

# Empirical Study on Risk Factors for Long-Only Equity Hedge Funds in China

Xi Zhao<sup>1</sup>, Wena Zhang<sup>1</sup>, Yi Hong<sup>2</sup>

<sup>1</sup>China Academy of Finance Research, Shanghai Jiao Tong University, Shanghai

<sup>2</sup>Department of Mathematical Sciences, Xi'an Jiaotong Liverpool University, Suzhou Jiangsu

Email: yi.hong@xjtlu.edu.cn

Received: Jan. 2<sup>nd</sup>, 2019; accepted: Jan. 18<sup>th</sup>, 2019; published: Jan. 25<sup>th</sup>, 2019

---

## Abstract

This paper utilizes the multi-factor model to identify risk factors of long-only equity hedge funds in China. These factors cover asset based style factors, liquidity factor, sentiment factors and macro factors. In addition to the conventional OLS model, this paper also employs Stepwise and LASSO regression models to select the subset of effective factors for each hedge fund sample. The results show that these factors can explain over 50% of the return variations of most long-only equity hedge funds in terms of adjust R<sup>2</sup>s with different sets of effective factors. Meanwhile, from the perspective of statistical significance ( $|t| > 2$ ), about 88%, 64% and 41% of long-only equity hedge funds are characterized mainly by the asset-based style factors, sentiment factors and macro factors respectively, whereas about 13% of hedge funds are explained significantly by the liquidity factor.

## Keywords

Hedge Fund, Performance-Attributed Factors, Return, Trading Strategy

---

# 中国股票多头私募证券投资基金的风险因子研究

赵 羲<sup>1</sup>, 张文鹭<sup>1</sup>, 洪 毅<sup>2</sup>

<sup>1</sup>上海交通大学中国金融研究院, 上海

<sup>2</sup>西交利物浦大学数学科学系, 江苏 苏州

Email: yi.hong@xjtlu.edu.cn

收稿日期: 2019年1月2日; 录用日期: 2019年1月18日; 发布日期: 2019年1月25日

## 摘要

本文利用多因子模型来甄别影响中国私募证券投资基金行业中股票多头基金业绩的风险因子。这些因子包括资产风格因子、流动性因子、市场情绪因子和宏观因子四类。除了经典OLS模型,本文还采用Stepwise和LASSO模型为每个基金样本选取有效因子组合。研究结果表明这些因子对大部分股票多头私募证券投资基金产品业绩的解释度(Adj R<sup>2</sup>)超过50%,而且不同基金的有效因子组合存在差异。同时,从统计显著性(|t|>2)来看,股票多头私募证券投资基金业绩受资产风格因子、市场情绪因子和宏观因子三类因子影响的样本比例分别为88%、64%和41%,而只有13%左右基金受流动性因子影响。

## 关键词

私募证券投资基金, 业绩解释因子, 收益, 交易策略

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

中国私募证券投资基金近年来快速发展,截至2018年11月底,备案的中国私募证券投资基金管理人达到8,966家,备案基金产品数量达到36,053只,资产管理规模达到2.26万亿人民币<sup>1</sup>,私募证券投资基金行业已成为我国构建多层次资本市场的一个核心支柱,同时私募证券投资基金作为一类重要资产,为居民财富的有效管理提供了新渠道和工具,因而受到众多投资者的青睐。由于中国资本市场的基本特征(如受宏观政策影响较大)和所处发展阶段(股票市场产品发行持续增加而提供风险对冲工具的衍生品市场则发展相对缓慢)等原因,目前主流的私募证券投资基金一般采用股票多头交易策略,这些基金的管理规模在整个私募证券投资基金行业占比达到80%以上。因此,本文通过实证分析来甄别影响股票多头策略私募证券投资基金的风险因子,并以此来剖析私募证券投资基金产品的业绩风险特征,为投资者进行有效资产配置方面提供有益参考。

基金业绩风险因子相关研究一直是基金研究领域的热点问题。这方面研究从20世纪90年代初开始一直持续到现在,其研究方法主要采用归因分析的方法,但不同学者采用的因子模型有所区别。Sharpe (1992) [1]最早采用多因子模型对美国公募基金做业绩归因,Fung和Hsieh (1997) [2]首次提出把Sharpe (1992) [1]的因子模型扩展到美国对冲基金(在中国这类基金被称为“私募证券投资基金”)对其进行业绩归因。Fung和Hsieh (1997) [2]认为对冲基金的收益来自于所投资的资产收益、策略和杠杆,作者通过主成份分析提取能够解释对冲基金收益差异的前5个主成份,并构建与每个主成份相关性最高的对冲基金组合,以此作为风格因子。Fung和Hsieh (2002) [3]进一步提出基于资产的风格因子,采用可交易资产构建对冲基金的风格因子,而Fung和Hsieh (2004) [4]则完善了这一分析框架,确定了7个基于资产的风格因子,其中股票资产风格因子采用市场(标准普尔500指数)和规模因子,固定收益资产风格因子采用10年国债收益率和信用利差,而趋势跟踪资产风格因子采用基于债券、货币、大宗商品等资产的回顾式期权。其研究发现这些因子可以解释超过90%的股票多空、固定收益及期货投资基金(CTA)等不同策略基金的收益和风险来源。

<sup>1</sup>根据基金业协会截止2018年11月底公布数据统计得到。

除上述资产风格因子外,流动性风险因子被发现是影响资产价格和基金业绩的一个重要因子。Pastor 和 Stambaugh (2003) [5]发现股票预期收益与其对市场总体流动性的敏感度有关。Brandon 和 Wang (2013) [6]发现管理能力强的对冲基金具有更好的业绩,但这种超额收益很大程度上来自于承担了更多的流动性风险。

