

## 说明

## 因子部分

文章介绍了两种因子扩散度的计算方法

$$QBD_t = \bar{\beta}_t^{High} - \bar{\beta}_t^{Low}$$

$$BD_t = \sqrt{\sum_{i=1}^n (w_{i,t} \cdot \beta_{i,t} - \sum_{i=1}^n (w_{i,t} \cdot \beta_{i,t}))^2}$$

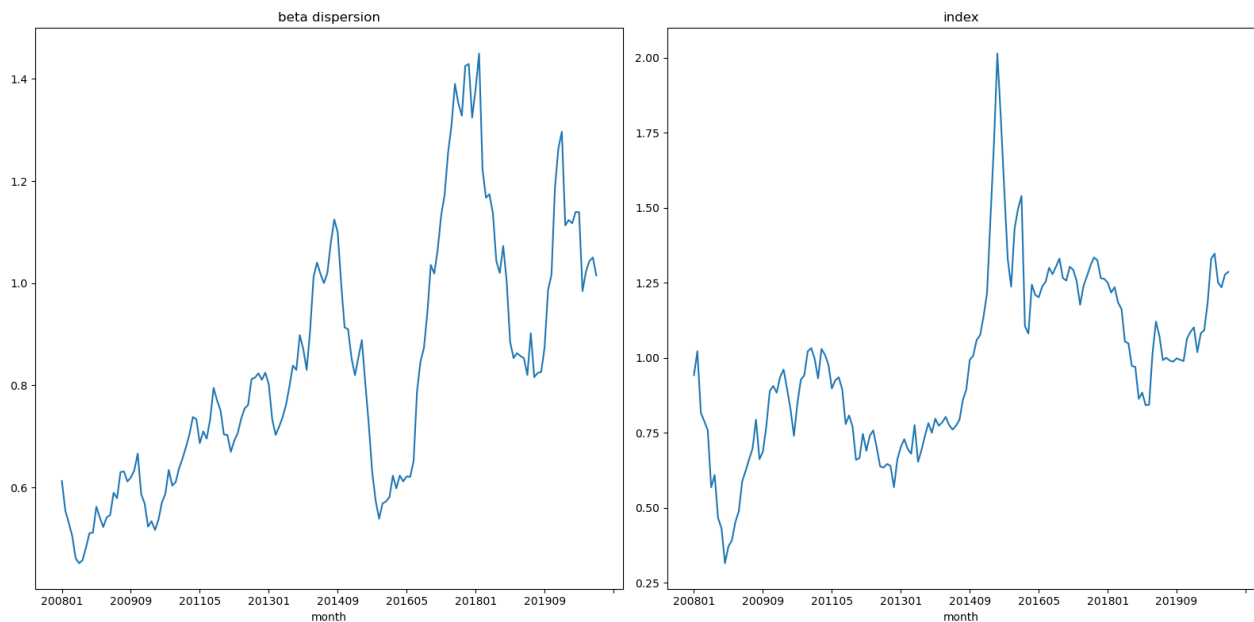
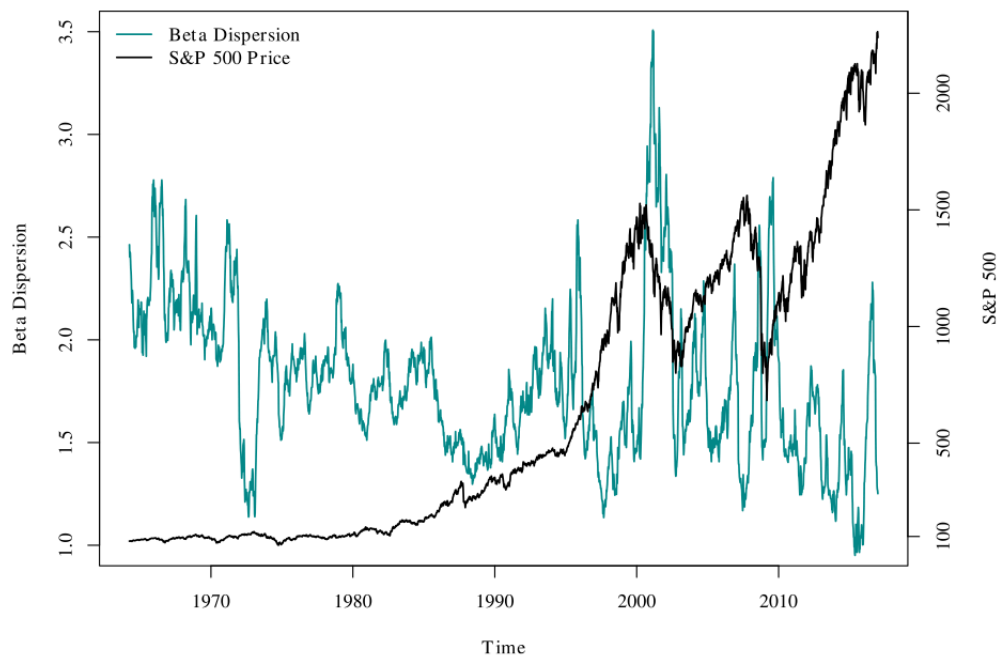
其中，做一点质疑。感觉 $BD_t$ 因子的写法应该是，可能是作者笔误。 $\sum w_{i,t} \cdot \beta_{i,t}$ 代表了整体的beta。beta减去整体beta才能表示其内部扩散情况。

