

Research on the Methods of Estimating the Position of Public Funds

Xingxing Chan*, Changjun Li

School of Mathematical Sciences, Ocean University of China, Qingdao Shandong

Email: *17860725195@163.com

Received: Jun. 12th, 2019; accepted: Jun. 26th, 2019; published: Jul. 3rd, 2019

Abstract

Fund position is the proportion of funds invested in the stock market to the assets that the fund can use. It's a reflection of market information and represents investors' expectations of the market. The position of the fund can be used as an indicator for investors to feel the future trend of the stock market. Public funds only disclose position information at the end of each quarter, which leads to an information asymmetry between investors and fund managers. In the paper, the daily return of public fund and CITIC industry index are taken as dependent and independent variables respectively. The coefficients of industry index are estimated by quadratic programming, stepwise regression and Lasso regression, and then the coefficients are summed up to obtain the position value of the fund. Errors of these three models are compared with the real fund positions published in the quarterly report. The results show that the errors of stepwise regression and Lasso regression are both smaller than that of quadratic programming. Therefore, the two methods are better than quadratic programming in solving such problems.

Keywords

Fund Position, Quadratic Programming, Stepwise Regression, Lasso Regression

公募基金仓位测算方法研究

产星星*, 李长军

中国海洋大学数学科学院, 山东 青岛

Email: *17860725195@163.com

收稿日期: 2019年6月12日; 录用日期: 2019年6月26日; 发布日期: 2019年7月3日

摘要

基金仓位是基金投入股市的资金占基金所能运用的资产的比例。它是市场信息的一个反映, 代表了

*通讯作者。

文章引用: 产星星, 李长军. 公募基金仓位测算方法研究[J]. 金融, 2019, 9(4): 307-314.

DOI: 10.12677/fin.2019.94039

投资者对市场的预期, 基金仓位的高低可以作为投资者判断股票后市走向的指标。由于公募基金只在每个季度末披露仓位和持仓信息, 这就导致了投资者与基金管理人之间的信息不对称性。本文将公募基金的日收益率与中信一级行业指数的日涨跌幅分别作为因变量和自变量, 利用二次规划、逐步回归和Lasso回归这三种方法估计中信行业指数的系数, 再对系数求和, 从而得到基金仓位值。再结合季报公布的真实基金仓位值, 比较上述三种方法得出的结果。结果表明逐步回归和Lasso回归这两种方法测算的误差比二次规划法测算的误差小, 从而得出这两种方法比二次规划法解决此类问题效果更好。

关键词

基金仓位, 二次规划, 逐步回归, Lasso回归

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

基金公司自成立以来, 作为专业的投资机构总能被市场关注。基金的持股情况更是一直广泛受到投资者的关注。由于基金仓位数据公布比较少, 一般只会在一季度结束后才会公布当季基金仓位, 所以存在一定的滞后性。对投资者来说, 仓位数据无法作为判断市场情绪的参考。因此, 如果能提前给出基金仓位的测算, 投资者就能做出更好的投资决策。

国外对于基金仓位问题的研究比较早, 比如投资组合理论[1], 资本资产定价模型[2]、Fama 五因子模型[3]。目前, 国内杨艳军[4]通过行业因子收益和规模因子收益对基金的收益进行了解释, 使用多因素优化方法建立了基金仓位配置模型。何青峰[5]通过最优化法, 选取了行业因子和规模因子对基金收益进行绩效分。马永谱[6]认为国内的基金仓位估算模型已经经历了四个阶段, 对比分析了每一代模型在算法上和误差上的不同之处, 并认为, 通过引入神经网络、因子分析、最优化等方法, 基金仓位测算的精准性会逐步提高, 模型将更加完善。海琴[7]运用 BP 神经网络构建基金仓位的测算模型, 选取单只基金“华夏成长证券投资基金”收益率时间序列作为研究数据, 证实了模型的有效性。曲荣华[8]为了解决指数间的共线性问题, 对每个指数分别进行简单线性回归得到各自的系数, 再通过最优化处理得到系数前的比重, 最后对各系数按比重加权求和得到仓位估计值。目前已有的研究主要是通过最优化法来解决变量间的多重共线性的问题, 本文在二次规划法的基础上引入逐步回归法和 Lasso 回归法, 通过剔除有共线性影响的变量来实现对基金仓位更准确的测算。

