

吹当时的央行政策，同时央行吹「货币宽松」政策，但央行同时保持「货币宽松」政策的流动性（想想下2013年的“钱荒”）。另外还加入了PPI（工业品价格指数），它可以看作是CPI的先导，是一个更加敏感的反应通胀的指标。

提到股市，不得不提另一个几乎全民参与建仓的市场——房地产。与股市的起起落落不同，十年来中国的房价一路涨涨涨，诺奖获得者Robert Shiller 早就提出中国的房价远远被高估，但他万万没想到房价这个泡沫始终没有破裂。Shiller写了一本畅销书《非理性繁荣》，提到了投资者的种种非理性行为，Shiller由于在行为金融学方面的贡献获得2013年诺贝尔经济学奖。房地产和股市都是投资（机）者聚集的地方，两个市场相互影响，房价的变化能反映投资者的情绪以及风险偏好。（下图为同获2013年诺贝尔经济学奖的Fama和Shiller）



股市中大部分投资者都相信可以通过研究历史而预测未来，因此他们利用技术分析或者经验判断等方法进行选股和择时。在这一部分，选取了动量、波动率、以及各组股票上月收益这些常用变量，预测各组下一个月的收益情况。

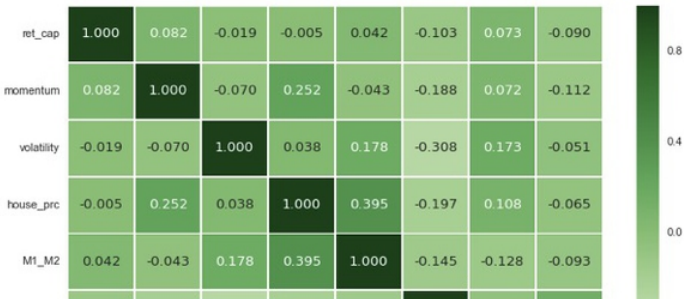
做个小结就是我会用以下变量（帮助我算命）：

- 1). 短期利率（Interest）：隔夜拆借利率的均值
- 2). CPI（月度环比）
- 3). 货币供应量：M1/M2
- 4). 房价（house_prc）：70个大中城市新建住宅价格指数（月度环比）
- 5). PPI（月度环比）
- 6). 动量（momentum）：上证指数近六个月平均收益率
- 7). 波动率（volatility）：上证指数上月历史波动率
- 8). 上月PB组合收益率差（ret_style）：低PB组合收益率-高PB组合收益率
- 9). 上月大盘股和小盘股收益率差（ret_cap）：上月中证500收益率-上证50收益率

由于宏观变量的发布滞后，这里取宏观变量滞后两期的值，即预测2016年6月的收益时，用的是2016年4月的宏观经济数据。

3. 变量检验

选取变量后，我们来check一下自变量的相关性（我真的不是故意用绿色的）。相关性均在[-0.5,0.5]范围内。



谢亚轩 2017-05-17 19:32

汇市2015年来最“平静”只是表象！
波动隐患暗流涌动

张一苇 2017-05-16 17:29



接着检验这些自变量（x）和我们的目标——下个月收益最大的组（y）是否有联系，如果某一变量在不同分组下分布相同，则这个变量一定是无效的。首先根据y的值将所有观测分成四组，对于每一个自变量x，分别检验这四组数据两两之间分布是否有显著差异。这里用了一个非参数检验方法：双样本Kolmogorov–Smirnov检验。为什么放弃常用的t检验呢？因为数据太少我真的无法保证它们正态分布。

原假设H0：两组数据服从同一分布

KS值为：

$$D_{n_1,n_2} = \sup |F_{1,n_1}(x) - F_{2,n_2}(x)|$$

其中 F_{1,n_1} 和 F_{2,n_2} 分别为两组数的经验分布， $\sup(f(x))$ 表示f(x)取最大值时x的取值。

$$D_\alpha = c(\alpha) \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

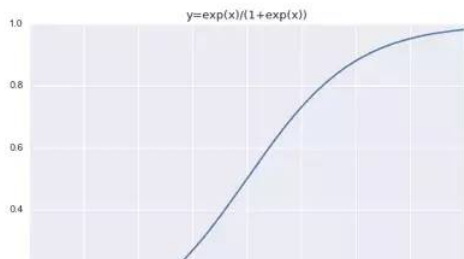
α	0.1	0.05	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.63	1.73	1.95