另一方面,投资者情绪也对资产价格和基金业绩产生重要影响,这方面研究最早可以回溯到上世纪 30 年代,而到本世纪初又有一些重要发现。Baker 和 Wurgler (2006) [7]发现美国股票市场情绪的大幅波动对一些估值较为主观且难以套利的股票收益影响更大,比如小市值股票、刚上市股票、高波动股票、不分红股票、极高成长性股票等。Baker、Wurgler 和 Yuan (2011) [8]发现除美国之外,加拿大、法国、德国、日本、英国股票市场中也有类似的现象。胡昌生和池阳春(2012) [9]对中国 15 个市场情绪变量对比分析发现大多数单一变量都能在一定程度上反应投资者情绪,但都不完整。该文对其中 4 个主要指标进行分析,结果显示换手率和新开户数是非理性情绪指标,而封闭基金折价率和消费者信心指数是理性情绪指标。Chen, Han 和 Pan (2015) [10]用情绪因子解释对冲基金的收益差异,发现对情绪因子暴露大的对冲基金的后续收益高于情绪因子暴露小的对冲基金。

此外,研究发现宏观因子也可以用于解释资产收益。这方面研究始于上世纪八十年代末。Fama 和 French (1989) [11], Keim 和 Stambaugh (1986) [12], Campell (1987) [13]等提出了一系列对于股票和债券市场回报具有解释和预测力的宏观经济指标,如美国国债利率,信用利差,股息率,利率结构等等。Bali 等(2011) [14]认为宏观经济周期影响资产价格变化,因此提出用宏观经济指标解释和预测对冲基金的收益差异,发现违约风险溢价和通货膨胀率可以有效预测对冲基金的收益差异。Avramov 等(2011) [15]以宏观经济变量(特别是信用违约利差和波动率指标)为基础预测对冲基金业绩,并以这种预测构建对冲基金组合产生样本外超额收益。Avramov 等(2013) [16]进一步发现信用违约利差、分红收益率、波动率指数和整体资金流等四个宏观变量可以有效解释和预测对冲基金的收益差异。

当前,中国证券投资基金的业绩风险因子研究也主要集中于归因研究,但主要体现在公募基金领域,如沈维涛和黄兴李(2001) [17]应用国外基金业绩评价中普遍采用的风险调整指数法、T-M 模型和 H-M 模型。李红权和马超群(2004) [18]尝试采用 JensenAlpha、T-M、H-M 等模型对公募证券投资基金的绩效表现进行衡量。屠新曙和朱梦(2010) [19]在 Fama-French 框架内对证券投资基金的业绩进行归因分析。郑琳(2013) [20]采用 Brinson 模型从资产配置,个股选择和交互角度进行对公募证券投资基金进行业绩归因。

由于中国私募证券投资基金起步较晚和信息披露不规范等原因,针对私募证券投资基金业绩的风险因子研究比较匮乏,本文的研究成果填补国内当前该领域的不足。此外,本文在统计方法方面亦做出贡献。大多数文献做业绩归因都是基于经典 OLS 模型,采用同一组因子去解释所有基金样本收益,然而即使相同策略类型的不同基金在投资方法也存在一定差异,因此应该选取不同因子组合来解释不同基金。为弥补以上不足,本文对私募证券投资基金样本做业绩回归分析时,分别采用 OLS, Stepwise 和 LASSO 方法,针对不同基金选取因子组合,并综合对比三种方法的结果,以加强检验结果的可靠性。

## 2. 数据

### 2.1. 私募证券投资基金样本数据

由于私募证券投资基金产品通常采用数据“主动”报备的方式,这就导致了基金产品信息披露不规范,而且数据提供商在收集整理过程中也容易出现数据不一致问题。鉴于此,本研究采用上海交通大学中国私募证券投资研究中心(CHFRC)的研究数据库来获得高质量的私募证券投资基金产品数据,特别是存续期较长的私募证券投资基金产品样本。该数据库整合了万得信息有限公司(上海)、融智投资顾问有限公司(深圳)、朝阳永续科技有限公司(上海)和大智慧科技有限公司(上海)四个主流数据服务商提供的私募证券投资基金产品的业绩数据,并针对基金净值缺失或错误以及分类错误等情况,研究中心通过一系列

数据清洗工作提高数据质量。在此基础上, 研究中心进一步研发出基于私募证券投资基金产品的基金策略数据库, 涵盖股票、固定收益以及管理期货等私募基金产品在内共 18 种交易策略。表 1 报告了在 2011/01/01 至 2018/10/31 样本期内, 各主要交易策略类型的私募证券投资基金产品数量统计。由表 1 可见, 不同策略类型的有效产品数量(如存续期高于 6 个月)差异很大, 且数据质量较好(如存续期大于 36 个月)的私募证券投资基金产品样本占比不高, 仅达到 22.05%。更重要地, 表 1 显示当前国内私募证券投资基金产品仍然以股票多头策略为主, 其在存续期高于 6 和 36 个月的基金样本中占比分别为 81.96%和 86.62%, 远高于其他策略的样本数量。因此, 本文将研究对象聚焦于采用股票多头策略的私募证券投资基金产品, 从而避免由于其他类型样本基金数量偏少而导致统计分析可靠性问题。

**Table 1.** Sample statistics of private equity funds with investment strategies

**表 1.** 不同私募证券投资基金策略的样本量统计

策略类型	策略说明	产品数量 (超过 6 个月存续期)	产品数量 (超过 36 个月存续期)
股票多头	股票多头组合, 没有任何做空或对冲	6476	1472
股票定向增发	参与股票定向增发, 持有到期解禁后退出	158	15
股票套利	以股票 ETF 套利或分级基金套利为主	101	19
股票市场中性	建立股票多头组合, 并用股指期货对冲市场 (沪深 300/中证 500)风险敞口	302	75
股票多策略	多个股票投资策略混合	183	26
债券多头	挑选并买入利率债或信用债组合, 信用债大部分持有到期	125	20
管理期货	通过各种方法交易期货, 包括股指期货和商品期货	381	49
宏观对冲	通过对宏观趋势的判断选择不同资产和策略实现收益	86	23
组合基金	以私募证券投资基金产品为投资标的	89	6

