

对股票可转债的一个探究性分析

张城

目录

- 一、 选题简介2
 - 1) 可转债介绍2
 - 2) 项目介绍2
- 二、 数据获取2
- 三、 数据清理与预处理3
- 四、 数据分析及可视化3
 - 1) 分类与统计分析3
 - a.分布与分类关系.....3
 - b.涨跌幅研究.....5
 - c.线性回归分析6
 - d.其他相关分析.....10
 - 2) 计算可转债内在价值11
 - 3) 投资策略建议15
- 五、 总结与展望16
- 附录：全部代码（含注释）17

一、 选题简介

1) 可转债介绍

可转债是一种可以转换为公司股票的债券，因为可转债既可以当作普通债券，获取固定收益；也可以转换为公司股票，获得风险收益，兼具攻守性，因此受到投资人的喜爱。

可转债具有以下优点：

1. 由于可以转换为股份，公司往往不需偿付。在股价持续低迷时还可以下调转股价促使投资者转股，进一步避免偿付，使得可转债相比普通债券违约风险更低
2. 部分可转债具有转股负溢价，存在直接或间接的套利空间
3. 存在回售条款，保护投资者
4. 不设涨跌幅限制，t+0 交易，相比对应股票交易更加灵活

但可转债也有以下缺点：

1. 作为债券，利率普遍较低
2. 下调转股价需要股东大会审议，可能难以通过，仍存在违约风险
3. 流动性往往不如股票，不适合大资金操作
4. 由于不设涨跌幅限制，可能使投资者承担更大的波动风险
5. 强制赎回条款一定程度上限制了投资者选择
6. 持有的可转债仓位不能作为打新额度
7. 涉及的规则条款复杂，一般投资者入门困难

2) 项目介绍

我的项目主要是爬取可转债与其对应股票的数据，并研究其统计规律，进行可视化；另外利用 **Black-Scholes** 模型计算可转债期权价值，从模型得到的低估情况和获得的统计结论提供投资建议，并反思模型的问题，做出总结和展望。

二、 数据获取

数据获取采用了 R 语言爬虫的方法，爬取主要分为两部分：

首先是在东方财富这一网站上获得可转债的大部分数据，使用了 **selenium** 以获得动态数据，由于我爬取的网站有两页，但两页的 **url** 相同，所以使用了 **selenium** 进行模拟点击的方法，以获得两页的数据。另外，模拟点击时发现了一个问题，就是翻到第二页后，立刻读取会读取到第一页的 **url**，可能是因为翻页速度过快，短时间网站没有响应。通过翻页后停顿一秒钟再读取的手段，可以解决这个问题。

但是由于该网站不能得到股票的波动率，分析时缺少必要的参数，之后我又进行了第二部分的爬取工作，这一部分是在网易财经（静态网页）上进

行的。由于该网站的 url 存在规律,于是在第一阶段获得股票代码的基础上,用拼接的方式,获得 url,爬取相应股票历史涨跌幅的数据,转化为数值型并计算其标准差即波动率,再将这一数据加入数据框。最后将数据保存为各种格式,进行下一部分的工作。

三、 数据清理与预处理

由于爬取的数据没有反映本来类型且存在缺失值,要进行清理与预处理:对于百分比数据,选择去除百分号并转化为数值型,再除以 100 变成小数;对于普通数值数据,直接转化为数值型会出现结果失真的情况,所以先转化为字符型再转化为数值型。对于日期数据,将其转化为日期型。最后去除含有缺失值的行(这部分主要是尚未上市交易的可转债)。

除此之外,为了可视化方便,选择将可转债名字中重复出现的“转债”字样去除,之后根据上市日期和当前日期(2018 年 5 月 27 日)计算可转债到期时间,并添加一列,以便于分析。

四、 数据分析及可视化

1) 分类与统计分析

a.分布与分类关系

首先作直方图,观察转股溢价率(conprerate),纯债溢价率(bondprerate),到期时间(t),波动率(move)这几个重要变量的分布关系

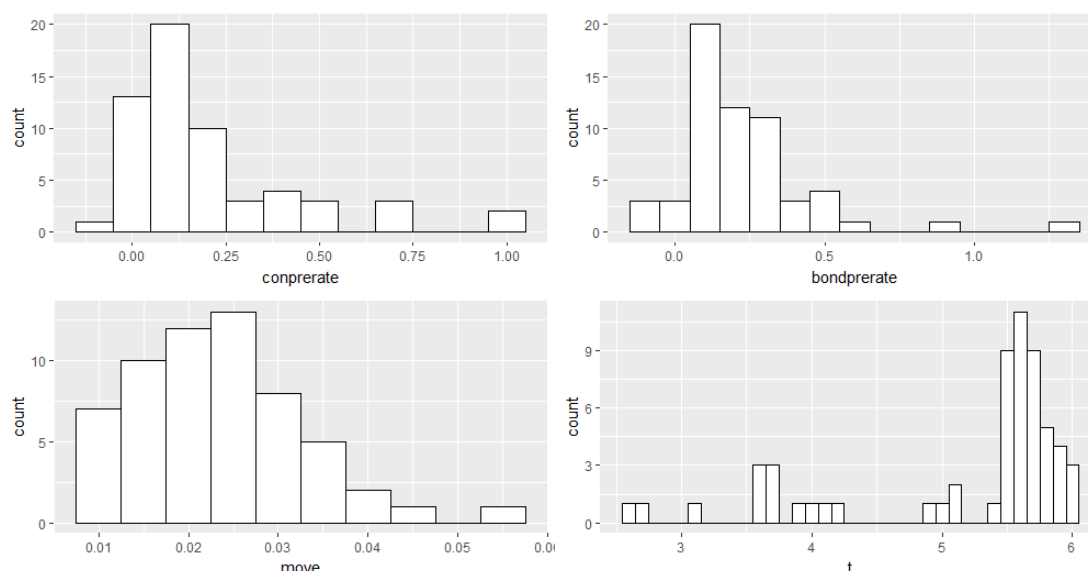


图 1 几个变量的直方图

可以发现转股溢价率和纯债溢价率在 10%附近的最多,且大多数在-5%-25%之间,波动率基本在 0.01-0.03 左右,到期时间以 5.5 年-6 年的居多,也有部分在 3 年,4 年或 5 年左右的,分布较为分散。接下来对可转债一些变

量的规律进行分析,首先是主观经验上相关性较强的“可转债涨跌幅”和“对应股票涨跌幅”的之间的关系。

由于可转债既可以兑付本息，也可以转股，其具有债券和股票期权两种性质，所以考虑先将其分为“债性为主”和“股性为主”两类，再分别分析。

具体分类方法是根据转股溢价率和纯债溢价率，进行 k-means 聚类分析，因为转股溢价率低，纯债溢价率高的可转债往往转股收益较大，呈现出股性；而纯债溢价率低，转股溢价率高的可转债往往兑付收益较大，呈现出债性。

对预处理后的 59 只可转债进行分类，发现 46 只可转债以股性为主，13 只以债性为主，进行可视化输出，如图：

图 2 可转债分类图

在此基础上添加可转债名字,得:(由于中间点过密,可视化效果并略差)