2. 数据及符号说明

2.1. 数据说明

本文的数据来源于 wind 金融终端。选择投资类型为普通股票型和偏股混合型的基金, 时间为 2018 年 1 月 1 日到 2018 年 12 月 31 日, 剔除 2018 年之后成立的基金以及正处于清盘期或者建仓期的基金。根据获取的 2018 年四个季度的基金净值数据计算日收益率作为因变量, 根据四个季度的中信 29 个一级行业指数每个交易日的收盘价数据计算日收益率作为自变量, 通过建立变量间的回归模型实现对基金仓位的测算。

2.2. 符号说明

表 1 给出了本文涉及到的相关符号的解释说明。

Table 1. Symbolic explanation

表 1. 符号说明

符号	释义
T	回归时间窗口的长度(本文 T 为一个季度)
N	行业指数的个数
$R_{i,t}$	第 i 个行业指数第 t 日的收益率
$R_{f,t}$	单只基金在第 t 日的收益率
γ_i	单只基金投资于指数 i 的股票资产比例
γ_d	单只基金投资于各指数的股票资产比例总和下限
γ_u	单只基金投资于各指数的股票资产比例总和上限
ε	模型拟合的残差项
λ	Lasso 回归的惩罚系数

3. 多元线性回归模型

以基金的日收益率作为因变量, 分别对中信一级行业指数以季度为周期进行回归, 模型为:

$$R_{f,t} = \gamma_1 R_{1,t} + \gamma_2 R_{2,t} + \cdots + \gamma_N R_{N,t} + \varepsilon$$

对得到的系数 γ_i , $i = 1, 2, \dots, N$ 进行求和即为测算的基金仓位。这里多元线性回归模型是作为下面要介绍的二次规划、逐步回归和 Lasso 回归这三种模型的理论基础。

4. 二次规划

二次规划模型可以通过多元线性回归模型转化得到, 它的本质就是带约束的线性回归。优化目标是使回归模型的均方误差最小。即

$$\min_{\gamma} \frac{1}{2T} \sum_{t=1}^T \left(\sum_{i=1}^N \gamma_i R_{i,t} - R_{f,t} \right)^2$$

由于基金投资于单个指数的股票资产比例是在(0,1)之间, 并且证监会对公募基金仓位有一定的范围限制, 因此可以得出系数的约束条件, 如下所示:

$$\begin{aligned} \gamma_d &\leq \sum_{i=1}^N \gamma_i \leq \gamma_u \\ 0 &\leq \gamma_i \leq 1, \quad i = 1, 2, \dots, N \end{aligned}$$

分别对每只基金进行二次规划模拟系数, 并计算每只基金对应模型的 R^2 值。选择 R^2 大于 0.5 的模型对系数进行求和, 得到对应基金的仓位值。图 1 和图 2 分别给出了用二次规划法测算的偏股混合型和普通股票型的基金仓位与季报公布的实际基金仓位值的误差分布。

从四个季度仓位测算的误差分布来看, 二次规划法测算的误差大部分分布在(5%, 15%)区间。因为基金仓位在短时间内变动本身不会太大, 所以如果模型测算的误差达到 5% 以上说明测算效果不好。

从结果来看, 由于二次规划法在测算系数时没有考虑自变量的多重共线性(即行业指数间的多重共线性), 所以产生了较大误差。图 3 是中信 29 个一级行业指数间的相关系数热图:

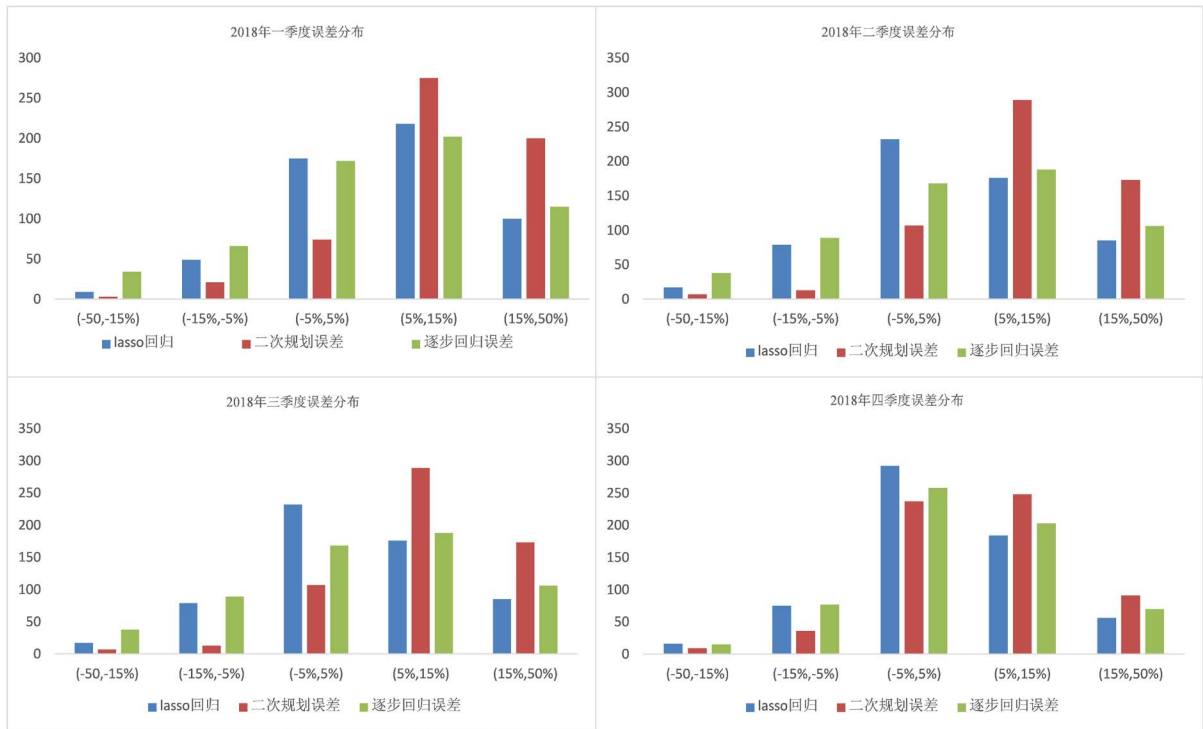


Figure 1. Error distribution of partial-equity hybrid funds
图 1. 偏股混合型基金误差分布

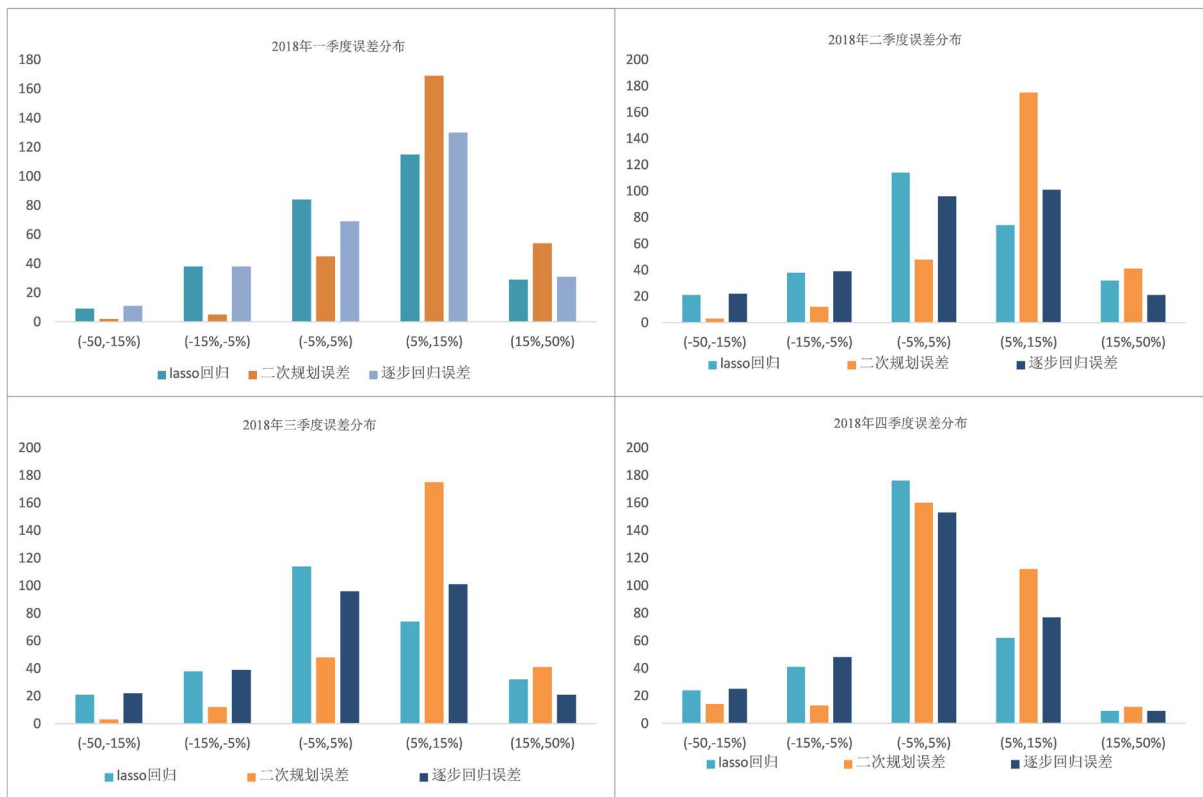


Figure 2. Error distribution of common stock funds
图 2. 普通股股票型基金误差分布

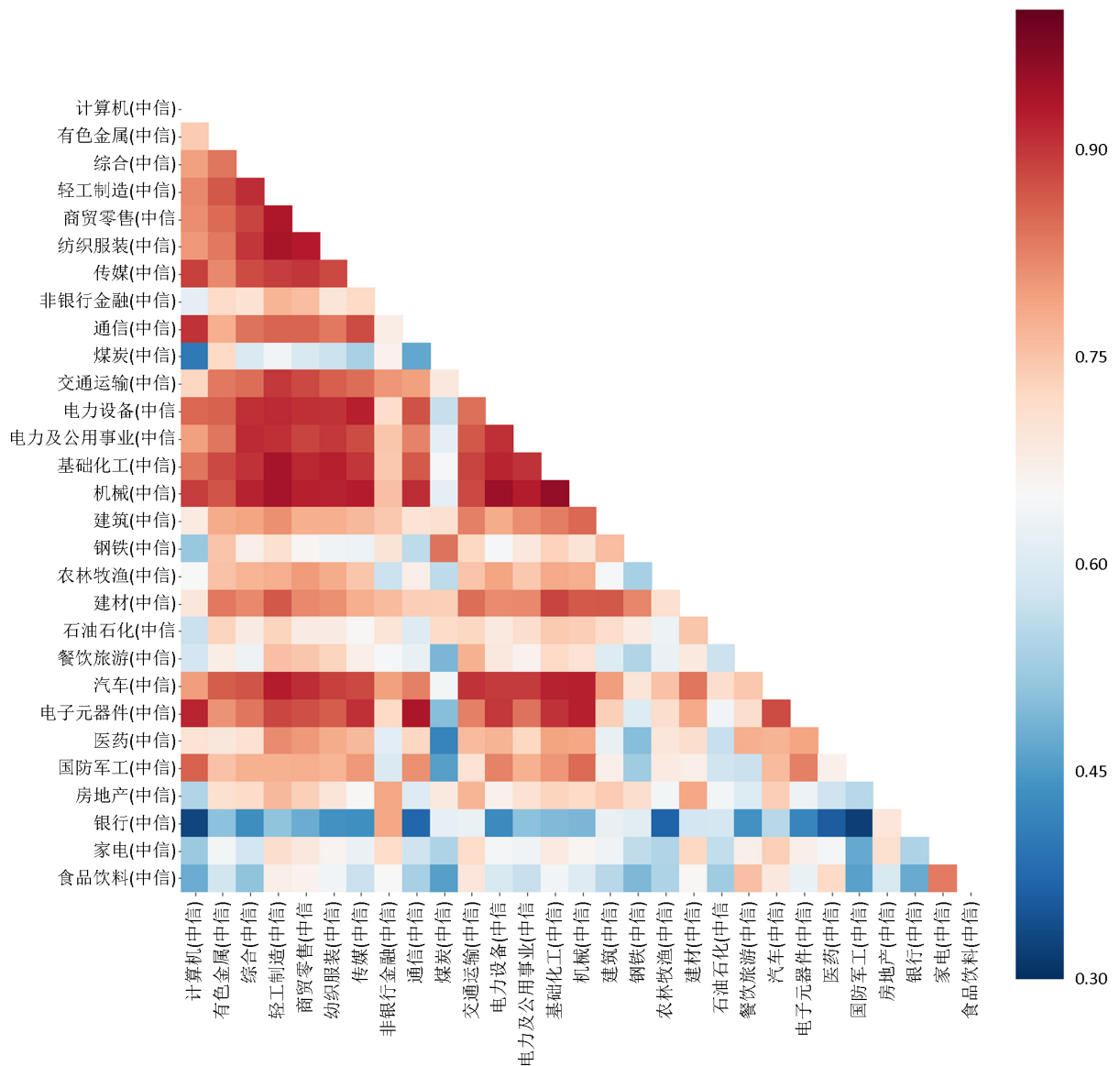


Figure 3. Relevance coefficient of industry index

图 3. 行业指数相关系数

从图 3 中可以看出行业指数之间存在较大的相关性, 所以对仓位进行测算时需要解决多重共线性的问题, 因此在二次规划法的模型基础上引入了改进的模型。

5. 改进的模型