进一步地, 表 2 报告了存续期超过 36 个月的 1472 个股票多头类私募证券投资基金产品的描述统计。这些私募证券投资基金样本的业绩具有如下特征。首先, 年化对数收益平均达到 4.7%左右, 但收益分布较宽, 最好和最差的基金收益差距很大。其次, 风险水平高, 年化波动率平均超过 20%, 而且标准差较小, 说明大部分基金的收益波动都比较高。因此基金风险调整后的收益(即夏普比率)不高, 只有 0.04, 其标准差为 0.48, 说明绝大部分基金样本的夏普比率都小于 1。最后, 以月度收益计算的最大回撤 0.17 左右, 中位数为 0.15, 均小于年化波动率。

**Table 2.** Descriptive statistics of performance of private equity funds

**表 2.** 私募证券投资基金产品业绩的描述统计

指标	平均值	中位数	最大值	最小值	标准差
年化收益	0.0470	0.0534	0.6407	-0.5254	0.1028
波动率	0.2295	0.2164	0.7518	0.0203	0.0846
夏普比率	0.0433	0.0658	2.3470	-1.9543	0.4810
最大回撤	0.1749	0.1541	0.9306	0	0.1594

## 2.2. 因子类型与构造

为避免选取大量无效因子, 本文归纳前述文献中相关研究所用到的有效因子, 选取 4 大类因子, 分别是资产风格因子、流动性因子、情绪因子、以及宏观因子。由于所研究对象是股票多头策略私募证券投资基金, 因此资产风格因子中只选取了 Fung 和 Hsieh (2004) [4]中解释股票类对冲基金的因子(即市场

因子和规模因子)。流动性因子采用经典的 Pastor 和 Stambaugh (2003) [5]定义的流动性因子。情绪因子则按照胡昌生和池阳春(2012) [9]的分析结果, 选取换手率、股票新开户数、封闭基金折价率和消费者信心指数, 同时增加融资融券余额作为补充。宏观因子在国外研究中证明有效的因子包括信用风险溢价、通货膨胀率、股息率、波动率指数(VIX)、整体资金流等 5 个指标, 由于中国波动率指数和中国私募证券投资基金的资金流动数据不可获取, 本文主要采用前面 3 个指标作为宏观因子。同时, 本文还增加了代表货币政策的广义货币 M2 供应量, 以及代表经济增长的工业增加值两个指标作为宏观因子。表 3 报告了本文采用各类因子的构建方法及数据来源。

**Table 3.** Factor types and definitions

**表 3.** 因子类型及来源说明

因子类型	因子	计算方法	数据来源
资产风格	HS300 全收益指数	沪深 300 全收益指数取 log 变化率	Wind
	规模因子	Fama-French 三因子中 Size 因子	CSMAR
流动性	流动性风险	Pastor-Stambaugh 整体流动风险测度	CSMAR
	封闭基金折价	取封闭式基金按成交额加权折价率	Wind
	上证换手率	取上证换手率	Wind
情绪	新增开户数	取新增开户数量 log 变化率	Wind
	消费者信心指数(CCI)	取 CCI 的 log 变化率	Wind
	融资融券余额	取融资融券余额 log 变化率	Wind
	信用风险(信用利差)	基于 AAA 信用债和 AA 信用债价格指数计算	Wind
	居民消费价格指数 CPI	取 CPI 环比 log 变化率	Wind
宏观	HS300 股息率	沪深 300 股息收益率	Wind
	广义货币 M2	取 M2 供应量环比 log 变化率	Wind
	工业增加值	取同比 log 变化率	Wind

部分因子计算方法说明如下:

### 2.2.1. Log 变化率

考虑时间序列平稳性以及量纲等问题, 将指数及实值统计量类数据取 Log (对数)变化率作为指标。回报率/变化率的计算公式如下:

$$\log(ret_t) = \log\left(\frac{x_t}{x_{t-1}}\right) \quad (1)$$

其中,  $x_t$  和  $x_{t-1}$  分别表示相应变量在  $t$  和  $t-1$  时刻的状态值。

### 2.2.2. 信用风险指标

本文采用中短期不同等级的信用利差指标衡量信用违约风险溢价, 即计算中债企业债 AAA 财富指数 ( $CB\_Index^{AAA}$ ) (1~3 年)的 Log 回报率与中债企业债 AA 财富指数 ( $CB\_Index^{AA}$ ) (1~3 年)的 Log 回报率的差值。其计算公式为:

$$Csprd_t = \log\left(\frac{CB\_Index_t^{AA}}{CB\_Index_{t-1}^{AA}}\right) - \log\left(\frac{CB\_Index_t^{AAA}}{CB\_Index_{t-1}^{AAA}}\right) \quad (2)$$

### 2.2.3. 封闭基金折价指标

取 2011 年 1 月起封闭式基金数据, 取折价率在-50%到+50%之间的基金产品以排除错误数据和特殊情况, 并依月末当日成交量作为权后重计算封闭基金折价率。

$$Cefdisrate = \sum_1^n \frac{volume_n * discontrate_n}{totalvolume} \quad (3)$$

其中  $volume$  为基金在月末日的成交量,  $discontrate$  为其折价率,  $totalvolume$  为样本封闭式基金的当天总成交量。