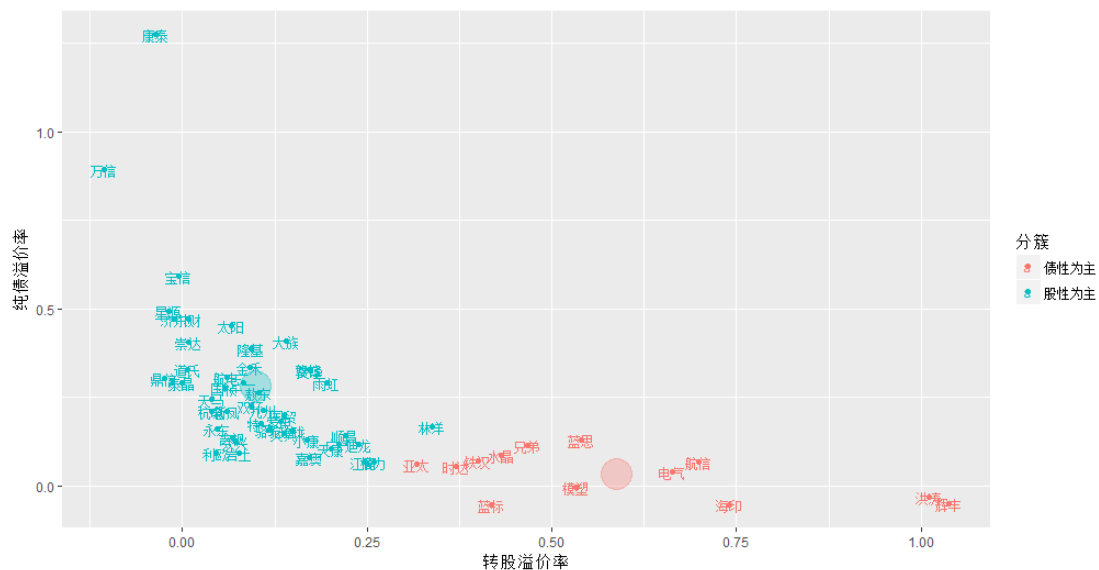


图 3 加入名字标签的可转债分类图

b.涨跌幅研究

分类后分别对两类可转债的涨跌幅与对应股票股价涨跌幅作出散点图并添加拟合线，进行比较：

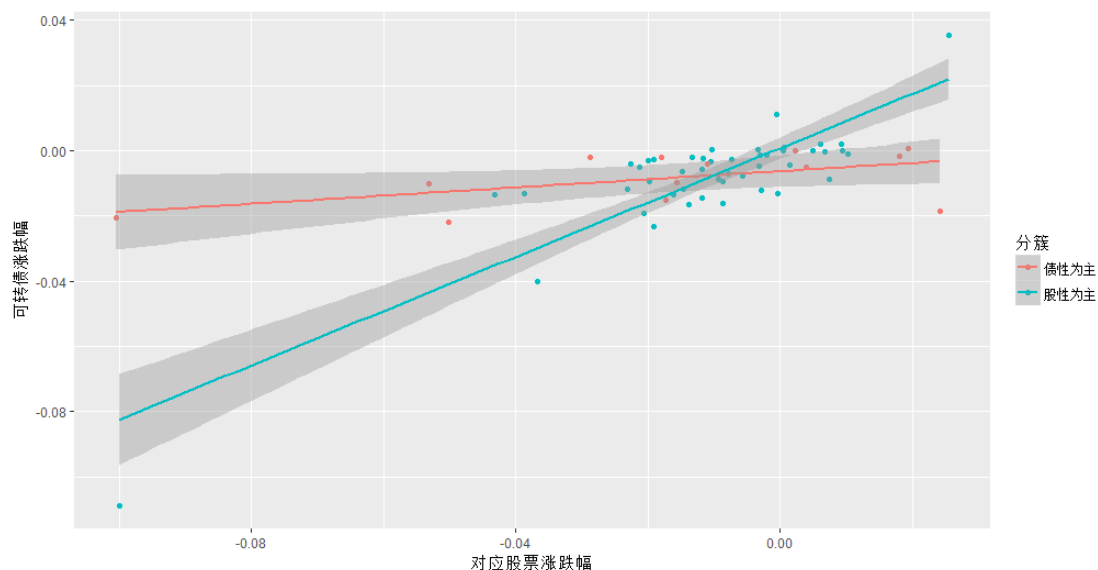


图 4 可转债与对应股票涨跌幅

可以发现对于股票的波动，股性可转债波动较大，而债性可转债波动较小，这进一步验证了我们的经验。接下来进行定量分析。

将其分别保存为 **bondlike**（债性为主）与 **stocklike**（股性为主）两个数据框，进一步从数值上分析其波动性质。通过计算可转债涨跌幅平均数（**bond_mean**），对应股票涨跌幅平均数（**stock_mean**），对应股票相对于可转债涨跌幅平均数倍数（**mean_div**），可转债涨跌幅标准差（**bond_sd**），对应股票涨跌幅标准差（**stock_sd**），对应股票相对于可转债涨跌幅标准差倍数（**sd_div**），可以获得下表：

表 1 可转债与对应股票涨跌幅统计关系

class	bond_mean	stock_mean	mean_div	bond_sd	stock_sd	sd_div
bondlike	-0.0085	-0.01745	2.053394	0.008061	0.034888	4.328001
stocklike	-0.00819	-0.01075	1.312898	0.018358	0.018896	1.02931

另外将其可视化，可得图 5：

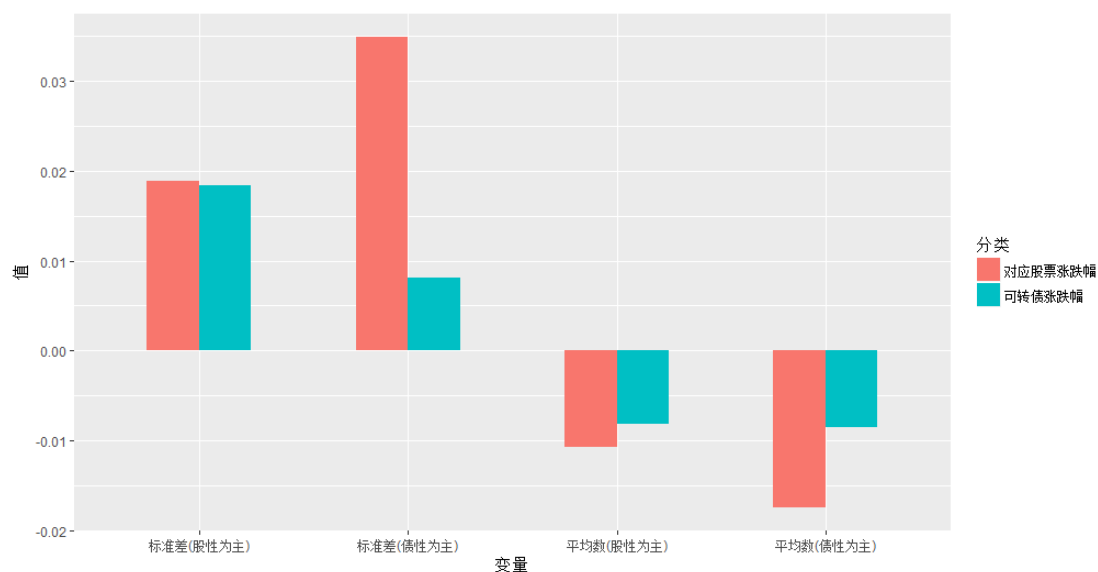


图 5 可转债与对应股票涨跌幅统计关系

根据表 1 和图 5 可以看出，在涨跌幅标准差方面，股性可转债与对应股票相当，而债性可转债远低于对应股票，约为其四分之一；在涨跌幅平均数方面，股性可转债依旧与对应股票大致相当，而债性可转债仍低于对应股票，约为其二分之一，而且都为下跌。由此我们可以发现，从整体上看，股性可转债与对应股票的波动情况大致相当，购买股性可转债，下跌情况时的波动风险与对应股票类似；而债性可转债的波动幅度和波动差异情况都明显低于对应股票，购买债性可转债很可能做到降低波动风险能够降低下跌情况时的波动风险。