$$BD_t = \sqrt{\sum_{i=1}^n w_{i,t} \cdot (\beta_{i,t} - \sum_{i=1}^n (w_{i,t} \cdot \beta_{i,t}))^2}$$

## 回归部分

原文的两个基本逻辑

成分指数中个股beta之差beta扩散度，beta扩散度可以表示了当前指数的一些性质。是一个很好的指标。



上图对应原文中的图。可以看出来，中国股市相对于美国股市beta扩散值更小。

#### OLS Regression Results

```

=====
Dep. Variable:          CHGPct      R-squared:                0.004
Model:                  OLS          Adj. R-squared:           -0.003
Method:                 Least Squares  F-statistic:              0.5472
Date:                  Sun, 04 Apr 2021  Prob (F-statistic):       0.461
Time:                  22:58:13       Log-Likelihood:          157.63
No. Observations:      156           AIC:                    -311.3
Df Residuals:          154           BIC:                    -305.2
Df Model:               1
Covariance Type:       nonrobust
=====

```

```

=====
coef      std err          t      P>|t|      [0.025      0.975]
-----

```

```

const      -0.0123      0.025      -0.487      0.627      -0.062      0.038
beta       0.0216      0.029      0.740      0.461      -0.036      0.079
=====
Omnibus:                7.117      Durbin-Watson:                1.806
Prob(Omnibus):          0.028      Jarque-Bera (JB):            8.483
Skew:                  -0.313      Prob(JB):                    0.0144
Kurtosis:              3.956      Cond. No.                    7.05
=====
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is
correctly specified.

```

当我们仅对beta分散度做预测性的检验，得到的结果并不理想。其p值的检验未通过。文中引入了一个虚拟变量 $D_{t-2}$ 表示提前两个时间点，市场的涨跌情况。我们也对该变量进行引入，得到如下的回归结果。其检验方程为：

```

                                OLS Regression Results
=====
Dep. Variable:                  CHGPct      R-squared (uncentered):
                                0.024
Model:                          OLS      Adj. R-squared (uncentered):
                                0.011
Method:                        Least Squares      F-statistic:
                                1.861
Date:                          Sun, 04 Apr 2021      Prob (F-statistic):
                                0.159
Time:                          23:04:53      Log-Likelihood:
                                158.90
No. Observations:              156      AIC:
                                -313.8
Df Residuals:                  154      BIC:
                                -307.7
Df Model:                      2
Covariance Type:               nonrobust

=====
                                coef      std err      t      P>|t|      [0.025      0.975]
-----
beta                          0.0205      0.011      1.850      0.066      -0.001      0.042
dummy                         -0.0271      0.016      -1.663      0.098      -0.059      0.005
=====
Omnibus:                      9.949      Durbin-Watson:                1.968
Prob(Omnibus):                0.007      Jarque-Bera (JB):            12.653
Skew:                        -0.411      Prob(JB):                    0.00179