### 2.2.4. Pastor-Stambaugh 流动性指标

根据 Pastor 和 Stambaugh (2003) [5] 提出的整体流动性风险测度计算。其公式为:

$$r_{i,d+1,t}^e = a_{i,t} + b_{i,t} r_{i,d,t} + c_{i,t} \text{sign}(r_{i,d,t}^e) v_{i,d,t} + \epsilon_{i,d+1,t} \quad (4)$$

其中  $r_{i,d,t}$  是股票  $i$  在  $t$  月  $d$  日的回报,  $r_{i,d,t}^e$  是  $r_{i,d,t}$  减去依市值加权后的市场投资组合回报的超额部分,  $v_{i,d,t}$  是股票  $i$  在  $t$  月  $d$  日的交易额。而  $c_{i,t}$  为所要估计的流动性风险指标。本文中使用的为月频的全市场流动性风险指标, 其为个股流动性风险均值。

### 2.2.5. 新开户数量变化率

取新开证券账户数量与存量的比值, 即:

$$rate_t = \frac{\Delta x_t}{x_{t-1}} \quad (5)$$

其中  $x_{t-1}$  为截止至  $t-1$  时刻的账户总开户数量,  $\Delta x_t$  为  $t$  时刻新开户数量。表 4 报告了各因子在 2011 年 1 月 1 日至 2018 年 10 月 31 日的描述性统计量。

Table 4. Descriptive statistics of factors

表 4. 因子统计描述

因子	均值	标准差	偏度	峰度	最小值	最大值
HS300 全收益指数	0.0018	0.0677	-0.0190	2.2313	-0.2362	0.2296
规模因子	0.0060	0.0553	0.0950	5.7531	-0.2354	0.2417
流动性风险	-0.5049	0.6644	-0.1930	2.6340	-2.8674	1.8099
封闭基金折价	0.0048	0.0108	2.0737	5.9155	-0.0176	0.0508
上证换手率	0.0057	0.0043	2.5391	6.5775	0.0019	0.0243
新增开户数	0.0106	0.0121	3.1922	12.4007	0.0017	0.0721
消费者信心指数	0.0018	0.0243	0.0463	0.2891	-0.0531	0.0773
融资融券余额	0.0436	0.1051	-1.0039	4.0152	-0.4259	0.2908
信用利差	0.0004	0.0013	-0.2456	2.0296	-0.0036	0.0048
居民消费价格指数 CPI	0.0019	0.0049	0.4654	0.6943	-0.0111	0.0159
HS300 股息率	0.0019	0.0031	2.0132	3.1963	0.0000	0.0138
广义货币 M2	0.0096	0.0101	0.4882	0.2650	-0.0128	0.0367
工业增加值	0.0802	0.0318	0.6022	1.7909	-0.0214	0.1931

表 5 报告了各因子之间相关性系数统计。可以看出, 这些因子之间的相关性普遍较低, 大部分相关系数在零左右, 相关系数高于 0.5 的只有两对因子。新增开户数与上证换手率之间相关性最高, 达到 0.75; 其次是 HS300 全收益指数与融资融券余额之间, 相关性系数为 0.58。因此线性回归分析中的共线性问题基本可以忽略。

Table 5. Correlation matrix of factors  
表 5. 因子间相关性系数统计

因子	HS300 全收益指数	规模因子	流动性风险	封闭基金折价	上证换手率	新增开户数	消费者信心指数	融资融券余额	信用利差	消费价格指数 CPI	HS300 股息率	广义货币 M2	工业增加值
HS300 全收益指数	1.00												
规模因子	0.08	1.00											
流动性风险	0.20	-0.09	1.00										
封闭基金折价	-0.04	0.15	0.03	1.00									
上证换手率	0.28	0.14	0.27	0.35	1.00								
新增开户数	0.12	0.18	0.25	0.28	0.75	1.00							
消费者信心指数	0.12	-0.04	-0.10	-0.22	-0.06	-0.08	1.00						
融资融券余额	0.58	0.30	0.08	0.02	0.12	-0.08	0.02	1.00					
信用利差	-0.06	0.11	-0.11	-0.14	-0.02	-0.07	0.06	0.06	1.00				
消费价格指数	-0.04	-0.13	-0.05	-0.17	-0.11	-0.19	0.25	-0.16	-0.14	1.00			
HS300 股息率	-0.06	-0.10	0.13	0.15	-0.07	-0.02	-0.09	-0.14	0.06	-0.18	1.00		
广义货币 M2	-0.04	-0.02	0.10	0.14	0.05	0.08	-0.15	0.06	-0.07	-0.06	-0.06	1.00	
工业增加值	0.04	0.00	-0.08	0.02	-0.24	-0.29	-0.08	0.42	0.02	-0.08	-0.05	0.21	1.00

### 3. 研究方法

#### 3.1. 研究方案设计

本文首先采用经典的 OLS 模型用全部因子解释每个私募证券投资基金样本的收益。这种方法存在两个潜在问题，一是由于因子数量较多而私募证券投资基金样本的收益数据时间序列长度不足，可能会导致回归结果有过拟合的现象；二是不同的私募证券投资基金由于投资策略差异可能具有不同的风险因子组合，因此用相同的因子组合去解释不同私募证券投资基金不合适，应该有针对性地筛选出不同的因子组合来解释不同私募证券投资基金。针对这些问题，本文分别采用 Stepwise 回归和 LASSO 回归的方法来挑选因子组合。Stepwise 回归的目标函数与 OLS 模型基本一致，可以直接对比解释度和因子显著性；而 LASSO 回归模型的目标函数中增加了惩罚项，并且最终追求交叉验证的结果最优，但导致其结果与 OLS 和 Stepwise 的结果无法直接对比。本文在用 LASSO 模型选取变量子集之后，再用 OLS 模型对变量子集进行回归检验，将其解释度与因子显著性与前两种方法结果对比分析。