另外可以观察到债性可转债对应股票涨跌幅标准差显著高于股性对应股票，且跌幅也比股性对应股票大，可能是由于这些股票的股价相对低迷，在下跌环境中可能有更差的表现，当然也可能只是因为样本不足产生的偶然现象。

c.线性回归分析

接下来，我们利用线性回归，研究两种可转债与对应股票涨跌幅的相关关系，首先是股性可转债，对其进行 pearson 相关系数分析，获得结果如下：

Pearson's product-moment correlation

```
data: stocklike$change_pct and stocklike$stock_cp
t = 10.997, df = 44, p-value = 3.29e-14
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7532152 0.9183091
sample estimates:
cor
0.8562814
```

p 值= $3.29 \times 10^{-14} < 0.05$ ，故认为可转债涨跌幅与对应股票涨跌幅是显著相关的，相关系数为 0.8562814。

之后对债性可转债进行 pearson 相关系数分析，获得结果如下：

```
Pearson's product-moment correlation

data: bondlike$change_pct and bondlike$stockcp
t = 2.1393, df = 11, p-value = 0.05567
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.01273807  0.84166359
sample estimates:
cor
0.5420516
```

p 值=0.05567>0.05，故认为可转债涨跌幅与对应股票涨跌幅并非显著相关。

由此可以发现股性可转债与对应股票涨跌相关性较强，存在较强的涨跌联动作用，而债性可转债与对应股票涨跌相关性较弱，在对应股票价格波动时，可转债的跟随波动情况不明显。

之后分别进行线性回归分析，对于股性可转债：

```
Call:
lm(formula = stocklike$change_pct ~ 1 + stocklike$stockcp)

Residuals:
    Min       1Q   Median       3Q      Max
-0.0263654 -0.0061129 -0.0009664  0.0064753  0.0217828

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0007552   0.0016312    0.463   0.646
stocklike$stockcp 0.8318984   0.0756495  10.997 3.29e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009589 on 44 degrees of freedom
Multiple R-squared:  0.7332,    Adjusted R-squared:  0.7272
F-statistic: 120.9 on 1 and 44 DF,  p-value: 3.29e-14
```

得到线性回归方程为 $y=0.8319x+0.0008$ ，y 为可转债涨跌幅，x 为对应股票涨跌幅，相关系数 R^2 为 0.7272。截距基本为 0，相关性较好。

之后进行残差分析：

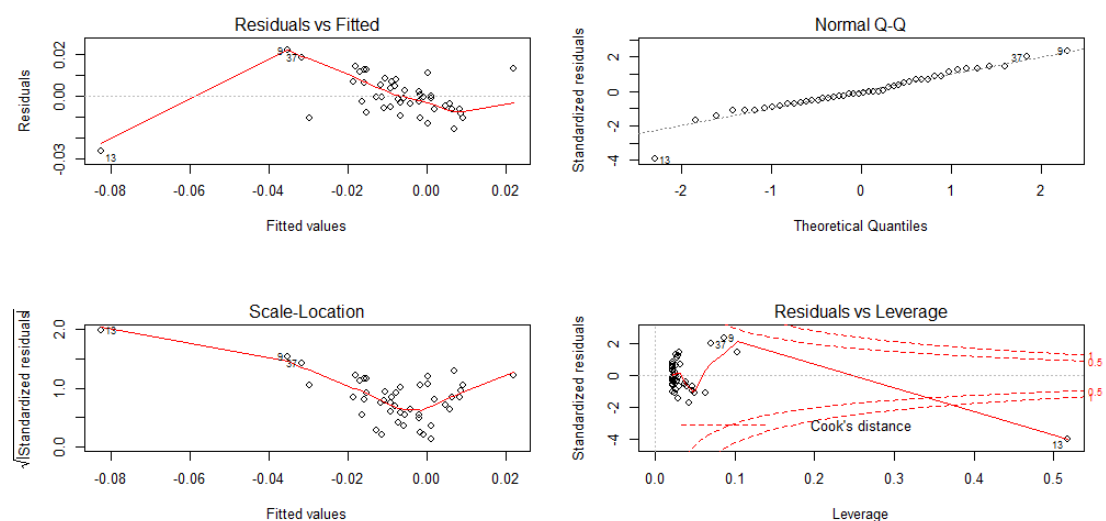


图 6

左上为拟合值 \hat{y} 对残差的图形, 可以发现数据点都基本均匀地分布在直线 $y=0$ 的两侧, 无明显趋势;

右上为 Normal QQ 图, 图中数据除了 13 号点, 分布趋于一条直线, 说明除了 13 号点, 残差是服从正态分布的;

左下图显示了标准化残差(standardized residuals)的平方根的分布情况. 最高点 13 号点为残差最大值点, 其他均较低;

右下为 Cook 距离(Cook's distance)图, 显示 13 号点以外, 点对回归的影响都不大。

由图可以认为, 在线性回归中, 第 13 号点为强影响点。

首先查看这个点对应的数据情况:

```
> stocklike[13,]
  number  code name close change pct stockcode stockname stockprice stockcp conprice convalue conprerate bondprerate
18      18 123007 道氏 115.2 -0.1088  300409  道氏技术    51.74   -0.1    45.21  114.444    0.0066    0.3282
  trigput trigred redprice strbondvaule  conddate  blistdate      move      t cluster
18    31.65   58.77    107    86.736 2018-07-04 2018-01-26 0.05671883 5.669406      2
```

发现这个点对应的股票跌停, 可转债也有 10%以上的跌幅。从数值情况分析, 这一点对应股票波动太大, 导致模型获得的残差也太大, 体现异常; 从内在原因上分析, 由于可转债无涨跌幅限制, 而对应股票有 10%的涨跌幅限制, 所以在特殊情况下, 10%不能体现股票的真实价格波动, 出现封死涨停或者封死跌停的情况, 而可转债的涨跌幅可能超出股票涨跌幅, 表现出可转债“价格发现”的特性。

综上所述, 由于-10%不足以体现股票价格波动, 考虑将这个异常点去掉, 重新进行线性回归分析, 结果如下:


```
Call:
lm(formula = del$change_pct ~ 1 + del$stockcp)

Residuals:
    Min       1Q   Median       3Q      Max
-0.0193861 -0.0036820  0.0000501  0.0034691  0.0231459

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.001321   0.001391  -0.950   0.348
del$stockcp  0.528424   0.087434   6.044 3.16e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.007784 on 43 degrees of freedom
Multiple R-squared:  0.4593,    Adjusted R-squared:  0.4467
F-statistic: 36.53 on 1 and 43 DF,  p-value: 3.158e-07
```

p 值= $3.158 \times 10^{-7} < 0.05$, 仍认为可转债涨跌幅与对应股票涨跌幅显著相关, 但得到线性回归方程为 $y = 0.5284x - 0.0013$ (y 为可转债涨跌幅, x 为对应股票涨跌幅), 截距依旧基本为 0, 但斜率有所下降。相关系数 R^2 下降到了 0.4467, 相关性明显减弱, 说明可转债涨跌幅波动呈现出一定的混沌性。

之后继续进行残差分析:

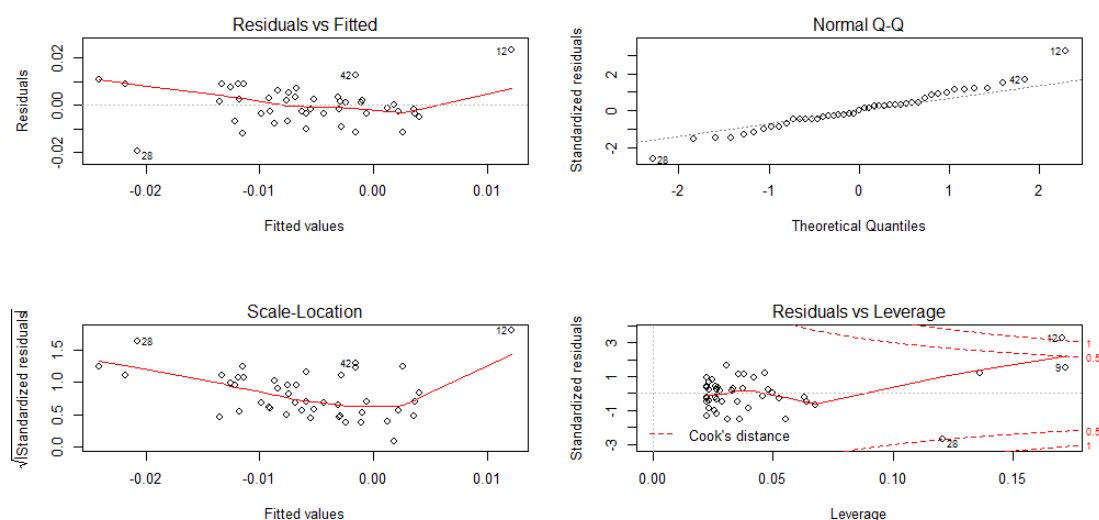


图 7

左上拟合值对残差图中,数据点都基本均匀地分布在直线 $y=0$ 的两侧,无明显趋势;

右上 Normal QQ 图,图中数据除了 12 和 28 号点,分布趋于一条直线,说明除了 12 和 28 号点有所偏离,残差是服从正态分布的;

左下标准化残差的平方根的分布情况图中,最高点 12 号点为残差最大值点;

右下 Cook 距离图显示,12 号点以外,点对回归的影响都在可控范围。

由图发现 12 号点呈现一定的异常性，仍旧查看这个点对应的数据情况：

```
> del[12,]
      number  code name close changept stockcode stockname stockprice stockcp conprice convalue conprerate bondprerate
14      14 123008 康泰 193.6   0.0353   300601 康泰生物    94.14  0.0255    46.89  200.768   -0.0357    1.2744
      trigput trigred redprice strbondvaule   condate  blistdate      move      t cluster
14      32.82   60.96      106      85.121 2018-08-07 2018-03-19 0.03697666 5.811872      2
```

发现这只可转债的涨跌幅超出对应股票涨跌幅 0.98%，原因很可能是其具有 3.57%的转股负溢价。由于 2 个多月后进入转股期即可转股套利，可能有资金选择持有可转债而非正股，降低了转股负溢价，导致可转债涨幅较大。这是可转债交易的正常情况，去除也会有过拟合嫌疑，因此将其归入分析的最终结果。

最后可以得出，股性可转债与对应股票涨跌具有相似的表现，整体上往往为对应股票涨跌幅二分之一左右，但在具体情况下可能呈现不一样的波动规律，难以用单一简单模型模拟其波动。另外，对于转股负溢价的可转债，可能获得超越正股的收益。

对于债性可转债，线性回归得：

```
Call:
lm(formula = bondlike$changept ~ 1 + bondlike$stockcp)

Residuals:
      Min       1Q   Median       3Q      Max
-0.015204 -0.001520  0.002360  0.004384  0.007821

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.006314   0.002212  -2.854   0.0157 *
bondlike$stockcp  0.125243   0.058543   2.139   0.0557 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.007075 on 11 degrees of freedom
Multiple R-squared:  0.2938,    Adjusted R-squared:  0.2296
F-statistic: 4.577 on 1 and 11 DF,  p-value: 0.05567
```

相关系数 R^2 为 0.2296，进一步验证了两者的线性相关度较低。这与债性可转债的低波动性也可能有关。

d.其他相关分析

之后考察其他联系较为复杂的变量，进行相关分析，这些变量为转股溢价率(conprerate)，纯债溢价率(bondprerate)，到期时间(t)，波动率(move)，采用绘制相关矩阵图的方式，如下：

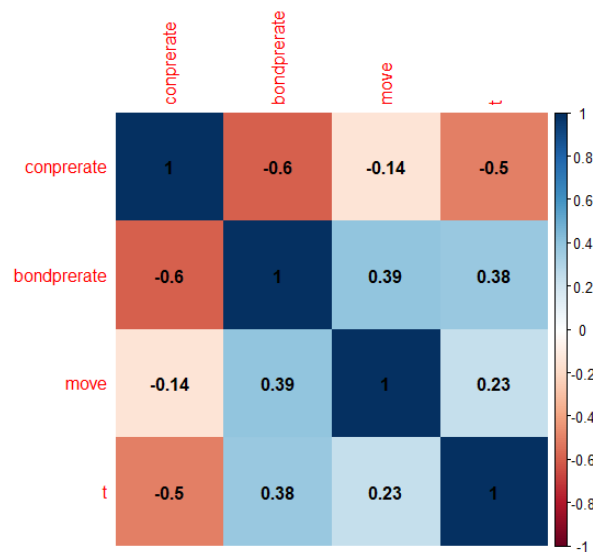


图 8 几个变量的相关矩阵图

可以发现除了转股溢价率和纯债溢价率，以及转股溢价率和到期时间存在一定的负相关关系，其他的相关关系都不明显，这主要还是这些变量之间的联系比较复杂，作用机制受多个因素影响的结果。

2) 计算可转债内在价值

这一部分我们利用 Black-Scholes 模型计算可转债的期权价值，并与其纯债价值相加由此计算可转债的内在价值，为了使用这一方法，首先做出如下假设：

1. 可转债在作为债券赎回时，认为其期权价值无限趋近于 0，而可转债具有较高纯债溢价率时，纯债价值等于行权购入股票的成本（提前支付），故可转债内在价值可以表示为纯债价值与期权价值的和
2. 可转债的期权价值分为四部分，转股期权+下调转股价期权+回售期权-强赎期权，但是下调转股价需要经股东大会审议，存在相当的外部因素和心理因素，难以计算，而回售和强赎由于条款复杂，且各个可转债的差异性大，也没有参与计算，故只计算转股期权价值
3. 可转债在上市 6 个月内不能转股，可以认为是一个欧式看涨期权，之后进入转股期，变为一个美式看涨期权，故可转债是一个奇异期权。但在实际计算时，我们认为美式看涨期权的持有者不会选择行权，而是持有或者卖出，将其看作一个欧式看涨期权处理，以运用 Black-Scholes 模型计算
4. Black-Scholes 模型的相关假设：①市场上没有无风险套利机会②市场没有交易费用③所研究股票在相应期限内不分红④市场允许卖空且资产是无限可分的⑤资产价格服从几何布朗运动模型
5. 采用历史数据法计算波动率，即对应股票前 30 个交易日涨跌幅的标准差，一年的总交易日数取 245 天，无风险利率取当日 5 年期国债收益率 3.462%。

计算公式如下：（引用自 https://en.wikipedia.org/wiki/Black–Scholes_model）

$$C(S_t, t) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)}$$
$$d_1 = \frac{1}{\sigma\sqrt{T-t}} \left[\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t) \right]$$
$$d_2 = d_1 - \sigma\sqrt{T-t}$$

其中 $C(S_t, t)$ 为期权价值， $N(d)$ 是 d 的标准正态分布的累积分布函数， S_t 为对应股票的当前价格， K 为行权价格，即转股价， σ 为波动率（每年）， $T-t$ 为到期时间（年）， r 为无风险利率。

