

# 国债期货的两种潜在获利逻辑讨论

债券的定价模型为派息+面值的现值计算，即：

$$V = \sum_{t=1}^T \frac{C}{(1+r)^t} + \frac{P}{(1+r)^T}$$

我国目前国债期货有两种品种，5 年期国债期货（合约标的：面值为 100 万元人民币、票面利率为 3% 的名义中期国债）与 10 年期国债期货（合约标的：面值为 100 万元人民币、票面利率为 3% 的名义中期国债）。我认为基于定价模型，存在以下两种逻辑，可以设计出相应模型，并考察其获利能力。所有的近月合约与远月合约均选用当季与下季合约（只相差 3 个月）。理由在于隔季甚至隔两季合约的流动性总是非常差的，会导致冲击成本极大。

## 逻辑主线 1：

由于央行的货币政策预判基本基于：通胀、经济状况、就业相关的数据。也就是说，央行决策者们，在看到数据后，基于他们对数据的解读，进而制定相应的货币政策。而利率是债券最重要的影响因素。如果可以一定程度上预测利率，则可以实现对债券价格的预测。

## 模型设计：

我们虽然无法读取决策者的想法，但是数据挖掘的算法是可以找到每次货币政策与前次宏观数据之间存在的相似性的。采用 GDP、上市公司利润增速同比作为经济状况的指标，CPI、核心 CPI（由于扣除食品、能源后的 CPI 可获得的历史数据从 2013 年开始，故，采用扣除食品后的 CPI 值替代）作为通胀指标，人均可支配收入、增速，作为就业近似指标（因为缺少就业状况的公开数据），再加上 M2 与 M2 同比（不确定这两个因子是否会有用）。由于宏观经济指标都是月度数据，因此所有的输入项与输出项均为月度数据。

5 年国债期货历史数据从 2013 年开始，10 年国债期货从 2015 年开始，历史数据太少，建模过度拟合的可能性极大。由于利率与国债期货价格的极强正相关性，可替代的方案为，预测下月 Shibor 的涨跌，从而实现对国债期货价格涨跌的预测。

由于央行决策者根据宏观数据，从而制定相应货币政策，影响利率市场，因此最适合的算法应该是基于决策树为基础的算法。我们选用 RandomForest 算法（简单而言就是一种加强型复合决策树模型）对 Shibor 的预测，并不预测具体数值，只预测 Shibor 的涨跌。最终对应到国债期货的交易策略上，则为基于利率预测的，每月第一个交易日和每月最后一个交易日的相应开仓与平仓。

## 数据预处理：

由于 GDP 为季度性指标，并且 GDP 为线性平滑数据。因此，我们考虑将每季度的 GDP 值作为该季度第二个月份的 GDP 值。该季度第一、第三季度的 GDP 采用加权平均的方式获得。遇到首位数据无法获得加权平均值的，直接用当月 GDP 值代替。

首先需要测试 Shibor 的哪个品种与国债期货相关性最高。理论上来说相关性应该都是非常高的。

	TF00C1	ShiborO/N	Shibor1W	Shibor1M	Shibor3M	Shibor6M	Shibor1Y
TF00C1	1						
ShiborO/N	-0.51878	1					
Shibor1W	-0.7623	0.870773	1				
Shibor1M	-0.74071	0.829437	0.944302	1			
Shibor3M	-0.82696	0.701781	0.879617	0.892298	1		
Shibor6M	-0.80013	0.630638	0.791652	0.816508	0.968324	1	
Shibor1Y	-0.82653	0.618112	0.788224	0.806855	0.952204	0.992672	1
	T00C1	ShiborO/N	Shibor1W	Shibor1M	Shibor3M	Shibor6M	Shibor1Y
T00C1	1						
ShiborO/N	0.077134	1					
Shibor1W	-0.41716	0.734265	1				
Shibor1M	-0.54281	0.674467	0.900583	1			
Shibor3M	-0.5867	0.61785	0.825175	0.957605	1		
Shibor6M	-0.60287	0.534049	0.747055	0.897802	0.982143	1	
Shibor1Y	-0.62391	0.423918	0.712023	0.84779	0.945705	0.98366	1

