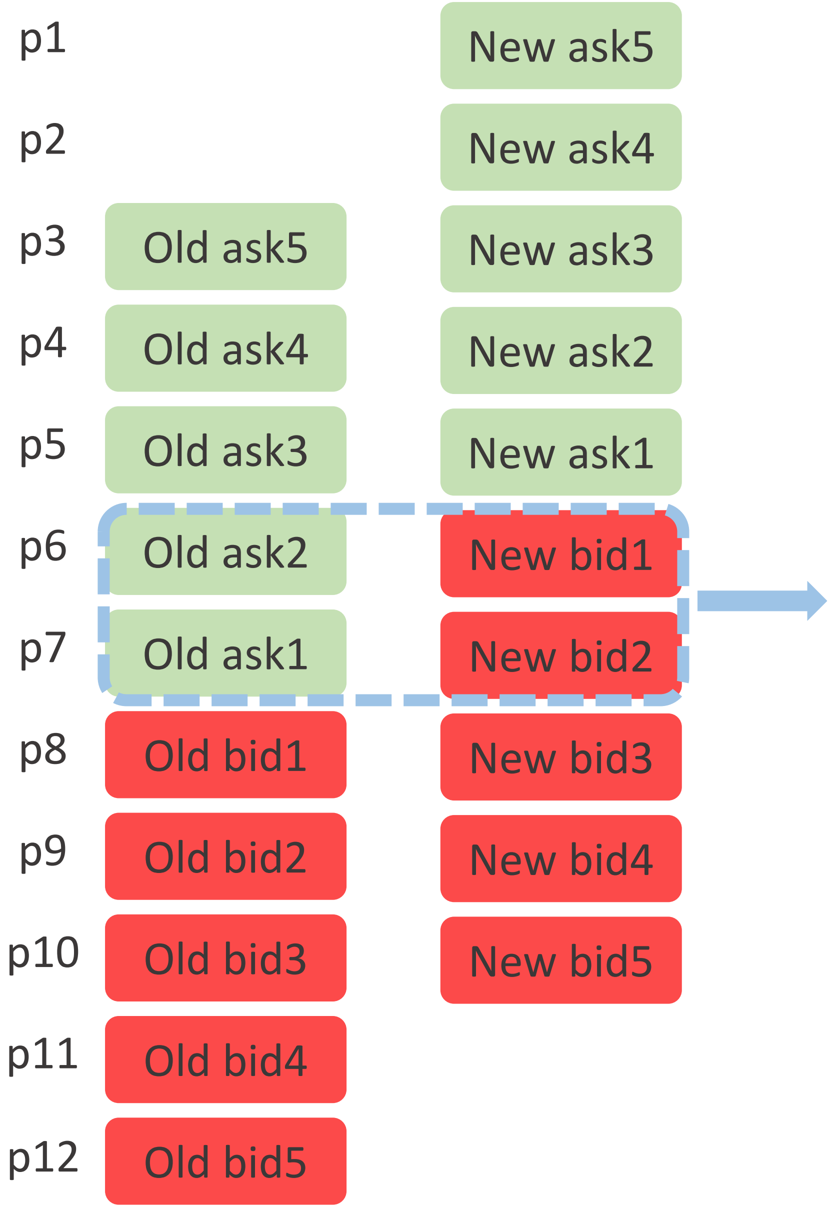


大于等于OldBidPrice1的档位上可以任意发生Trade，不限制isExhausted
因为价格可能逐步从OldBid1变到NewBid1，期间每一个档位都可能发生成交