#### 3.2. 回归模型介绍

##### 3.2.1. OLS 回归模型

Ordinary Linear Squared, 最小二乘法是以优化

$$\min \left\{ \frac{1}{N} \|y - X\beta\|_2^2 \right\} \quad (6)$$

为目标的线性回归模型。OLS 回归模型表述如下：

$$y_i = a + \sum_{k=1}^n b_k x_{ki} + \varepsilon_i \quad (7)$$

若将  $\{b_1, b_2, \dots, b_n\}$  记为向量形式  $\beta$ ,  $x$  和  $y$  组成的矩阵分别为  $X$ ,  $Y$ , 则  $\beta$  的估计值为  $(X^T X)^{-1} Y$ 。

本模型及其他模型中均为同期因子回归。

##### 3.2.2. Stepwise 回归模型

本文采用基于 Akaike information criterion (AIC) 的双向 Stepwise 回归。AIC 是一种衡量统计模型拟合优劣程度的判据：

$$AIC = 2k - 2 \ln(L) \quad (8)$$

其中  $k$  为模型中参数的数量，而  $L$  为模型的极大似然估计的似然函数值。

由于在回归模型中，加入解释变量会使得残差(RSS)单调减小，为防止过拟合以及评估有效解释变量，可以参考 AIC 的值来决定所用解释变量。筛选过程为在 OLS 回归模型中，通过增加/删减解释变量，计算每一次回归的 AIC，力求降低 AIC 直到最小值，根据线性回归模型的理论，此时模型作为有效，从而确定合理的解释变量集以及回归结果。

##### 3.2.3. LASSO 回归模型

###### 1) LASSO 回归模型描述

LASSO 即最小绝对值收敛和选择算子(Least absolute shrinkage and selection operator)，是一种同时进行特征选择和正交化的回归分析方法。最初由斯坦福大学统计学教授 Robert Tibshirani (1996) [21] 于 1996 年基于 Leo Breiman 的非负参数推断(Nonnegative Garrote, NNG)提出。相比于普通线性回归模型(OLS 或 GLS)，在应对可能存在多重共线性的数据集时，其预测准确性和可解释性有一定增强。附录 A 详细介绍 LASSO 回归模型。

## 2) 利用 LASSO 模型选取变量子集进行 OLS 回归

由于 LASSO 回归有变量选择的作用, 对每个基金产品进行 LASSO 回归后, 进一步将被 LASSO 选取的变量再做 OLS 回归, 结果可与其它回归方式进行比较。

### 3.3. 线性回归模型的误差调整

在实证分析阶段, 将上述回归模型分别应用于股票多头类私募证券投资基金产品样本。即对于每一基金样本, 对选取的解释变量集进行如下多因子 OLS 回归以确定模型解释度和因子显著性:

$$y_i = \tilde{a} + \sum_{k=1}^k \tilde{b}_k \tilde{x}_{kt} + \tilde{\varepsilon}_i \quad (9)$$

其中  $\tilde{x}$  表示选取的解释变量集。从表 5 可知, 解释变量共线性问题基本可以忽略。同时, 进一步对模型系数  $\tilde{a}$  和  $\tilde{b}$  的标准差运用 Newey-West 方法进行自相关性和异方差调整, 以识别统计显著的相关风险因子。最后, 在 R 环境里完成上述回归模型的计算与参数估计。

## 4. 实证研究结果及分析

### 4.1. 模型解释度

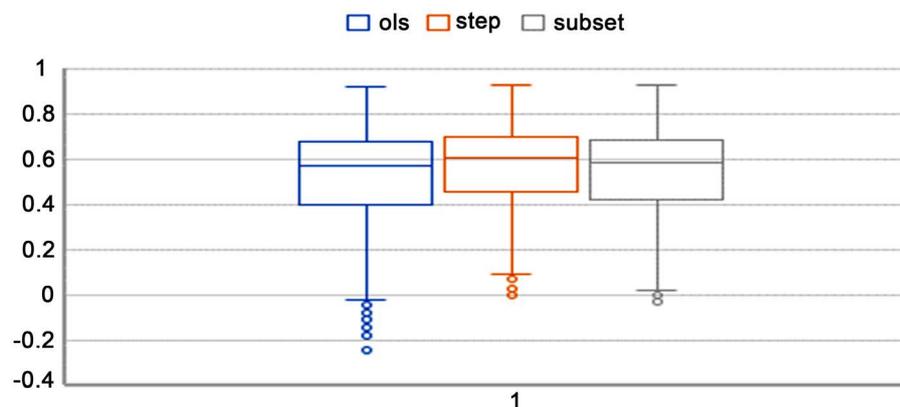
首先检验模型的解释度。表 6 报告了上述三个模型的 Adj R<sup>2</sup> 统计结果。结果显示三个模型对大部分基金样本都有较高解释度。它们的调整后 R<sup>2</sup> 的均值和中位数接近, 而且都超过 50%, 说明模型所选取的因子能够有效解释大部分股票多头策略基金的收益业绩。

**Table 6.** Model Adj R<sup>2</sup> statistics

**表 6.** 模型 Adj R<sup>2</sup> 统计结果

回归模型	平均值	中位数	最大值	最小值
OLS	0.5286	0.5712	0.9267	-0.2437
Stepwise	0.5780	0.6102	0.9348	0.0000
LASSO (subset)	0.5349	0.5893	0.9351	-0.0286

从 Adj R<sup>2</sup> 分位图(图 1)来看, Stepwise 回归模型解释度更好一些, 调整后 R<sup>2</sup> 集中在 45%到 70%之间, 而 OLS 模型的解释度稍差一些, 有少量调整后 R<sup>2</sup> 为负, 说明对少量基金样本的解释度偏低, 而 LASSO 回归模型的解释能力则介于上述两个模型之间。