带入计算期权价值，并计算可转债当前价与内在价值溢价率。将计算得到的期权价值绘制直方图，并添加密度曲线，得：

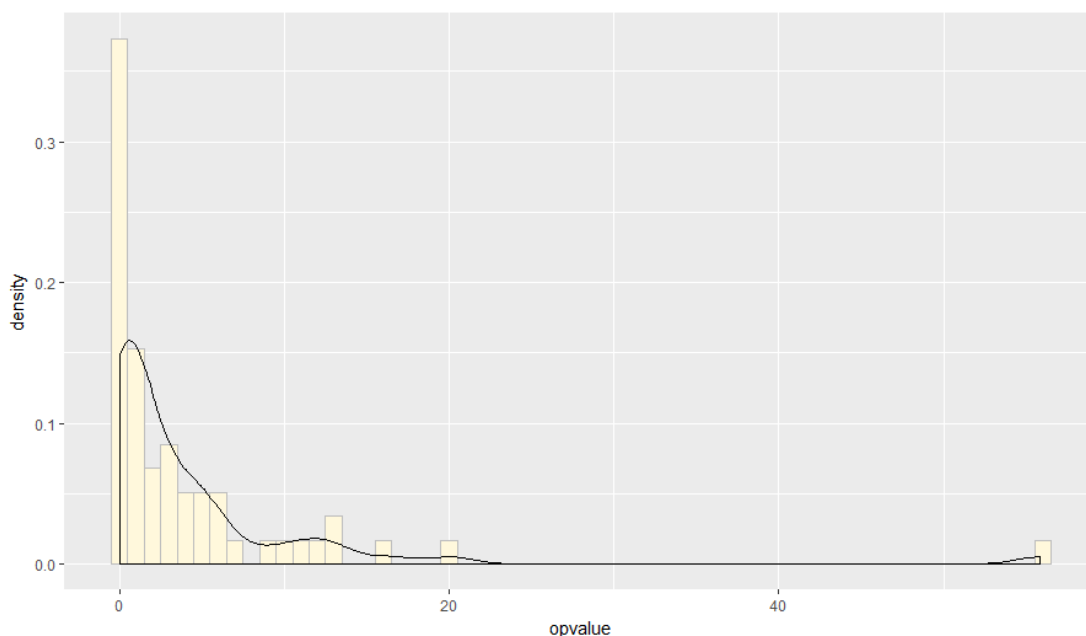


图 9 期权价值分布图

可以发现期权价值约等于 0 的可转债是最多的，可能是因为这部分可转债有大量债性可转债，期权价值较低，所以对股性可转债和债性可转债分别进行可视化：

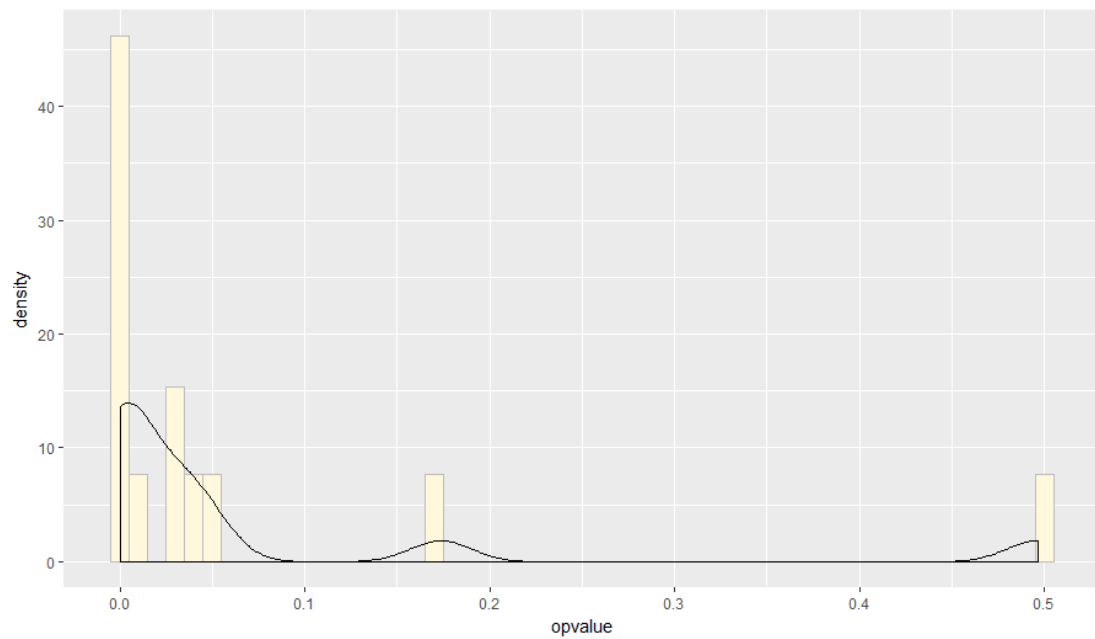


图 10 债性可转债期权价值分布图

发现债性可转债期权价值基本在 0-0.05 之间，期权价值普遍较低。

对于股性可转债：

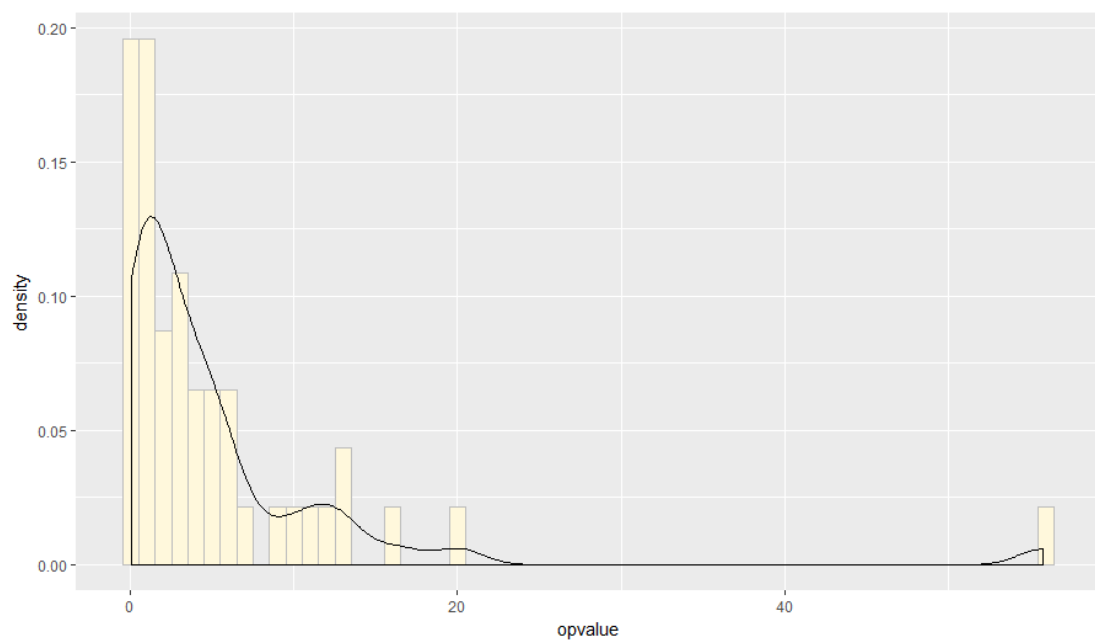


图 10 股性可转债期权价值分布图

发现可转债期权价值大部分在 0-5 之间，整体比债性可转债要高出不少。

之后绘制可转债内在价值溢价率分布图：

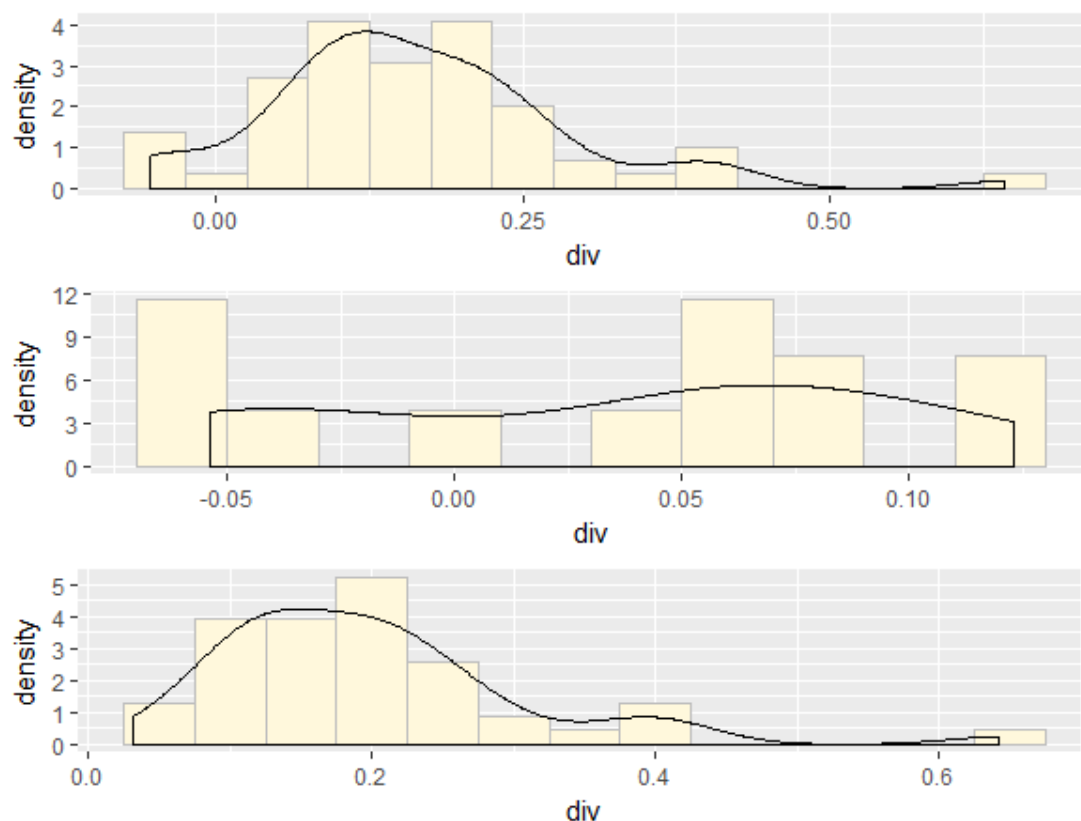


图 10 内在价值溢价率分布图
(上方为总体，中间为债性可转债，下方为股性可转债)

发现债性可转债内在价值都在-7%-13%之间，且分布较均匀，部分有负溢价，呈现普遍低估。而股性可转债都为正溢价，溢价在 15%左右的居多，整体溢价率高于债性可转债，相对高估。

这可能是因为债性可转债价格低，纯债溢价率低，在期权价值基本可忽略的情况下，内在价值溢价也较低，而股性可转债纯债溢价率高，期权价值难以补足其溢价，导致内在价值溢价也较高。

从广泛出现的较高溢价，考虑到市场的有效性，也说明这一模型存在诸多问题：

1. 计算忽略了下调转股价期权，回售期权以及强赎期权，但对于部分可转债来说，这些期权的影响较大。例如对于股价持续低迷，且接近回售期限的可转债，大股东有相当的动机下调转股价避免回售，而回售的可能性也很高，导致下调转股价期权和回售期权的价值不可忽视。
2. 可转债进入转股期为美式看涨期权，但计算中我们忽略了持有者行权的可能性，将其简化欧式看涨期权。实际上，美式看涨期权的价格高于欧式看涨期权，简化计算得到的期权价值偏小。而且从前提上，一些可转债具有转股负溢价，持有者很可能行权进行套利，证明这一简化存在较大的问题。
3. Black-Scholes 模型的假设很可能无法完全符合：

- ① 市场具有交易费用，虽然这一影响可能并不大
 - ② 虽然发行可转债的股票往往有较高的资金需求，分红可能性不高，但在相应期限内仍有分红的可能性
 - ③ A 股市场很多时候并不能借入足够的股票进行卖空，且卖空可能费率过高，这也导致了很多可转债具有转股负溢价，却无法做空正股并做多可转债进行无风险套利
 - ④ 资产并非是无限可分的，A 股只能按一手 100 股的整数倍买入
 - ⑤ 不能保证资产价格是完美的几何布朗运动模型，实际情况很可能与理论产生差异，例如 A 股有 T+1 和涨跌幅限制，导致实际资产价格变化偏离理论模型。
4. 根据一个月涨跌幅计算的波动率可能无法反映年波动率的实际情况

3) 投资策略建议

根据我们的研究结果，可以提炼出几个可能有效的投资策略：