可以发现 Shibor1 年与 10 债 5 债期货价格负相关性最高。因此，我们就选择 Shibor1 年作为模型的预测目标。Shibor1 年的数据，每月收盘利率减去上月收盘利率，如果得到负值，则标记为利率下行，如果正值，则标记为利率上行，如果得到零，则参考 Shibor 半年在该月的涨跌值，假如还是为零，则参考 Shibor3M 的变化值。这一顺序的原因在于，与二者与债券相关度差不多，但 6M 与 1 年的相关度高于 3M。如果还是出现利率变化为零，则参考 Shibor1M，Shibor1W。此外，一定要选出利率是上行了还是下行了的原因在于，哪怕基准利率是不变的，但市场上的利率一定不会是固定的，至少也有极其微小的变化。这就足以引起债券价格的波动。

在未来国债期货的历史数据足够多的时候，我们可以考虑直接以国债期货价格作为预测目标，这样相应的获利能力应该会更（因为除了利率本身，还有风险偏好、资金喜好的问题，从而导致利率的变动无法 100% 等价于反向债券价格变动）。

### 具体操作与结果：

我们将历史数据随机分成两组，一组大小为 80% 作为模型训练组，一组大小为 20% 作为模型精度测试组。

经过不同模型参数的测试，得到的最优模型的树级别为 4。该参数的模型输出评测为：

```

Accuracy : 0.9167
95% CI : (0.73, 0.9897)
No Information Rate : 0.6667
P-Value [Acc > NIR] : 0.004871

Kappa : 0.8235
McNemar's Test P-Value : 0.479500

Sensitivity : 1.0000
Specificity : 0.8750
Pos Pred Value : 0.8000
Neg Pred Value : 1.0000
Prevalence : 0.3333
Detection Rate : 0.3333
Detection Prevalence : 0.4167
Balanced Accuracy : 0.9375

'Positive' Class : down

```

模型的精度很惊人，达到了 91.67%，其他几项指标（实在不知道如何翻译）也达到了相当

高的精度。可惜的是，由于 RandomFores 算法的特性（构建多重决策树，最后叠加各平级决策树的输出结果），我们没法获得最终该模型究竟使用了我们输入的哪几个因子，每个因子的影响权重如何。故而无法对模型进行进一步的解读。在历史数据越来越多的时候，可以进行进一步的测试，以确定到底是模型真正有效，还是有历史数据建模中过度拟合的可能，又或者是随机分配出的测试组碰巧是一组幸运的采样。

不考虑仓位，止盈止损。仅仅在预测结果出来后的下月第一个交易日套用模型预测对当月的两种国债期货进行建仓，并在月底最后一个交易日平仓，交易测试结果：

交易时间	TF00C1	T00C1	预测下月 利率	根据预测交易 TF获利点数	根据预测交 易T获利点数
2017-01-26	99.0750	95.8850			
2016-12-30	99.4000	97.3200	up	0.325	-1.435
2016-11-30	100.7100	100.0200	up	1.31	-2.7
2016-10-31	101.7050	101.2450	up	0.995	-1.225
2016-09-30	101.6250	101.1900	up	-0.08	0.055
2016-08-31	101.8100	101.1300	down	-0.185	0.06
2016-07-29	101.4750	101.0650	down	0.335	0.065
2016-06-30	101.0100	100.3900	down	0.465	0.675
2016-05-31	101.0400	100.0000	down	-0.03	0.39
2016-04-29	100.6000	99.5150	up	-0.44	0.485
2016-03-31	101.0100	99.8000	down	-0.41	-0.285
2016-02-29	101.3500	100.7000	down	-0.34	-0.9
2016-01-29	100.7900	100.3650	down	0.56	0.335
2015-12-31	100.6800	100.0350	down	0.11	0.33
2015-11-30	100.5950	99.3500	up	-0.085	0.685
2015-10-30	99.9150	98.8300	down	0.68	0.52
2015-09-30	99.2150	97.3150	down	0.7	1.515
2015-08-31	97.6500	97.0000	up	-1.565	0.315
2015-07-31	96.6050	95.5450	up	-1.045	1.455
2015-06-30	95.9600	94.7550	up	-0.645	0.79
2015-05-29	95.6550	94.9200	down	0.305	-0.165
2015-04-30	98.4600	97.9550	down	-2.805	-3.035
2015-03-31	96.9000	95.6500	down	1.56	2.305
2015-02-27	98.3940		down	-1.494	
2015-01-30	97.2980		down	1.096	
2014-12-31	96.7720		up	-0.526	
2014-11-28	96.9000		up	0.128	
2014-10-31	95.5340		down	1.366	
2014-09-30	94.7520		down	0.782	
2014-08-29	93.3860		down	1.366	
2014-07-31	93.1580		down	0.228	
2014-06-30	94.4280		down	-1.27	
2014-05-30	94.1680		down	0.26	
2014-04-30	92.8240		down	1.344	
2014-03-31	92.5460		down	0.278	
2014-02-28	92.5900		down	-0.044	
2014-01-30	92.2540		up	-0.336	
2013-12-31	91.8020		up	-0.452	
2013-11-29	91.7820		up	-0.02	
2013-10-31	93.7400		up	1.958	
2013-09-30	94.4620		up	0.722	
			<b>总点数</b>	<b>5.101</b>	<b>0.235</b>

