

咨讯 ~ 快讯~ 行售 、 专栏 | 一起交易 √ 洗股宝 苗金头条 下载 金融招聘 Q 登录 I 注册

三 评论(0)

☆ 此藏

上一篇:

还在等贵金属"回 调"?看顶级交易吊 是如何"抢上车"的

下一篇·

【内参】银监会于学 军关于中国当前经济 金融形势问题的讲话

分享到





微信

基于logistic回归模型的风格轮动预测

(2016-07-28 22:04

□ 0

0. 序

这篇文章基于两个信仰:

a. 作为一个(其实是半个)学金融的人,相信股市反映了实体经济

b. 作为一个量化交易者(划掉最后三个字, 我并不交易), 相信可以根据历史数据 预测未来

(请不要告诉我这两个信仰很扯淡好嘛?我不听我不听~~

(既然是信仰, 有时真的毫无道理可言

这篇文章没有复杂的模型,也并没有研究什么高深的问题,大概只能靠图片和括 号中的第二人格(其实这才是本性)吸引大家了。下图源自网络,侵删。



1. 风格轮动

言归正传,就从有效市场小王子Fama的三因素模型说起。 Fama French 的三因 素模型是资产定价领域的经典,在传统的CAPM模型中加入市值因子(SMB)和账面 市值比因子(HML),用于解释股票的超额收益。市净率PB的倒数就是这里的账面市 值比,因此利用市值和市净率可以对股票的收益率进行预测。

$$R_{i,t} - R_{f,t} = \beta_i (R_{m,t} - R_{f,t}) + s_i SMB_t + \nu_i HML_t + \alpha_i$$

根据Fama French的结论,投资小市值低PB的股票可以获得显著的超额收益,但 股市不会一直这么按套路走, 时不时就来个二八转换, 大盘股就上位了, 令广大投 资者措手不及。大盘股VS小盘股、高PB股VS 低PB股、随着市场条件的改变、不 同风格股票的相对强弱也并不相同,这就是风格轮动。例如2014年12月,银行券 商等大蓝筹股异军突起,一时间众多小盘股难以望其项背。砖家们都说股市反映实 体经济, 那么问题来了, 股市这么任性, 实体经济真的知道吗? (你猜

2. 变量选取

首先对市场进行划分,流通市值前20%的股票为大盘股,后20%为小盘股,市净率 前20%为成长股,后20%为价值股,由此得到:高市值&低PB(group0),高市 值&高PB(group1),低市值&低PB(group2),低市值&高PB(group3)四组 股票。所以我们的目标是——预言下个月收益最高的组。(哦,原来是个算命的。

实体经济的指标那么多。哪些能够对股市,尤其是股市风格的转换产生影响呢? 首 先想到的就是资金的流动性。CPI 和M1/M2(货币供应量的比值) 两项指标能反 血虫前的货币政等 而到家后吨了资金的机合成本 犬甘里阿方垢供到家后吨了铂









以目的可以中以來,同时中以來了與本的的大學,心共定層校所自己中以來」及 期的流动性(遥想下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映通胀的指标。

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。

									_
ret_cap	1.000	0.082	-0.019	-0.005	0.042	-0.103	0.073	-0.090	0.8
momentum	0.082	1.000	-0.070	0.252	-0.043	-0.188	0.072	-0.112	
volatility	-0.019	-0.070	1.000	0.038	0.178	-0.308	0.173	-0.051	0.4
house_prc	-0.005	0.252	0.038	1.000	0.395	-0.197	0.108	-0.065	
M1_M2	0.042	-0.043	0.178	0.395	1.000	-0.145	-0.128	-0.093	0.0
		-		-					



Interest	-0.103	-0.188	-0.308	-0.197	-0.145	1.000	-0.055	0.151	-0.4
ret_style	0.073	0.072	0.173	0.108	-0.128	-0.055	1.000	0.200	
PPI-CPI	-0.090	-0.112	-0.051	-0.065	-0.093	0.151	0.200	1.000	-0.8
	ret_cap	momentum	volatility	house_prc	M1_M2	Interest	ret_style	PPI-CPI	

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:

$$D_{n1,n2} = \sup |F_{1,n1}(x) - F_{2,n2}(x)|$$

其中 $F_{1,n1}$ 和 $F_{2,n2}$ 分别为两组数的经验分布, $\sup(f(\mathbf{x}))$ 表示 $f(\mathbf{x})$ 取最大值时 \mathbf{x} 的取值。

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

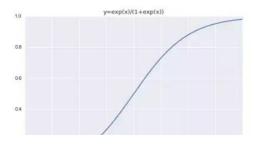
$$\frac{\alpha}{c(\alpha)} \quad \begin{array}{ccccc} 0.1 & 0.05 & 0.01 & 0.005 & 0.001 \\ c(\alpha) & 1.22 & 1.36 & 1.63 & 1.73 & 1.95 \end{array}$$

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{

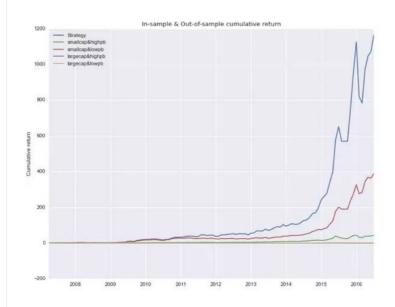
$$\{G_1, G_2, ..., G_k\}$$

,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。

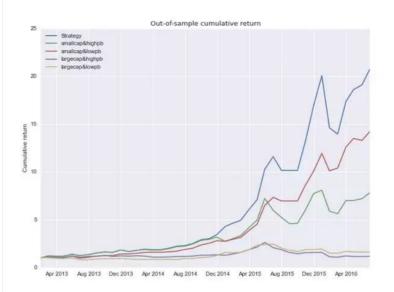
$$P(G_k|x,w) = \frac{\exp(w_k^T x)}{\sum_j \exp(w_j^T x)}$$

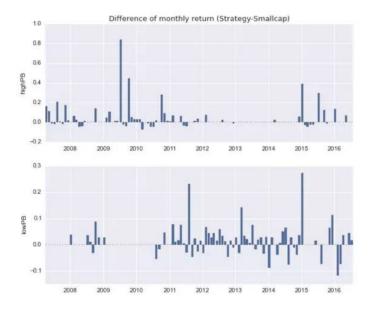
5. 结果分析

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。



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股的表现在四组中最优秀,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!

点击原文可查看详细代码~









相关信息 联系我们 关注我们 加入我们

关于华尔街见闻 广告投放 微博@华尔街见闻

使用条款 意见反馈 邮件订阅 版权声明 信息安全举报 RSS订阅

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

招聘专页