为了降低行业指数之间的多重共线性, 下面采用逐步回归法和 Lasso 回归法来拟合基金日收益率与行业指数日涨跌幅之间的回归模型。

5.1. 逐步回归

逐步回归基本思想是有进有出, 每次引入或剔除自变量, 都对原模型和新模型进行 F 检验。如果新自变量的引入能够显著提升模型的解释能力, 那么引入该自变量; 如果原自变量的剔除不会显著降低模型的解释能力, 那么剔除该自变量。直到不再有变量被选入或剔除为止, 保证最后所得的回归子集是最

优子集。通过对行业指数与单只基金的相关系数从大到小排序, 依次引入模型。当 p 值小于 0.05 则保留变量, 否则剔除该变量。

在逐步回归的过程中, 逐渐剔除掉共线性较强的成份(行业指数), 保留相对独立的成份, 假设基金只在这些成份上进行配置, 将这些成份的回归系数加总得到基金仓位预测值。

逐步回归的模型得到的 R^2 值均大于 0.9, 保留下来的变量系数也通过显著性检验。图 1 和图 2 分别给出了用逐步回归法测算的偏股混合型基金和普通股票型基金误差的分布。可以看出此方法下模型误差在(-5%, 5%)区间的数量增加, 而在(5%, 15%)区间的误差数量明显减少, 说明逐步回归的结果相对于二次规划法的结果误差小。

表 2 给出易方达国防军工这只基金通过逐步回归法得到的系数值, 可以看出该基金在国防军工行业上的投资占比明显大于其他行业, 属于行业配置较为集中的基金。因此用逐步回归法测算的结果比较符合实际, 但对于行业配置比较分散的基金则估计效果就不太理想。

Table 2. Estimation coefficient

表 2. 估算系数

	20180331	20180630	20180930	20181231
国防军工	67.00%	72.71%	69.81%	69.61%
电子	34.00%	48.15%	33.49%	0
计算机	0	0	13.40%	33.45%
通信	-23.33%	0	-25.66%	0
机械	42.76%	-21.05%	0	0
汽车	17.95%	-20.32%	-46.00%	-39.79%
综合	10.34%	0	-33.80%	-20.33%
...				