后面的档位按照$OldBid_i$ 对应到 $NewBid_{i+k}$, $k = NewBid_1 - OldBid_1$

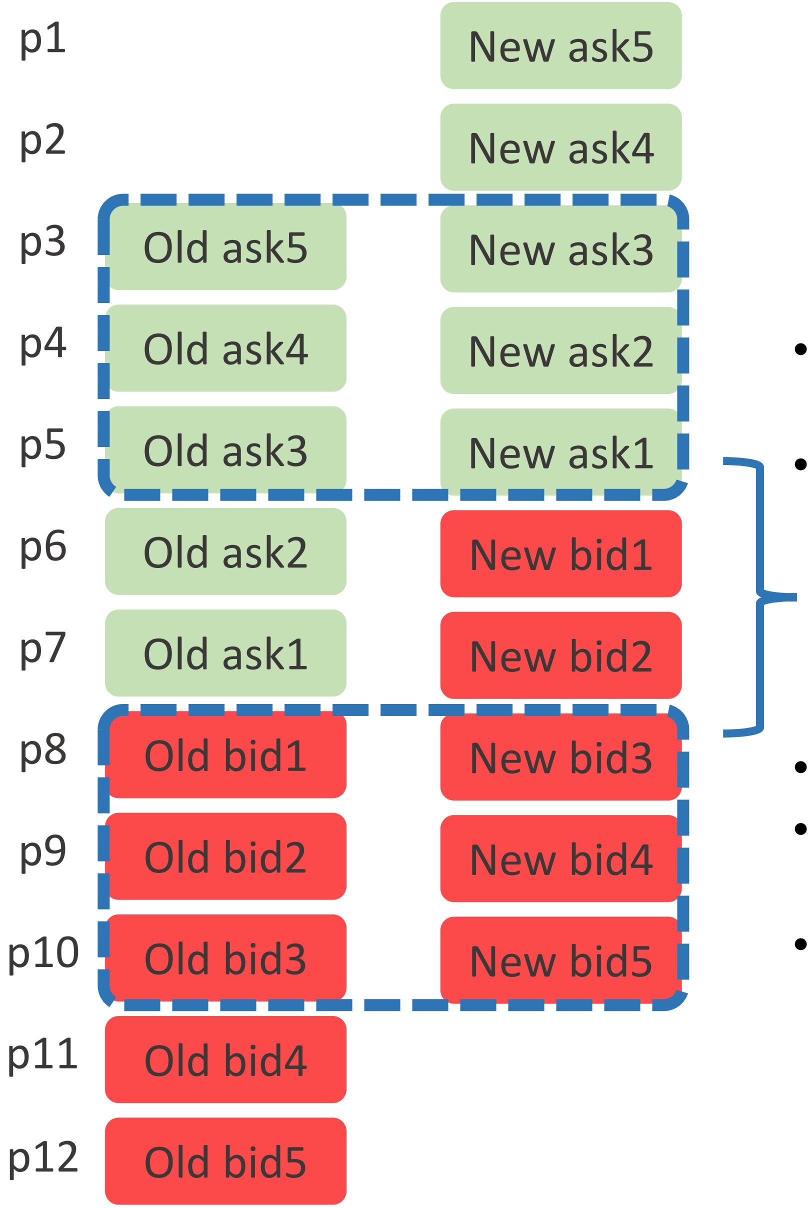
OldBid4,OldBid5舍弃不看

处理多个可行解



- 这种情况下容易出现多解
- 因为这两个价位上
  - ❖ 既有可能是主买跟ask成交的，
  - ❖ 也有可能是主卖跟bid成交的，
  - ❖ 两个方向上怎么分配量，对turnover没影响
  - ❖ 所以成为自由变量
- 但是从结果来看，这500ms内应该是主买方的势力更强
- 因此，我们先将成交量分配给OldAsk上挂单变化（这是确定性比较强的），剩下的量则优先分配给NewBid，而不是继续分配给OldAsk。
- 这一分配比较主观，可能会高估主动方向的成交。