1. 分析发现债性可转债具有低波动的性质，可以考虑买入纯债溢价率为负的可转债，并根据溢价率大小决定仓位，构造投资组合，可以在回撤相对较小的前提下，等待转债到期或开启回售，赚取利息和纯债溢价。为了进一步控制回撤，还可以采取网格交易的方法，在转债下跌时分批买入，获得更大的安全边际。如果出现下调转股价，或者一致预期下调转股价的情况，还可获得超额收益。纯债溢价率为负的可转债如下：

```
> group1 <- bond[bond$bondprerate<0, ]
> group1[c(2:9,13,24)]
```

	code	name	close	change	pct	stockcode	stockname	stockprice	stockcp	bondprerate	div
54	127004	模塑	91.280	-0.0185		700	模塑科技	4.67	0.0241	-0.0056	-0.00554648
58	128013	洪涛	90.299	-0.0039		2325	洪涛股份	4.49	-0.0110	-0.0318	-0.03183300
59	127003	海印	90.500	-0.0021		861	海印股份	2.73	-0.0180	-0.0539	-0.05386192
60	128012	辉丰	84.480	-0.0102		2496	辉丰股份	3.21	-0.0531	-0.0519	-0.05193706
67	123001	蓝标	90.360	-0.0218		300058	蓝色光标	6.25	-0.0502	-0.0536	-0.05359199

这些转债同样具有内在价值负溢价，这一策略主要风险是债券违约风险。

2. 根据内在价值溢价率，选取负溢价的低估品种，发现获得的结果与第一种策略相同，故不做赘述
3. 根据内在价值溢价率，选取股性可转债中溢价较低的低估品种，排序后观察发现有 5 只可转债的溢价率小于 10%，考虑以这五只可转债，根据溢价率大小决定仓位，构造投资组合，期待溢价中枢上移获取收益。

```
> group3 <- stocklike[stocklike$div<0.1, ]
> group3[c(2:9,12,24)]
```

	code	name	close	change	pct	stockcode	stockname	stockprice	stockcp	conprerate	div
4	128038	利欧	94.644	-0.0070		2131	利欧股份	2.50	-0.0079	0.0449	0.08791199
5	128037	岩土	94.024	-0.0043		2542	中化岩土	7.03	0.0014	0.0767	0.08434698
44	113502	嘉澳	94.840	-0.0011		603822	嘉澳环保	36.80	0.0102	0.1721	0.03164424
61	113010	江南	100.710	-0.0013		601199	江南水务	4.93	-0.0020	0.2461	0.06676807
70	110030	格力	107.000	-0.0046		600185	格力地产	6.15	-0.0032	0.2596	0.06779097