```

在引入新的虚拟变量后，两个系数p值都变得显著。与原文不同的是，此处引入的虚拟变量为 $D_{t-1}$ 。虚拟变量表示了对不同的市场情况。原文认为美国市场的涨跌具有一定的联机效应（240页倒数第二段），即，在beta扩散度变大之前，市场会先出现征兆，此时的虚拟变量 $D_{t-2}$ 就是征兆的代理变量。我通过回归发现不然，国内的市场并没有这种所谓联级效应，反而，使用 $D_{t-1}$ 作为代理变量回归之后会让beta扩散度与虚拟变量都变得显著。

在显著的条件下，可以做一定的解释。虚拟变量为负，与原文一致，说明当出现了市场的崩溃之后，其后面延续的行情依旧可能是崩溃的状态。

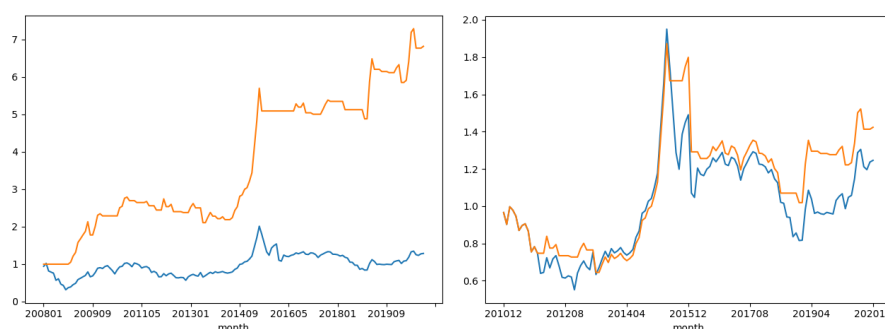
beta扩散度为正，与美股市场相反。说明，美股在崩溃之后beta越大的点回继续出现强烈的奔溃，从而反应在市场上。A股为正，说明出现beta越大的点回出现一定的调整。也正是这种调整，让A股市场的Beta扩散度是明显小于美国市场的。可以理解为，美国走28行情（即指数的行情反映在极少数的股票上），而中国普遍走一种行情（指数成分股走出的行情与成分股比较类似）。

接下来，文章使用了很多方法来说明beta分散度因子的有效性。包括，行业的分析，情绪的分析，技术指标因子分析，时间序列分区的分析。在排除了这几种情况下，beta扩散度依然显著，说明了beta扩散度的有效性。

## 策略部分

由于文中数据充足，可以做长的跳步进行策略的实现。优矿只提供了2008年之后的成分股数据。有效数据只有12年。文中基于预测的数据都是先进行20年的回归做初始化，然后，每次预测都增加时间点，预测的时长一只保证了20年以上。

因此，我们出了提供了基于预测的策略还有基于逻辑（回归结果）的策略。



在六个月的rolling下，计算第一种beta扩散度。

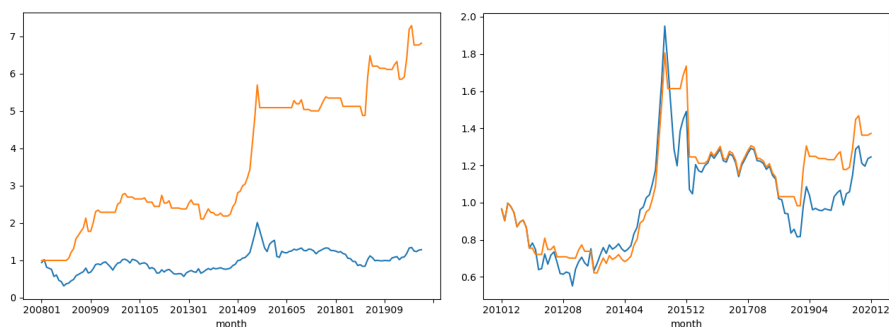
左侧是基于逻辑的策略，即上一个时间段的市场是在涨的，且当前beta值大于历史beta值的中位数。右侧是基于35天的预测，若预测大于0，则买入。