**Figure 1.** Adj R<sup>2</sup>Quantiles of OLS, stepwise, and LASSO regression models

**图 1.** OLS, Stepwise 和 LASSO 回归模型的 Adj R<sup>2</sup> 分位图

## 4.2. 不同基金的风险因子差异

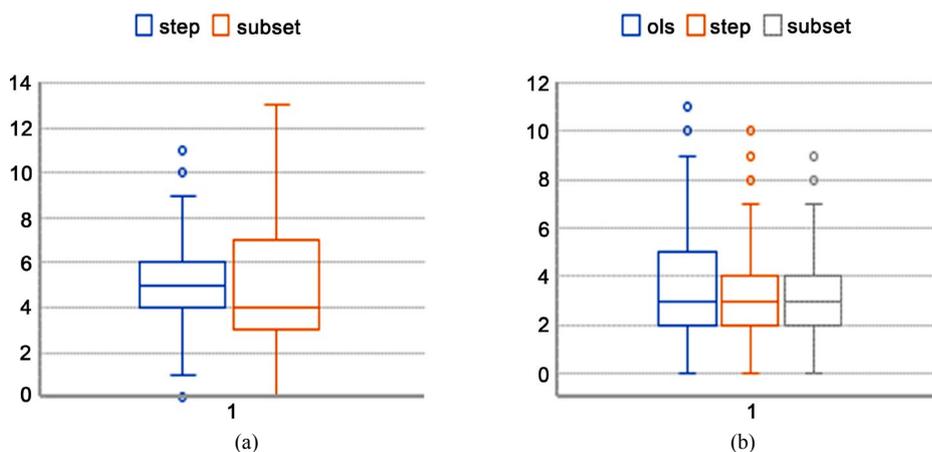
对比三种回归模型所选择的因子数量(表 7), 可以看出 Stepwise 模型和 LASSO 模型选择的因子数量远远少于 OLS 模型因子数量, 且没有影响其解释度。LASSO 所选取的因子数量与 Stepwise 相似, 但因子数量的中位数最小, 只有 4 个。从分位图上看(图 2(a)), Stepwise 模型中大部分基金样本只选择了 4~6 个因子, LASSO 模型选择了 3~7 个因子, 而按照之前解释度的分析, 这些因子可以解释大部分基金样本 50%以上的收益。

**Table 7.** Statistics of selected factors with three models (significant factors in parenthesis)

**表 7.** 三种模型选择因子数量统计结果(括号内为显著的因子数量)

回归模型	平均值	中位数	最大值	最小值
OLS	13 (3.46)	13 (3)	13 (11)	13 (0)
Stepwise	4.94 (3.18)	5 (3)	11 (10)	0 (0)
LASSO(subset)	5.07 (2.74)	4 (3)	13 (9)	0 (0)

从显著的因子数量来看(表 7), 三种方法的显著因子数量都比较小, 平均显著因子数量在 3 个左右。这说明针对大多数基金样本个体而言, 能够显著解释其收益的因子并不多, 大部分因子针对单个基金样本收益的解释力并不显著。而比较三个模型的统计显著因子数量的分位图(图 2(b)), 三个模型的显著因子数量非常相近, 大部分基金样本的因子显著数量只有 2~4 个。



**Figure 2.** Selected factors in Stepwise and LASSO Model (a) vs Significant Factors in Models (b)

**图 2.** Stepwise 和 LASSO 选中因子数量(a)和不同模型的统计显著因子数量(b)对比

从基金的角度看, 不同基金选取的因子数量和因子存在较大差异。这说明虽然研究样本都是股票多头策略, 但基金之间的风险特征还存在明显差异, 而这种风险特征差异可能来源于具体投资策略和风险管理能力方面的差异。

## 4.3. 不同风险因子的解释力差异

模型回归结果中, 不同风险因子的显著性有较大差异, 这种显著性差异是因子对私募证券投资基金样本收益解释力不同的表现。表 8 报告了每个因子在不同基金样本进行模型回归结果中显著( $|t| > 2$ )的次数。由表 8 可以看出, 四大类因子中, 资产风格类因子的解释力最高, 接近 90%的基金样本收益受到至少一个资产风格类型因子的影响。其中影响最大的是市场因子(沪深 300 全收益指数), 有 80%左右的基

金样本显著受市场因子影响,这也符合股票多头策略的基本特征,即较明显的市场风险因子暴露。规模因子(Size)也较为显著,有超过 50%的基金样本显著受规模因子影响,可见股票多头策略的收益有相当一部分是来源于小市值股票的风险溢价。所有个体因子中,除市场因子和市值因子外,其他因子的单独解释能力都不高,大部分因子能够显著影响基金样本的比例不到 20%,只有个别因子仅能显著影响 30%左右基金样本业绩。

从大类因子层面看,情绪因子的整体解释力比较高,超过 60%以上的基金样本收益显著受到至少一个情绪类因子的影响。投资者情绪对基金收益解释力高符合 Baker 和 Wurgler (2006) [7]的研究发现。美国股票市场情绪的大幅波动对部分估值主观且难以套利的股票收益影响更大,比如小市值股票、刚上市股票、高波动股票、不分红股票、极高成长性股票等。中国股票市场的中小市值股票普遍波动大,分红少,而且估值高又无法做空。同时中国股票多头私募证券投资基金倾向于投资中小市值股票,因此中国情绪因子对股票多头策略私募证券投资基金解释力较高符合实际情况。

宏观因子的解释力相对较弱,总体上有 40%左右的基金样本显著受到至少一个宏观因子影响,但单个宏观因子中,除了工业增加值之外,可以解释的基金样本比例只有 10%左右。主要原因可能是过去中国宏观经济增长和货币政策与股市表现没有直接关系。孙华好和马跃(2003) [22]的研究发现中国股市没有发挥货币传导机制作用。梁琪和滕建州(2005) [23]研究显示中国股市与经济增长没有必然因果关系。