5.2. Lasso 回归

Lasso 回归是引入 L1 正则化项的线性回归, 是多元线性回归的有偏估计, 适用于共线性数据, 它可以通过将系数进行压缩直到零从而达到剔除部分自变量的目的。Lasso 回归的结果对正则化系数 λ 取值敏感。实际使用过程中, 需要通过交叉验证的方法选择合适的 λ 值。

以 K 折交叉验证为例, 将样本随机分成 K 份, 每次取 K-1 份作为训练样本拟合 Lasso 回归模型, 计算该模型在剩余 1 份验证样本的均方误差; 重复 K 次, 得到该模型在 K 组验证样本的均方误差均值。对一定范围内可能的 λ 值进行遍历, 选择验证样本均方误差均值最小的 λ 值, 作为最终的 Lasso 回归正则化系数。Lasso 回归的损失函数为:

$$J(\gamma) = \frac{1}{2T} \sum_{t=1}^T \left(\sum_{i=1}^N \gamma_i R_{i,t} - R_{f,t} \right)^2 + \lambda \sum_{i=1}^N |\gamma_i|$$

这里根据交叉验证结果得出惩罚系数 $\lambda = 0.1$ 时 Lasso 回归得到的样本均方误差最小, 且回归的 R^2 值均在 0.9 以上。

图 1 和图 2 分别给出了用 Lasso 回归法测算的偏股混合型基金和普通股票型基金的误差分布。该模型得出的误差分布与逐步回归的误差分布相似, 比二次规划法的误差小, 说明这两种方法在此问题上得到的验证结果都优于二次规划法的结果。

图 4 和图 5 是根据对时间进行滚动得到的求和仓位再加权平均与实际平均仓位的对比图。因为基金仓位只在每一季度末的季报公布, 所以 2018 年只有四个季度末的数据。



Figure 4. Position comparison of partial-equity hybrid funds

图 4. 偏股混合型基金仓位对比

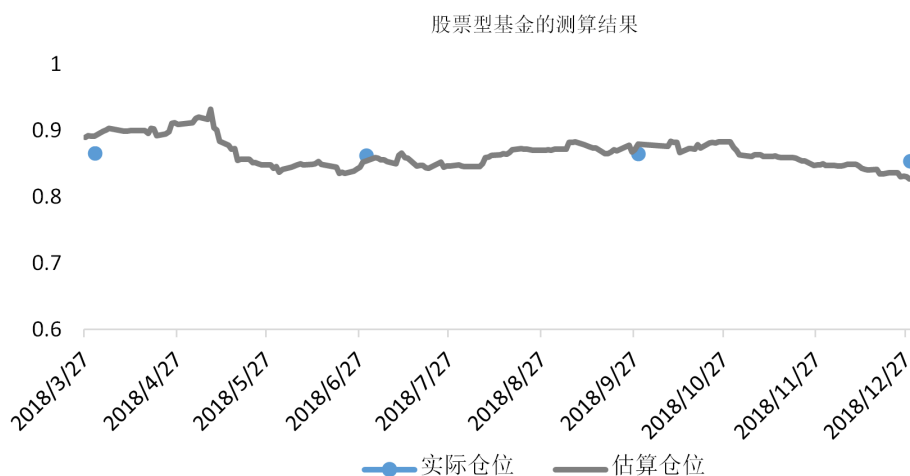


Figure 5. Position comparison of common stock funds

图 5. 普通股票型基金仓位对比

图 4 和图 5 折线图与实际仓位值对比可以看出 Lasso 回归测算的基金的平均仓位非常接近真实值。

6. 模型结果对比

表 3 给出了这三种方法测算的平均仓位与实际四个季度报告公布的平均仓位的绝对误差数据, 从表中可以看出逐步回归和 Lasso 回归的误差相比二次规划法的误差要小, 误差基本都在 0 到 5% 之间。说明逐步回归和 Lasso 回归测算的结果对于判断市场基金平均仓位走势有比较大的参考价值。

综合来看, 这三种方法都可以测算出公募基金的仓位。二次规划法可以得到基金在每个指数上的投资比例, 但是这种方法面临的一个问题是自变量之间的多重共线性。逐步回归法通过逐步筛选变量能够降低多重共线性, 它对于行业配置比较集中的基金的仓位的测算效果比较好, 但是对于行业配置分散的基金仓位的测算效果就不太理想。Lasso 回归通过对系数加入惩罚项, 避免估计值过大影响效果, 并且通