由于数据量太少会导致预测不够准确。不过与原文fig.5对应，也可以看出。基于预测的效果其实不会跑过基准多少。

基于逻辑的策略是一种选择策略，只对有把握的点买入，就会出现较好的效果。

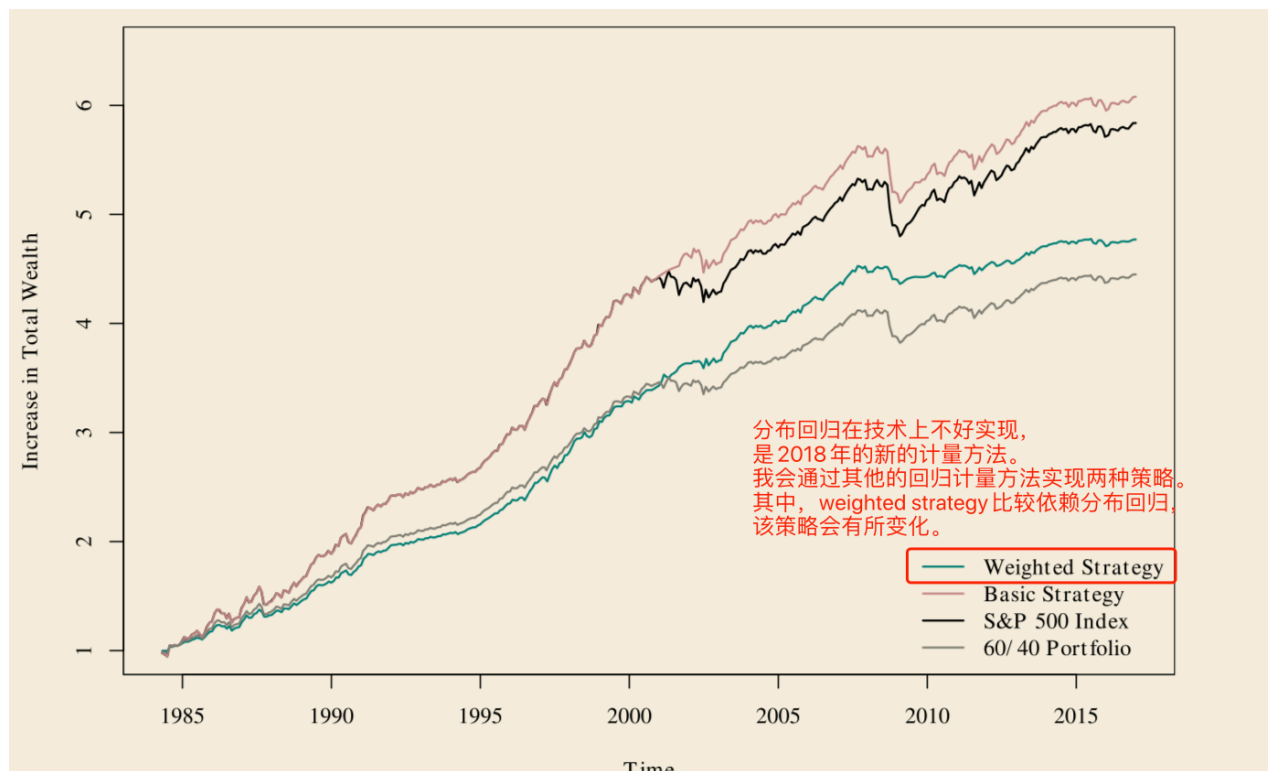
此处正对signal说明，dummy是进行过滞后处理的变量，beta在预处理过程中就进行了处理，因此，此时的signal就不需要进行滞后处理。

另一种beta 扩散度的计算方法基本走出了一样的结果。



说明两种beta扩散度的计算方法都是有效的。

## 关于文中策略部分使用分布回归的说明及其他说明

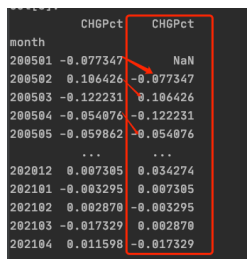


这篇文章给出来的策略效果其实不是很好（紫线对比黑线）。另，这篇文章是一篇的重点是计量的部分，不是策略驱动的。

文中股票数据来自wind，指数数据来自优矿。

可以通过修改Beta\_1.py中，106行的month\_step参数来进行月份的调整。实验展示部分使用的是6个月数据。

## 补充说明：



| month  | CHGPct    | CHGPct    |
|--------|-----------|-----------|
| 200501 | -0.077347 | NaN       |
| 200502 | 0.106426  | -0.077347 |
| 200503 | -0.122231 | 0.106426  |
| 200504 | -0.054076 | -0.122231 |
| 200505 | -0.059862 | -0.054076 |
| ...    | ...       | ...       |
| 202012 | 0.007305  | 0.034274  |
| 202101 | -0.003295 | 0.007305  |
| 202102 | 0.002870  | -0.003295 |
| 202103 | -0.017329 | 0.002870  |
| 202104 | 0.011598  | -0.017329 |