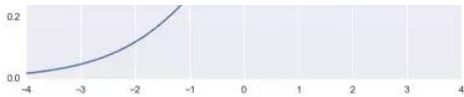
当KS值>临界值时，拒绝原假设，否则不能拒绝原假设。下图为KS检验各组的P值（这次就用红色吧）。所有变量都在某一或者某几组对比中有效（此处取10%）。



4. logistic 回归模型

普通的线性回归模型并不擅长分类问题，预测结果可能在[0,1]之外，并且对于离群值(Outlier)十分敏感。而logistic 模型在线性模型的基础上加入sigmoid函数（如下图），确保结果在[0,1]之间，并且函数在0.5左右的斜率陡然升高，结果落在0.5左右的概率很小。





假设 y 表示分类结果， $y=\{0, 1\}$ ， x 为解释变量， w 是待估参数，那么给定一个观测，分类结果为1的概率为：

$$P(y = 1|x, w) = \frac{\exp(w^T x)}{1 + \exp(w^T x)}$$

显然，这个观测分类结果为0的概率为：

$$P(y = 0|x, w) = \frac{1}{1 + \exp(w^T x)}$$

假设有 N 个观测，用极大似然估计（MLE）估计参数，似然函数如下，可以用梯度下降法(Gradient Descent)，牛顿法(Newtons Method)等数值方法求解，此处就不赘述了。

$$L(w) = \prod_{i=1}^N p_i(w)^{y_i} (1 - p_i(w))^{1-y_i}$$

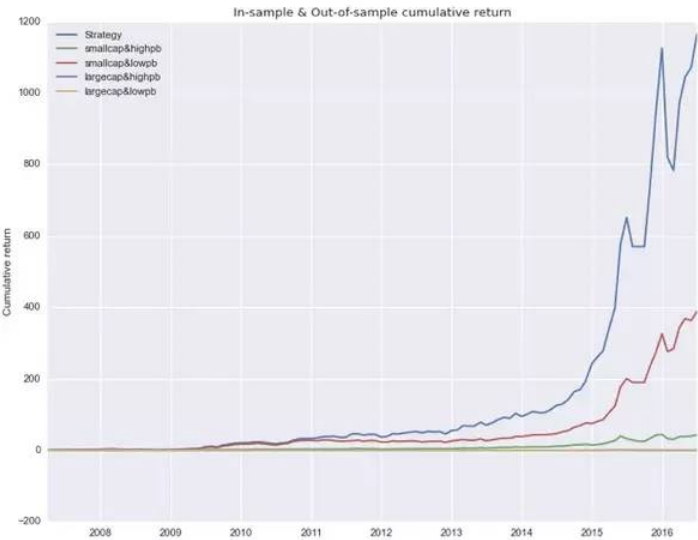
由于一共有四组股票，这里用到的是多项logistic回归(Multinomial Logistic Regression)，模型与上面的类似，假设 y 分为 k 组{ G_1, G_2, \dots, G_k }，给定某一个观测，属于第 K 组的概率如下，同样运用极大似然估计求解。