处理多个可行解



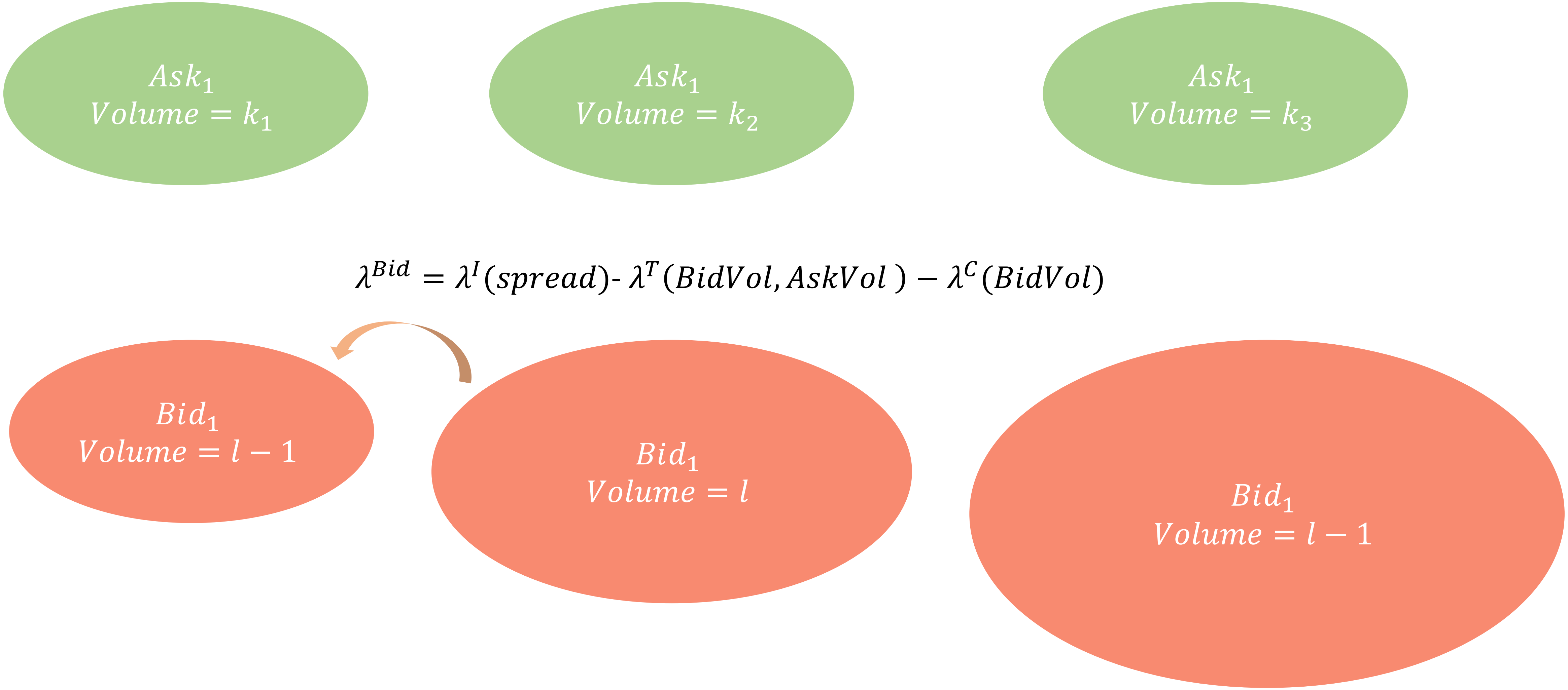
- 此外，如果在后面的档位挂单量全部是增加，Cancel是可以被优化到0的，这样的话也容易出现多解。
- 我们在Objective Function上加一个小的惩罚项：

$$Obj_{new} = Obj_{old} + 0.01 * \left(\frac{\sum w_i * Traded@Bid_i}{\sum w_i * Traded@Bid_i} + \frac{\sum w_j * Traded@Ask_j}{\sum w_j * Traded@Ask_j}\right)$$

- 取这样的i和j，使得：$Bid_i \leq OldBid_1, Ask_j \geq OldAsk_1$
- 乘以0.01的意思是说后面这项less important，我们在能优化Obj_{old}的时候优先优化Obj_{old}，Obj_{old}达到最优以后再考虑后面这项
- 其中w_i是递增的，也就是说除非NewBid高于OldBid1，否则我们尽可能分配volume给靠前的价位

将挂单量建模为连续时间非时齐的马尔可夫转移过程

考虑Bid1和Ask1上的挂单量的增加和减少过程



$$\lambda^{Bid} = \lambda^{I}(spread) - \lambda^{T}(BidVol, AskVol) - \lambda^{C}(BidVol)$$

将挂单量建模为连续时间非时齐的马尔可夫转移过程

转移概率矩阵P(t)满足：

$$
\begin{cases}
P(t) = e^{Qt} \\
Q_{l,l+1} = \lambda^I \\
Q_{l,l-1} = \lambda^T + \lambda^C \\
Q_{l,l} = -\sum_{k \neq l} Q_{l,k} \\
Q_{l,k} = 0 \quad o.w.
\end{cases}
$$

Q是连续时间马尔可夫转移过程的最小生成子

Q在时间上积分得到转移概率矩阵$P(t) = e^{Qt}$

λ^I, λ^T, λ^C 分别表示Insert, Trade和Cancel的Intensities

将挂单量建模为连续时间非时齐的马尔可夫转移过程

出于合理化的考虑，该马尔可夫转移过程应该是非时齐的，即$\lambda^I, \lambda^T, \lambda^C$不是常数

首先对于$\lambda^{\mathrm{T}}_{Bid}$，

- ❖ 它应该与bid上的Volume相关。因为bid上的Volume较大时，主动卖方倾向于等待， 反之则倾向于抢单
- ❖ 它应该与ask上的Volume相关。因为ask上的Volume较大时，新增卖单若在Ask1挂单等待则Priority较低，因此更倾向于挂aggressive单与Bid1成交
- ❖ 所以λ^T_{Bid}是$BidVol$的减函数， $AskVol$的增函数：$\lambda^T_{Bid}(BidVol \downarrow, AskVol\ \uparrow)$