这一策略主要风险是股票下跌导致可转债跟随下跌、溢价没有上升反而进一步降低、以及可转债违约等风险。

4. 探究中发现对于一些转股负溢价的可转债，可能出现可转债涨幅大于正

股的情况，考虑可以买入，等待转股负溢价消失时卖出。若对应股票下跌，转股负溢价可以作为“安全垫”起到一定保护作用，若对应股票上涨，可转债也会跟随上涨，因为到转股期后负溢价必然趋于 0。这一策略还可以结合其他量化策略，作为股票交易的替代获得更高的 **sharp**，还可借助可转债 T+0 交易的特性反复操作多次获取利润。这里列出转股负溢价的可转债：

```
> group4 <- bond[bond$conprerate<0, ]
> group4[c(2:9,12,18,24)]
```

	code	name	close	change	pct	stockcode	stockname	stockprice	stockcp	conprerate	condat	div
3	113506	鼎信	109.04	-0.0078		603421	鼎信通讯	24.44	-0.0057	-0.0247	2018-10-22	0.2037495
9	123009	星源	125.79	-0.0087		300568	星源材质	35.86	-0.0094	-0.0182	2018-09-13	0.2925843
14	123008	康泰	193.60	0.0353		300601	康泰生物	94.14	0.0255	-0.0357	2018-08-07	0.3738503
28	123005	万信	162.80	-0.0130		300168	万达信息	23.87	-0.0004	-0.1059	2018-06-25	0.6417004
30	113503	泰晶	112.88	-0.0089		603738	泰晶科技	20.32	0.0074	-0.0006	2018-06-21	0.2125516
42	110039	宝信	136.33	-0.0401		600845	宝信软件	24.93	-0.0367	-0.0053	2018-05-23	0.4247555
43	110038	济川	130.21	0.0011		600566	济川药业	54.03	0.0006	-0.0110	2018-05-17	0.1974540

对于临近转股日或以及开始转股的可转债，这一策略更加有效，可以将转股日期与转股负溢价一同作为决定仓位的依据。这一策略主要风险是股票下跌导致可转债跟随下跌的风险。

五、 总结与展望

本文通过对可转债的统计分析，发现了一些可转债的统计规律，例如债性可转债的低波动性特征，股性可转债与对应股票涨跌幅联动的直观性和复杂性；通过其内在价值的计算，构造了一个可转债定价模型，并以此出发提出了一些投资建议。但由于数据有限，能进行的分析有限，之后希望获取更多可转债的相关数据，包括 tick 数据或其他 K 线数据，构造更加复杂也更加精确的模型来描述变量间相关关系。

另外利用 Black-Scholes 模型计算可转债的期权价值时做出了诸多假设与简化，之后希望加入其他几个忽略掉的期权价值，使用最小二乘蒙特卡洛法 (LSMC)，二叉树等计算美式期权定价更加泛用的方法，获得更加精确的结果，并研究可转债定价的优化方案，尝试用于实时计算中。

之后也会考虑对提出的策略建议进行回测，尝试挖掘出高 **sharp** 有实盘交易价值的策略。

可转债由于其复杂性和较低流动性常常被许多投资人忽视，但被忽视也意味着更多未发掘的价值。可转债的灵活性也为其带来了相当的优势，其中蕴含的交易机会值得我们重视。

附录：全部代码（含注释）

```
library(xml2)
library(rvest)
library(RSelenium)
library(RJSONIO)
library(jsonlite)
library(mongolite)
library(VIM)
library(plyr)
library(gridExtra)
library(ggplot2)
library(graphics)
library(cluster)
library(corrplot)

#数据获取(东方财富)
url0 <- 'http://quote.eastmoney.com/center/fullscreenlist.html#bp_8'
number=code=name=close=change_pct=stockcode=stockname=stockprice=stockcp=con
price=convalue=conprerate=bondprerate=trigput=trigred=redprice=stbondvaule=condat
e=blisdate=move=stchpct=sd=NULL

remDr = remoteDriver('localhost',4444L,browserName='firefox')
remDr$open()
remDr$navigate(url0)

for(i in 1:2)
{
web <- remDr$getPageSource()[[1]] %>% read_html()
number <- c(number,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
number') %>% html_text())
code <- c(code,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-Code') %>%
html_text())
name <- c(name,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-Name') %>%
html_text())
close <- c(close,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-Close') %>%
html_text())
change_pct <- c(change_pct,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
ChangePercent') %>% html_text())
stockcode <- c(stockcode,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
UnderlyingStockCode') %>% html_text())
stockname <- c(stockname,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
UnderlyingStockName') %>% html_text())
```

```

stockprice <- c(stockprice,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
UnderlyingStockPrice') %>% html_text())
stockcp <- c(stockcp,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
UnderlyingStockCP') %>% html_text())
conprice <- c(conprice,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
ConversionPrice') %>% html_text())
convalue <- c(convalue,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
ConversionValue') %>% html_text())
conprerate <- c(conprerate,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
ConvertiblePremiumRate') %>% html_text())
bondprerate <- c(bondprerate,web %>% html_nodes('tbody') %>% html_nodes('.listview-
col-BondPremiumRate') %>% html_text())
trigput <- c(trigput,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
TriggerPriceOfSpecialPut') %>% html_text())
trigred <- c(trigred,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
TriggerPriceOfSpecialRedemption') %>% html_text())
redprice <- c(redprice,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
Redemptionprice') %>% html_text())
strbondvaule <- c(strbondvaule,web %>% html_nodes('tbody') %>% html_nodes('.listview-
col-StraightBondValue') %>% html_text())
condate <- c(condate,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
ConversionDate') %>% html_text())
blistdate <- c(blistdate,web %>% html_nodes('tbody') %>% html_nodes('.listview-col-
BAdinLISTDATE') %>% html_text())

```

```

pg <- remDr$findElement(using = "css", selector =
'span#listview_next.next.paginate_button')
remDr$executeScript("arguments[0].click();", list(pg))
Sys.sleep(1)
}

```

#数据获取(网易财经)

```
urlst <- 'http://quotes.money.163.com/trade/ljsysj_'
```

```
for(i in stockcode)
```

```
{
```

```
url <- paste(urlst,i,".html#01b07",sep = "")
```

```
webst <- read_html(url)
```

```
stchpct <- c(stchpct,webst %>% html_nodes('table') %>% html_nodes('tr') %>%
html_nodes('td:nth-child(7)') %>% html_text())
```

```
stchpct=as.numeric(stchpct)
```

```
stchpct <- na.omit(stchpct)
```

```
sd <- c(sd(stchpct))
```

```

move <- c(move,sd)
stchpct = sd = NULL
}

```

#构造数据框，并储存于各种媒介

```

bond=data.frame(number,code,name,close,change_pct,stockcode,stockname,stockprice,stockcp,conprice,convalue,conprerate,bondprerate,trigput,trigred,redprice,strbondvaule,condate,blistdate,move)
print(bond)
write.table(bond, file = "c:/bond.txt", row.names = F, quote = F)
write.csv(bond,"c:/bond2.csv")
jsdata <- toJSON(bond)
write_json(jsdata,"c:/bond3.json")
con <- mongo(collection="bond",db="test",url="mongodb://localhost")
con$insert(bond)

```

#数据预处理

```

bond <- read.table("C:/Users/14113/Desktop/learn/data science/bond/bond.txt",
head=TRUE)
bond$change_pct <- as.numeric(sub("%", "", bond$change_pct))/100
bond$stockcp <- as.numeric(sub("%", "", bond$stockcp))/100
bond$conprerate <- as.numeric(sub("%", "", bond$conprerate))/100
bond$bondprerate <- as.numeric(sub("%", "", bond$bondprerate))/100
bond$close=as.numeric(as.character(bond$close))
bond$stockprice=as.numeric(as.character(bond$stockprice))
bond$conprice=as.numeric(as.character(bond$conprice))
bond$convalue=as.numeric(as.character(bond$convalue))
bond$trigput=as.numeric(as.character(bond$trigput))
bond$trigred=as.numeric(as.character(bond$trigred))
bond$redprice=as.numeric(as.character(bond$redprice))
bond$strbondvaule=as.numeric(as.character(bond$strbondvaule))
bond$condate=as.Date(bond$condate)
bond$blistdate=as.Date(bond$blistdate)
bond$move=bond$move/100
is.na(bond)
bond <- na.omit(bond)