特别需要指出的是,流动性因子在国外对冲基金的因子研究中表现比较显著,但对国内股票多头策略私募证券投资基金的解释力较弱。本文回归结果中,只有 13%左右的基金样本受流动性因子影响显著。主要原因可能是国内股票多头策略主要是基本面选股并长期持有策略为主,并不是市场流动性的提供者,因此受市场流动性的影响较小。

最后,回归模型中除所有因子之外还有截距项,其代表剔除风险因子后基金获取的真实阿尔法(Alpha)收益。三个模型结果显示有 20%~30%的基金产生能够 Alpha 收益。这一比例并不高,说明大部分私募证券投资基金的收益主要通过承担常规风险而获得,但也有部分基金具备独特投资能力来获得额外收益。

#### 4.4. 鲁棒性检验

为了进一步检验部分因子(特别是情绪因子)可能与市场因子存在某种非线性关系,即情绪因子有可能本质上也是市场因子的一种表现形式,本文通过增加市场因子的平方作为控制变量,即沪深(HS) 300 全收益指数收益的平方,来检验其他因子是否依然有效。检验结果由表 9 所示,即使增加市场因子的平方作为控制变量,其他因子依然有效,其显著解释的基金样本数量没有根本性变化。

## 5. 结语

本文采用资产风格因子、情绪因子和宏观因子等因子对中国私募证券投资基金的股票多头策略产品做业绩回归分析,进而甄别影响这些基金收益的主要风险因子。研究发现采用少量因子对大部分基金产品业绩的解释度(Adj  $R^2$ )超过 50%。同时,从统计显著性( $|t|>2$ )来看,在四大类因子中,资产风格因子显著性最高,对接近 90%的基金收益影响显著;情绪类因子的显著性也较高,对 60%以上的基金收益影响显著;宏观类因子的显著性较低,对 40%左右的基金收益影响显著;流动性因子的显著性最低,只对 13%左右的基金收益影响显著。股票市场因子是最普遍的因子,80%以上股票多头策略基金在市场因子上都有显著暴露。除股票市场因子外,各基金的因子暴露有所差异,原因是虽然同属于股票多头策略,但不同基金经理的投资方法/能力有较大区别。

本文的研究结果具有几个方面的意义和应用价值。首先,本文检验了股票多头策略的私募证券投资基金的风险因子有效性,明确股票多头基金收益的主要来源。其次,本文通过 Stepwise 和 LASSO 模型

**Table 8.** Significance statistics among all factors in explanatory power towards individual funds in three models ( $|t| > 2$ )  
**表 8.** 因子解释力差异显著性统计结果( $|t| > 2$ )

回归模型	截距			资产风格因子				情绪因子				宏观因子				流动性	
	Alpha	HS300	规模因子	封闭基金折价	上证换手率	新增开户数	消费者信心指数	融资融券余额	总体	信用利差	消费价格指数CPI	HS300股息率	广义货币M2	工业增加值	总体	流动性风险	
OLS	332	1193	848	299	344	346	299	508	198	147	164	123	399	728	220		
Stepwise	257	1199	835	259	282	309	282	451	174	127	124	109	336	633	188		
LASSO (subset)	475	1163	796	210	221	214	223	415	129	84	94	77	253	466	149		
平均	355	1185	826	256	282	290	268	458	167	119	127	103	329	609	186		
平均值占比	24%	81%	56%	17%	19%	20%	18%	31%	11%	8%	9%	7%	22%	41%	13%		

**Table 9.** Significance statistics among all factors in explanatory power towards individual funds in three models ( $|t| > 2$ )—robustness test  
**表 9.** 因子解释力差异显著性统计结果( $|t| > 2$ )——鲁棒性检验

回归模型	截距			资产风格因子				情绪因子				宏观因子				流动性	
	Alpha	HS300	规模因子	封闭基金折价	上证换手率	新增开户数	消费者信心指数	融资融券余额	总体	信用利差	消费价格指数CPI	HS300股息率	广义货币M2	工业增加值	总体	流动性风险	
OLS	335	1207	477	282	329	370	320	520	218	175	165	149	407	784	243		
Stepwise	250	1197	441	264	285	320	298	478	184	161	145	122	353	679	217		
LASSO (subset)	465	1156	348	180	186	200	202	405	135	85	91	60	234	439	148		
平均	350	1187	422	242	267	297	273	468	179	140	134	110	331	634	203		
平均值占比	24%	81%	29%	16%	18%	20%	19%	32%	12%	10%	9%	7%	23%	43%	14%		

尝试区别不同基金的风险因子,从而更有针对性地刻画不同私募基金经理的风险特征。再次,私募证券投资基金的历史相对较短而且很多都是月度数据,而解释因子的长期历史数据容易获取,可以用因子数据来模拟私募证券投资基金业绩来做更长期的分析。更进一步地,可以通过对因子的不同假设来对采用不同策略的私募证券投资基金业绩做情景分析。最后,本文提供了一个类似于 Fung 和 Hsieh (2004) [4] 研究框架,可以用于分析私募证券投资基金获取阿尔法收益的能力强弱,从而筛选出真正优秀或独特的基金。

总之,本文的研究框架及实证结果将有助于投资者根据对宏观和市场环境的判断来选择和配置私募证券投资基金。同时,监管机构也可以应用该研究框架来监控宏观环境及政策变化对私募证券投资基金整体的影响,并结合实证结果评估中国私募证券投资基金行业发展状况。