对于λ^I_{Bid}，

- ❖ 它几乎是一个常数，但当$spread$不是一个tick的时候可能会有挂单挂到spread之间
- ❖ 所以λ^I_{Bid}是$spread$的减函数

对于λ^C_{Bid}，

- ❖ 它和Bid上的Volume相关。因为Volume越大，挂单后等待时间越长，那些Priority排名靠后的交易者越有可能丧失耐心而撤单。
- ❖ 所以λ^C_{Bid}是$BidVol$的增函数

因此 $\lambda^{Bid} = \lambda^I(spread)$- $\lambda^T(BidVol, AskVol\,) - \lambda^C(BidVol)$

将挂单量建模为连续时间非时齐的马尔可夫转移过程

总得来说，转移概率矩阵P(t)满足：

$$\begin{cases} P(t) = e^{Qt} \\ Q_{l,l+1} = \lambda^{I}(spread) \\ Q_{l,l-1} = \lambda^{T}(BidVol, AskVol\,) + \lambda^{C}(BidVol) \\ Q_{l,l} = -\sum_{k\neq l} Q_{l,k} \\ Q_{l,k} = 0 \quad o.w. \end{cases}$$

用MLE进行参数估计, $S = \{BidVol, AskVol, spread\}$, 发生一次Trade/Cancel/Insert为一次事件,事件之间的时间间隔为Δt_i

$$\hat{\Lambda}_i(S) = \left((\Delta t_i|S)\right)^{-1}$$

$$\hat{\lambda}_i^{I}(S) = \hat{\Lambda}_i(S)\frac{\#\{\text{Inserted}|spread\}}{\#\{spread\}}$$

$$\hat{\lambda}_i^{C}(S) = \hat{\Lambda}_i(S)\frac{\#\{\text{Canceled}|BidVol\}}{\#\{BidVol\}}$$

$$\hat{\lambda}_i^{\mathrm{T}}(S) = \hat{\Lambda}_i(S)\frac{\#\{Traded|BidVol, AskVol\}}{\#\{BidVol, AskVol\}},$$

可以这样估计是因为三个指数分布的事件中第一个发生的事件也为指数分布，lambda等于三者lambda之和

(更新：这三种事件现在都条件于最新一次mp增减的方向$LastMpChangeDirect$)

参数估计

- 需要将状态空间离散化
- 0-10为一档（占25%），10-20为一档（占20%），20-30为一档（占17%），30-40为一档（占13%），40-50为一档（占9%），50以上为一档（占16%）
- Spread分为等于10（占85%）和大于10（占15%）两种（沪铜的tick size为10）
- 共有6*6*2=72种不同的状态空间

- 右图种，从数据上可以看到Ask上Traded的强度和和Bid上的Volume呈现负相关，
- 和Ask上的Volume呈现正相关

- 用所得的lambda作为下一个500ms的Trade Volume的估计，R方为0.033

OLS Regression Results

Dep. Variable:	TradedOnOldBid1	R-squared (uncentered):	0.033
Model:	OLS	Adj. R-squared (uncentered):	0.033
Method:	Least Squares	F-statistic:	2.365e+05
Date:	Fri, 01 Nov 2024	Prob (F-statistic):	0.00
Time:	09:37:41	Log-Likelihood:	-1.9300e+07
No. Observations:	6969113	AIC:	3.860e+07
Df Residuals:	6969112	BIC:	3.860e+07
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
bid_trade_lambda	0.8880	0.002	486.284	0.000	0.884	0.892

Omnibus:	15778566.108	Durbin-Watson:	1.687
Prob(Omnibus):	0.000	Jarque-Bera (JB):	345212682353.994
Skew:	21.164	Prob(JB):	0.00
Kurtosis:	1092.514	Cond. No.	1.00