Table 3. Error comparison of average position
表 3. 平均仓位误差对比

截面日		20180331			20180630			20180930			20181231		
仓位及 误差(%)		估算 仓位	真实 仓位	误差	估算 仓位	真实 仓位	误差	估算 仓位	真实 仓位	误差	估算 仓位	真实 仓位	误差
二次 规划	偏股	95.15	84.67	10.48	91.77	80.80	10.97	90.49	80.56	9.93	83.32	77.01	6.31
	普通	93.79	84.38	9.41	95.09	86.08	9.01	92.77	85.26	7.51	89.22	84.67	4.55
逐步 回归	偏股	87.07	83.15	3.92	85.90	80.80	5.1	83.56	80.56	3	80.55	77.01	3.54
	普通	91.77	86.36	5.41	87.56	86.08	1.48	85.49	85.26	0.23	84.41	84.67	-0.26
Lasso 回归	偏股	87.76	83.57	4.19	85.69	80.80	4.89	85.51	80.56	4.95	80.11	77.01	3.1
	普通	89.91	86.51	3.4	89.12	86.08	3.04	87.41	85.26	2.15	84.81	84.67	0.14

过绝对值压缩来剔除变量,降低多重共线性,但是这种方法剔除变量是完全基于数据的基础上进行的,并没有涉及单个基金实际持仓信息,从而有可能剔除关键变量。

这三种模型都是通过建立基金日收益率与指数日涨跌幅的回归模型,从而可以测算基金的仓位。从绝对误差角度来看,逐步回归和 Lasso 回归测算结果的误差比二次规划法测算的误差都小,所以如果想要估算单只基金的仓位变动情况或者预判整个后市的动向变化,采用逐步回归和 Lasso 回归效果会更好。当涉及行业配置比较集中的基金仓位的测算时,用逐步回归法得到的结果会更符合实际预期。对行业配置分散的基金进行仓位测算时可以用 Lasso 回归。

7. 结论

为了避免被投资者或者竞争者摸清自身的投资策略,基金经理可能会在季末进行临时调仓而达到“粉饰”季报的目的,或者基金经理更换人选导致投资风格大变都会影响模型的估计效果。但模型整体误差近似呈现正态分布,所以在计算平均误差的时候可以通过增加参与测算的基金的数量来降低误差。

目前已有的资料中关于基金仓位的研究比较少,原因在于基金净值和持仓相关的信息可获取的非常少,对于研究的开展比较困难。本文主要就是基于二次规划法、逐步回归以及 Lasso 回归这三种方法根据可获得的公募基金的净值数据来测算基金的仓位。根据仓位的变动趋势来判断后市动向从而作出相应的投资决策。在后续研究中,可以增加时间跨度、数据量,再结合基金持仓方面的信息进一步优化方法。

参考文献

- [1] Markowitz, H.M. (1952) Portfolio Selection. *The Journal of Finance*, 7, 77-91.
<https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- [2] Treynor, J.L. (1965) How to Rate Management Investment Funds. *Harvard Business Review*, 43, 63-75.
- [3] Fama, E.F. and French, K.R. (2015) A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, 116, 1-22.
<https://doi.org/10.1016/j.jfineco.2014.10.010>
- [4] 杨艳军, 刘琦. 开放式股票型基金仓位配置模型应用研究[J]. 中南大学学报(社会科学版), 2011, 17(5): 49-54.
- [5] 何青峰. 探究开放式股票型基金仓位配置模型应用[J]. 中国外资, 2012(24): 24.
- [6] 马永谱 19 条微博: 基金仓位测算之我见[J]. 股市动态分析, 2011(23): 66-67.
- [7] 海琴, 谢怀军. 开放式股票型基金的仓位预测研究[J]. 中国证券期货, 2013(8): 36-37.
- [8] 曲荣华, 马莉媛, 詹从赞, 童行伟. 基金仓位估算的修匀模型[J]. 数理统计与管理, 2014, 33(5): 932-941.