红框部分是shift(1)是数据下移，以index为200502为例，下移后数据改点对应200501时间点的数据。同理，index为200503数据下移后是200502的数据。所以，没有用到未来数据。只用到了过去的的数据。

where all variables correspond to Eq. (4) extended by the dummy  $D_{t-2}$  times the beta dispersion. The dummy becomes one if the market return in  $t-2$  is negative and becomes zero if it is positive. (原文240页倒数第三段第一句)

dummy变量的使用方法是与原文保持一致。上一个时段在涨，即dummy == 0。

使用dummy=0的逻辑是通过回归dummy变量的系数符号与显著性得到的。可以看到dummy系数为负且显著，说明dummy=1时，对收益的影响时负的。就是前一期处于下跌状态对下一期收益时负的。

所以，就很自然的得到了dummy==0的策略。

beta的系数大于0，说的是beta越大，下一期收益越大。因此，很自然的用beta序列的中位数来做阈值。

两个条件叠加，自然的生成了策略。

## 周数据说明

方法与上述保持一致，Beta1\_week.py在65行可以通过修改week的长度。我们选择大致6个月时间，即大概24周的时间。由于运行时间过长，加入了多进程来加快对数据的处理。

在对周数据做处理的过程中，发现仅对周数据滞后一期，其回归结果并不显著。由于有滞后一月的结果作为目前的经验。其滞后的变量为前四周市场的总收益。如果前四周的收益相加小于0，则标示为1，其余情况标示为0。

### OLS Regression Results

```
=====
=====
Dep. Variable:          CHGPct      R-squared (uncentered):
      0.005
Model:                OLS          Adj. R-squared (uncentered):
      0.002
Method:               Least Squares    F-statistic:
      1.719
Date:                 Tue, 06 Apr 2021    Prob (F-statistic):
      0.180
```

```

Time:                22:30:43    Log-Likelihood:
1164.3
No. Observations:    664    AIC:
-2325.
Df Residuals:        662    BIC:
-2316.
Df Model:             2

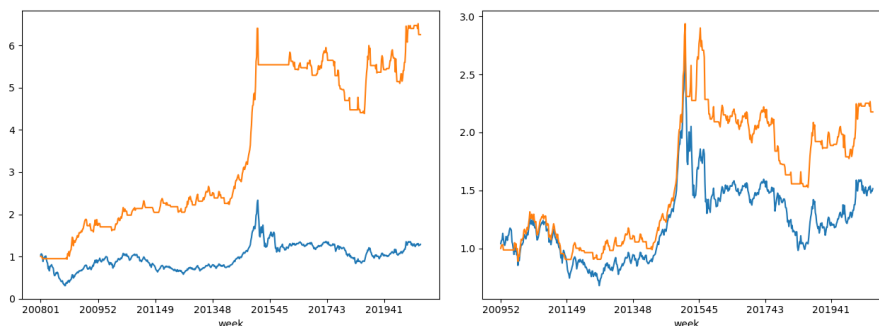
Covariance Type:      nonrobust

=====
                coef      std err          t      P>|t|      [0.025      0.975]
-----+-----
beta            0.0042      0.002       1.742     0.082     -0.001      0.009
dummy          -0.0061      0.004      -1.613     0.107     -0.013      0.001
=====

Omnibus:                59.647    Durbin-Watson:                1.935
Prob(Omnibus):           0.000    Jarque-Bera (JB):            186.607
Skew:                    -0.399    Prob(JB):                     3.01e-41
Kurtosis:                 5.471    Cond. No.                      2.47
=====

```

从回归结果可以看出来，dummy非常接近显著状态了。同时，beta也显著了。在显著条件下，可以进行策略构建。其策略与月数据的策略构建一致。



注:通过切换Beta2\_week文件的82行与84行来表示不同的beta扩散度计算方法。实验下来基本一致。