$$P(G_k|x, w) = \frac{\exp(w_k^T x)}{\sum_j \exp(w_j^T x)}$$

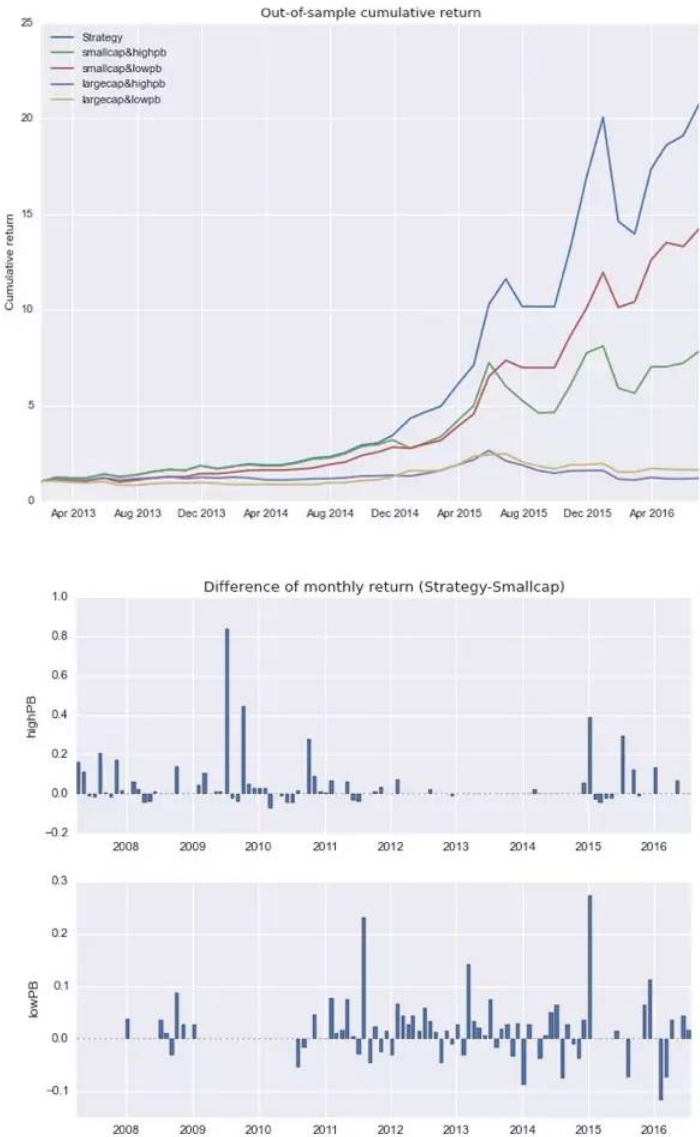
5. 结果分析

时间为2007年2月-2016年5月，共112个月，采用滚动窗口预测，估计期为70个月，样本外时间为42个月。

说了这么多废话（我并不承认），上累积收益率图（看看算命算准没）。敲黑板敲黑板，注意这里的收益率为组合内股票收益率的均值，相当于等额买入组内所有股票，并且不考虑任何交易费，以及涨跌停、停牌等不能交易的情况。蓝色为按预测结果买入相应组合的累积收益率。（由于大盘股的表现并不尽如人意，两条表示大盘股收益的曲线难以用肉眼区分，图中只有四条曲线）



上面是全样本的结果，样本可能内会过度拟合（overfitting），那么上样本外累积收益率图。可以发现，小盘低PB股的表现四组中最优秀，这也符合大家的一般认知。根据预测结果构造的策略无论在样本内还是样本外都跑过了所有的单一组合（撒花~）。值得注意的是，模型成功得捕捉了2014年12月的风格转换，买入了大盘股（我也不知道它为什么这么机智）。样本外预测的胜率为57%，跑过了25%这个掷骰子的概率，收益率也跑过了无脑买入小盘&低PB股的策略。既然小盘股那么受欢迎，最后也将策略每个月的收益与两组小盘股做了比较，胜率也都高于80%。



6 在过去几年中，许多投资者投资小市值股票而获得丰厚的回报（走上了人生巅峰），然而市场上唯一不变的就是变化。无脑买入小盘股的策略固然有效，但根据经济和市场的不同态势，动态调整股票和仓位，才能更好地顺应市场。一起动起来吧！

点击阅读原文可查看详细代码~



关注 [Ricequant](#) - 指纹长按识别二维码

Ricequant (米筐科技) 一个为你量身打造的量化策略平台
提供在线的策略编写、回测、模拟实盘、交易及分析功能
[www.ricequant.com](#)
微信号: [Ricequant](#)

☆ 收藏

微博

微信

相关新闻

• 对冲基金经理：收费的魔术
2017.01.25

• 如何用宏观因子量化预测外汇走势？高盛模型终于公开了
文/ WEEX 2017.02.27

• 如何用宏观因子量化预测外汇走势？高盛模型终于公开了
文/ 高盛集团 2017.02.27

• 差价合约CFD——统一便捷的交易模式，对接全球市场
文/ 花街老司机 2017.04.19

• 《王者荣耀》和《阴阳师》胜负已分，赞赏功能被取消难阻腾讯股价再创历史新高
文/ 王雅媛 2017.05.10

RiceQuant

Ricequant

总篇数 17 篇

暂无简介

+关注

请发表您的评论

相关信息

关于华尔街见闻
使用条款
版权声明

联系我们

广告投放
意见反馈
信息安全举报

关注我们

微博@华尔街见闻
邮件订阅
RSS订阅

加入我们

招聘专页

下载华尔街见闻

声明： 未经许可，任何人不得复制、转载、或以其他方式使用本网站的内容。

评论前请阅读网站“[跟帖评论自律管理承诺书](#)”

友情链接

新浪 腾讯财经 财经网 一财网 澎湃新闻 界面新闻 证券之星 全景财经 观察者网 经济网 黄金头条 猎聘网 华夏时报 陆家嘴金融网

© 2010 - 2017 华尔街见闻-上海阿牛信息科技有限公司 wallstreetcn.com 沪ICP备13019121号

<https://wallstreetcn.com/articles/255661>

6/6