**逻辑主线 2:**

一个国家所有市场中利率应该是统一的。一旦出现不同品种隐含的利率不统一，就存在套利机会。

**模型设计:**

10 年期国债远近合约的差价 D10, 与 5 年期远近合约的差价 D5, 由于二者票面利率均为 3%, 因此 D10 与 D5 同时指向的是近 3 个月的利率走向 (可近似看做 3 个月的票息价值在近 3 个月利率环境下的折现值)。那么意味着二者价格应该存在极强的相关性, 或者说总是有稳定的差值存在。一旦脱离, 就会有套利机会。我们可以统计出当月 10 年期国债与下月 10 年期国债的价差 D10, 当月 5 年期国债与下月 5 年期国债的差价 D5。比较 D10 与 D5 的历史数据的差值关系 DD。一旦 DD 脱离其正常区间 (或者是脱离区间立刻建仓 1/3 两个西格玛区间时加仓 1/3, 3 个西格玛时, 重仓), 则进行对冲套利。

**具体操作与结果:**

作为非机构用户, 可获得的历史数据具有局限性。理想状态下, 最好是测试所有历史数据的 1 分钟数据甚至每笔成交数据。从而得出精确的所有套利机会。本文采用可获得的所有 1 分钟历史数据 (最近 20 个交易日), 以及所有可获得的每日历史数据。通过数据整理、比对, 发现 10 年期国债远近合约差价与 5 年期国债远近合约差价确实存在正态分布:

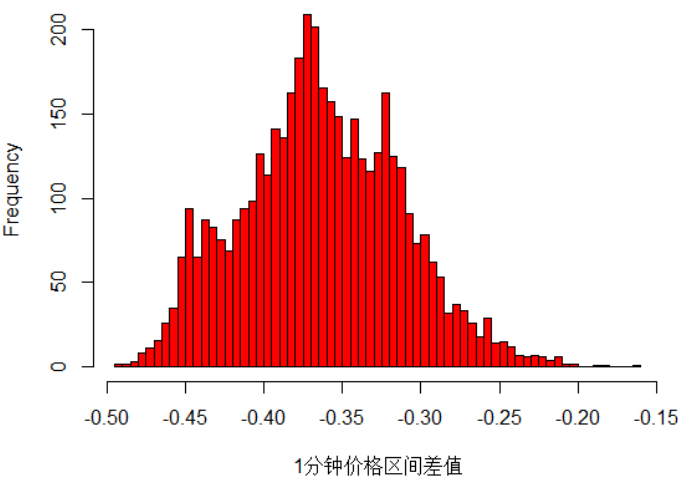
1 分钟数据:

10 债远近合约差价与 5 债远近合约差价的差值	
Mean	-0.361162731
Standard Error	0.000773458
Median	-0.365
Mode	-0.375
Standard Deviation	0.050836819
Sample Variance	0.002584382
Kurtosis	-0.202246763
Skewness	0.173416196
Range	0.331
Minimum	-0.495
Maximum	-0.164
Sum	-1560.223
Count	4320
Confidence Level(95.0%)	0.001516374

两个西格玛区间:

-0.46283637      -0.259489093

10债远近合约差价与5债远近合约差价的差值



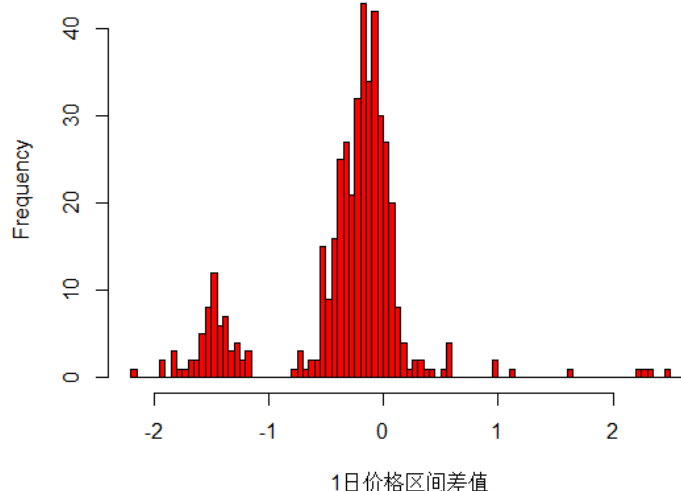
10 债远近合约差价与5 债远近合约差价的差值

Mean	-0.309393258
Standard Error	0.028352106
Median	-0.19
Mode	-0.2
Standard Deviation	0.598088333
Sample Variance	0.357709654
Kurtosis	4.904776275
Skewness	0.138712381
Range	4.83
Minimum	-2.195
Maximum	2.635
Sum	-137.68
Count	445
Confidence Level(95.0%)	0.055720998

两个西格玛区间

-1.505569924 0.886783407

10债远近合约差价与5债远近合约差价的差值



因此，站在长线角度，当“10 债远近合约差价与 5 债远近合约差价的差值”大于 0.886783407 时，可以考虑做空这一差值，即，空 10 债下季 + 多 10 债当季 + 空 5 债当季 + 多 5 债下季 + 多 10 债下季 + 空 10 债当季 + 多 5 债当季 + 空 5 债下季。当“10 债远近合约差价与 5 债远近合约差价的差值”小于 -1.505569924 时，可以考虑做多这一差值。即，多 10 债下季 + 空 10 债当季 + 多 5 债当季 + 空 5 债下季。

同理可用于 1 分钟级别的短期交易（最近 2 月的区间在 -0.46283637，0.886783407）。但是需要注意的是，由于 1 分钟级别的分布区间来源于短期的历史数据统计，仓位必须较低。此外，需要定期（比如每月重大宏观数据公布后）检测、修正短期波动区间上下沿的变化。

具体交易策略，初步的设想，以做空差值为例：在均值+2 个西格玛处下单建仓，挂 3 个西格玛+10 点止损。如果价格到达均值+ 3 个西格玛，加仓，止损位置不变。在均值+一个西格玛处止盈所有仓位的三分之一仓位。并在均值+一个西格玛+10 点处挂止损。当价格回到均值时，止盈剩余仓位的一半，均值+10 点处挂止损。假如差值还继续缩小，甚至向区间的左侧移动。则在均值 - 一个西格玛处止盈全部仓位。具体还需要有合适的回测软件进行精细的调试，尤其是每次的止损点数需要更精确的统计测算。

### 后续值得继续延伸的方向：

根据成交额，探讨资金容量。

获得更加细化的历史成交数据，例如 1 分钟数据甚至粉笔成交。并寻找到合适的回测软件，调试具体交易细节。从而得出以上两个模型的实际潜在收益率。

此外，逻辑 1 中的 RandomForest 模型在历史数据越来越多的时候，可以进行对债券价格变动的直接预测，而不必通过对利率的预测间接预测债券价格，从而提高策略的精度。