BidGroup	AskGroup	TradedOnOldAsk1
0	0	0.524579
	1	0.590058
	2	0.616785
	3	0.704849
	4	0.727197
	5	0.716479
	6	0.777757
1	0	0.410013
	1	0.501544
	2	0.610453
	3	0.677829
	4	0.652386
	5	0.645777
	6	0.700624
2	0	0.301754
	1	0.455929
	2	0.541943
	3	0.584562
	4	0.568196
	5	0.570242
	6	0.560837
3	0	0.342282
	1	0.513379
	2	0.555424
	3	0.585894
	4	0.546241
	5	0.573919
	6	0.546726

因子逻辑

- 计算完三种状态转移概率后，我们来构造因子
- 看到这篇论文： Huang W, Lehalle C A, Rosenbaum M. Simulating and analyzing order book data: The queue-reactive model[J]. Journal of the American Statistical Association, 2015, 110(509): 107-122. 说，基于两个很Trivial的假设（Volume量有上限、而且一定时间内的增长是有上限的），订单簿上的马尔可夫过程是ergodic遍历的，可以完全由
$$\rho = \frac{\lambda_I}{\lambda_T + \lambda_C}$$
决定。因此我构造了这个变量作为因子。
- 用2023年1月到6月的数据做参数估计，在7月到12月做test，结果如图。

	bid_rho_minus_ask_rho
pearson	0.098375
clipIC	0.099285
spearman	0.083863
long95	0.011468
short05	-0.001979
long99	0.001691
short01	-0.000075

因子逻辑

- 还有一种构造因子的方法是，用上述模型估计出Ask1，Bid1上的Volume被耗尽的概率（ExhaustedRatio）
- 在上面这个模型里面，只要发生一次事件，所有的转移概率都可能改变，因此没有解析解
- 所以只能用Monte Carlo做多次模拟
- 对测试次数进行敏感性分析，认为100次基本够了，再多影响速度
- 因子值使用AskExhaustedRatio减去BidExhaustedRatio

```
BidExhaustedRatio, AskExhaustedRatio = monte_carlo(31,12,10,-1,10000)
BidExhaustedRatio, AskExhaustedRatio
```
✓ 4.4s

```
(0.0217, 0.3076)
```

```
BidExhaustedRatio, AskExhaustedRatio = monte_carlo(31,12,10,-1,1000)
BidExhaustedRatio, AskExhaustedRatio
```
✓ 0.5s

```
(0.026, 0.277)
```

```
BidExhaustedRatio, AskExhaustedRatio = monte_carlo(31,12,10,-1,100)
BidExhaustedRatio, AskExhaustedRatio
```
✓ 0.0s

```
(0.02, 0.32)
```

	AskBidExhaustedRatio
pearson	0.122739
clipIC	0.124464
spearman	0.140685
long95	0.018712
short05	-0.018067
long99	0.006445
short01	0.000000

改进：Hawkes过程

- 用历史上同状态空间的样本来估计当前的lambda，不足以解释Trade量的波动
- 可以引入更短期（一分钟）的信息，来优化这个估计

- Basic Idea：过去一小段时间内发生的交易、撤单、挂单和价格变化，可能会影响后一段时间的交易、撤单和挂单的强度。具有这种特征的计数过程是self-exciting processes，或称Hawkes processes。

$$\lambda(t) = \mu + \sum_{T_i<t} \gamma(t - T_i)$$

其中$\gamma(u)$是核函数，是u的减函数，表征距离越远的事件对当前时刻的影响越小

- 核函数选用Exponential kernel:

$$\gamma(u) = \alpha\beta e^{-\beta u}$$

这个核在pandas里最好实现，调用ewm即可

- Mutually exciting processes 互激活过程
也就是说成交事件可能不仅受过去一段时间交易的影响，
还有可能收到撤单、挂单的影响

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{D} \sum_{T_j<t} \gamma_{ij}(t - T_j)$$

- 设定当mp改变的时候，认为前面所有的影响都被清空
- 这样改进完反而IC差一点

	bid_rho_minus_ask_rho_hawkes
pearson	0.080824
clipIC	0.081660
spearman	0.054487
long95	0.027870
short05	-0.002227
long99	0.004625
short01	-0.000349

改进：长周期和短周期估计的平衡

- 将Lambda改为服从一个Gaussian(μ, σ^2)，其中μ就按照之前估计的$\hat{\lambda}_i^{\mathrm{T}}(S) = \hat{\Lambda}_i(S)\frac{\#\{Traded|BidVol,AskVol\}}{\#\{BidVol,AskVol\}}$
- 而σ^2则是$\mathrm{Var}(Traded|BidVol, AskVol)$