```

#去除转债字样与添加到期时间

```

bond[,3] <- gsub("转债","",bond[,3])
t=expiredate=NULL
for(i in 1:length(bond$number))
{
years=6+as.numeric(difftime(bond[i,19], "2018-05-27", units="days"))/365
expiredate <- c(years)
}

```

```
t <- c(t,expiredate)
}
bond <- cbind(bond,t)
```

#图 1

```
h1<-ggplot(bond,aes(x=conprerate)) + geom_histogram(binwidth=0.1, fill="white",
colour="black")
h2<-ggplot(bond,aes(x=bondprerate)) + geom_histogram(binwidth=0.1, fill="white",
colour="black")
h3<-ggplot(bond,aes(x=move)) + geom_histogram(binwidth=0.005, fill="white",
colour="black")
h4<-ggplot(bond,aes(x=t)) + geom_histogram(binwidth=0.1, fill="white", colour="black")
grid.arrange(h1,h2,h3,h4)
```

#聚类分析

```
prerate = as.data.frame(bond[,12:13])
km = kmeans(prerate,2, nstart=25)
prerate$cluster = factor(km$cluster)
centers=as.data.frame(km$centers)
prerate$name = bond$name
```

#图 2

```
g1 <- ggplot(data=prerate, aes(x=conprerate, y=bondprerate, color=cluster )) +
  geom_point() + theme(legend.position="right") +
  geom_point(data=centers, aes(x=conprerate,y=bondprerate, color=as.factor(c(1,2))),
  size=10, alpha=.3, show.legend=FALSE) +xlab("转股溢价率")+ylab("纯债溢价率")
```

#图 3

```
g2 <- g1 + geom_text(aes(label=name), size=3) + scale_color_discrete(name="分簇",breaks=c("1", "2"),labels=c("债性为主", "股性为主"))
```

#图 4

```
bond$cluster = factor(km$cluster)
g3 <- ggplot(bond,aes(x=stockcp, y=change_pct, color=cluster)) +geom_point()
+stat_smooth(method=lm) +xlab("对应股票涨跌幅")+ylab("可转债涨跌幅")+
scale_color_discrete(name="分簇",breaks=c("1", "2"),labels=c("债性为主", "股性为主"))
```

#分类及分析统计性质

```
bondlike <- bond[bond$cluster==1, ]
stocklike <- bond[bond$cluster==2, ]
class <- c("bondlike","stocklike")
bond_mean <- c(mean(bondlike$change_pct),mean(stocklike$change_pct))
stock_mean <- c(mean(bondlike$stockcp),mean(stocklike$stockcp))
mean_div
```

<-

```

c(mean(bondlike$stockcp)/mean(bondlike$change_pct),mean(stocklike$stockcp)/mean(st
ocklike$change_pct))
bond_sd <- c(sd(bondlike$change_pct),sd(stocklike$change_pct))
stock_sd <- c(sd(bondlike$stockcp),sd(stocklike$stockcp))
sd_div <-
c(sd(bondlike$stockcp)/sd(bondlike$change_pct),sd(stocklike$stockcp)/sd(stocklike$chan
ge_pct))
stat1 <- data.frame(class,bond_mean,stock_mean,mean_div,bond_sd,stock_sd,sd_div)
write.csv(stat1,"c:/change_stat.csv")

```

#图 5

```

para <- c("平均数(债性为主)","平均数(债性为主)","标准差(债性为主)","标准差(债性为主)","
平均数(股性为主)","平均数(股性为主)","标准差(股性为主)","标准差(股性为主)")
value <-
c(mean(bondlike$change_pct),mean(bondlike$stockcp),sd(bondlike$change_pct),sd(bondli
ke$stockcp),mean(stocklike$change_pct),mean(stocklike$stockcp),sd(stocklike$change_pc
t),sd(stocklike$stockcp))
class <- c("可转债涨跌幅","对应股票涨跌幅","可转债涨跌幅","对应股票涨跌幅","可转债涨
跌幅","对应股票涨跌幅","可转债涨跌幅","对应股票涨跌幅")
stat2 <- data.frame(para,value,class)
g4 <- ggplot(stat2, aes(x=para,y=value,fill=class)) + geom_bar(stat="identity",
width=0.5, position=position_dodge(0.5))+xlab(" 变 量 ") + ylab(" 值 ") +
scale_fill_discrete(name="分类")

```

#相关分析

```

cor.test(stocklike$change_pct,stocklike$stockcp,method="pearson")
cor.test(bondlike$change_pct,bondlike$stockcp,method="pearson")

```

#股性可转债线性回归分析及残差分析及图 6

```

lm.sol<-lm(stocklike$change_pct ~ 1+stocklike$stockcp)
summary(lm.sol)
op<-par(mfrow=c(2, 2))
plot(lm.sol)
par(op)

```

#去除异常值后重新分析及图 7

```

del<-stocklike[-13,]
lm.resol<-lm(del$change_pct ~ 1+del$stockcp)
summary(lm.resol)
op2<-par(mfrow=c(2, 2))
plot(lm.resol)
par(op2)

```

#债性可转债线性回归分析

```

lm.sol2<-lm(bondlike$change_pct ~ 1+bondlike$stockcp)
summary(lm.sol2)

```

#图 8

```
other <- bond[c(12,13,20,21)]
corbond <- cor(other)
corrplot(corbond,method="shade",shade.col=NA,addCoef.col="black")
```

#计算内在价值

```
opvalue=div=v=d=NULL
for(i in 1:length(bond$number))
{
  d1 <- log(bond[i,8]/bond[i,10])/bond[i,20]/sqrt(bond[i,21]*245/30)+(0.03462+bond[i,20]*bond[i,20]/30*245/2)*sqrt(bond[i,21]/245*30)/bond[i,20]
  d2 <- d1-bond[i,20]*sqrt(bond[i,21]*245/30)
  p1 <- pnorm(d1, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
  p2 <- pnorm(d2, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
  op <- bond[i,8]*p1-bond[i,10]*p2*exp(-0.03462*bond[i,21])
  v <- c(op)
  d <- c((bond[i,4]/(bond[i,17]+op)-1))
  opvalue <- c(opvalue,v)
  div <- c(div,d)
}
bond <- cbind(bond,opvalue,div)
```

#图 9

```
g5<-ggplot(bond,aes(x=opvalue,y=..density..)) + geom_histogram(binwidth=1,
fill="cornsilk", colour="grey")+geom_density()
```

#图 10 及图 11

```
bondlike <- bond[bond$cluster==1, ]
stocklike <- bond[bond$cluster==2, ]
g6<-ggplot(bondlike,aes(x=opvalue,y=..density..)) + geom_histogram(binwidth=0.01,
fill="cornsilk", colour="grey")+geom_density()
g7<-ggplot(stocklike,aes(x=opvalue,y=..density..)) + geom_histogram(binwidth=1,
fill="cornsilk", colour="grey")+geom_density()
```

#图 12

```
g8<-ggplot(bond,aes(x=div,y=..density..)) + geom_histogram(binwidth=0.05, fill="cornsilk",
colour="grey")+geom_density()
g9<-ggplot(bondlike,aes(x=div,y=..density..)) + geom_histogram(binwidth=0.02,
fill="cornsilk", colour="grey")+geom_density()
g10<-ggplot(stocklike,aes(x=div,y=..density..)) + geom_histogram(binwidth=0.05,
fill="cornsilk", colour="grey")+geom_density()
grid.arrange(g8,g9,g10)
```

```
#投资策略建议
group1 <- bond[bond$bondprerate<0, ]
group1[c(2:9,13,24)]
group2 <- bond[bond$div<0, ]
arrstocklike <- arrange(stocklike,stocklike[,24])
group3 <- stocklike[stocklike$div<0.1, ]
group3[c(2:9,12,24)]
group4 <- bond[bond$conprerate<0, ]
group4[c(2:9,12,18,24)]
```