## 基金项目

中国博士后基金面上项目(2016M590351)、西交利物浦大学重点项目建设专项资金(KSF-A-02)。

## 参考文献

- [1] Sharpe, W.F. (1964) Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, **3**, 425-442.
- [2] Fung, W. and Hsieh, D.A. (1997) Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds. *Review of Financial Studies*, **10**, 275-302. <https://doi.org/10.1093/rfs/10.2.275>
- [3] Fung, W. and Hsieh, D.A. (2002) Asset-Based Style Factors for Hedge Funds. *Financial Analysts Journal*, **58**, 16-27. <https://doi.org/10.2469/faj.v58.n5.2465>
- [4] Fung, W. and Hsieh, D.A. (2004) Hedge Fund Benchmarks: A Risk-Based Approach. *Financial Analysts Journal*, **60**, 65-80. <https://doi.org/10.2469/faj.v60.n5.2657>
- [5] Pástor, L. and Stambaugh, R.F. (2003) Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*, **111**, 642-685. <https://doi.org/10.1086/374184>
- [6] Brandon, R.G. and Wang, S. (2013) Liquidity Risk, Return Predictability, and Hedge Funds' Performance: An Empirical Study. *Journal of Financial and Quantitative Analysis*, **48**, 219-244. <https://doi.org/10.1017/S0022109012000634>
- [7] Baker, M. and Wurgler, J. (2006) Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, **61**, 36. <https://doi.org/10.1111/j.1540-6261.2006.00885.x>
- [8] Baker, M.P., Wurgler, J. and Yuan, Y. (2011) Global, Local, and Contagious Investor Sentiment. Social Science Electronic Publishing.
- [9] 胡昌生, 池阳春. 投资者情绪:理性与非理性[J]. 金融评论, 2012(6): 46-62.
- [10] Chen, Y., Han, B. and Pan, J. (2015) Sentiment Risk and Hedge Fund Returns. Social Science Electronic Publishing.
- [11] Fama, E. and French, K. (1989) Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics*, **25**, 49-62. [https://doi.org/10.1016/0304-405X\(89\)90095-0](https://doi.org/10.1016/0304-405X(89)90095-0)
- [12] Keim, D.B. and Stambaugh, R. (1986) Predicting Returns in the Bond and Stock Markets. *Journal of Financial Economics*, **17**, 357-390. [https://doi.org/10.1016/0304-405X\(86\)90070-X](https://doi.org/10.1016/0304-405X(86)90070-X)
- [13] Campbell, J.Y. (1987) Stock Returns and the Term Structure. *Journal of Financial Economics*, **18**, 373-399. [https://doi.org/10.1016/0304-405X\(87\)90045-6](https://doi.org/10.1016/0304-405X(87)90045-6)
- [14] Bali, T.G., Brown, S.J. and OnurCaglayan, M. (2011) Do Hedge Funds' Exposures to Risk Factors Predict Their Future Returns? *Journal of Financial Economics*, **101**, 36-68. <https://doi.org/10.1016/j.jfineco.2011.02.008>
- [15] Avramov, D., Kosowski, R., Naik, N.Y. and Teo, M. (2011) Hedge Funds, Managerial Skill, and Macroeconomic Variables. *Journal of Financial Economics*, **99**, 672-692. <https://doi.org/10.1016/j.jfineco.2010.10.003>
- [16] Avramov, D., Barras, L. and Kosowski, R. (2013) Hedge Fund Return Predictability under the Magnifying Glass. *Journal of Financial and Quantitative Analysis*, **48**, 1057-1083. <https://doi.org/10.1017/S0022109013000422>
- [17] 沈维涛, 黄兴李. 我国证券投资基金业绩的实证研究与评价[J]. 经济研究, 2001(9): 22-30.
- [18] 李红权, 马超群. 中国证券投资基金绩效评价的理论及实证研究[J]. 财经研究, 2004, 30(7): 56-65.
- [19] 屠新曙, 朱梦. 基金绩效评价的 Fama-French 三因素模型检验[J]. 金融经济研究, 2010, 25(1): 103-112.

- 
- [20] 郑琳. 中国证券投资基金绩效归因分析的实证研究[D]: [硕士学位论文]. 上海: 上海交通大学, 2013.
- [21] Tibshirani, R. (1996) Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society, Series B (Methodological)*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [22] 孙华好, 马跃. 中国货币政策与股票市场的关系[J]. 经济研究, 2003, 7(2): 44-53.
- [23] 梁琪, 滕建州. 股票市场, 银行与经济增长: 中国的实证分析[J]. 金融研究, 2005(10).

## 附录 A: LASSO 回归模型

LASSO 回归模型通过在 OLS 算法中最小化残值的计算中加入一个 L1 范数作为惩罚约束,一定程度上应对解释变量的多重共线性和过拟合问题。其拉格朗日形式为:

$$\min \left\{ \frac{1}{N} \|y - X\beta\|_2^2 + \lambda \|\beta\|_2 \right\}$$

其中  $N$  为观察数量,  $y$  为被解释变量向量,  $\beta$  为系数向量,  $\lambda$  为变参向量,  $X$  为解释变量矩阵。本文中的被解释变量为私募证券投资基金指数以及个体样本的月度收益,解释变量集为表 2 所列因子。LASSO 计算得到满足公式(A1)的最优  $\lambda$ , 解释变量子集以及系数向量。

应用 LASSO 和交叉验证对每个基金样本分析会分别得到一组最优模型参数,包括最优  $\lambda$  和解释变量数量,同时也得到对应的回归系数和解释度。下面以股票多头产品其中一个的业绩归因检验为例说明。首先,附录图 1 以信用利差、沪深(HS) 300 收益率,融资融券余额和流动性风险四个解释变量为例,显示了在不同  $\lambda$  值下这四个解释变量的解释力差异。图中每条不同颜色的线都是一个解释变量,从右到左与横轴相交,越早相交的变量解释力越强。由附录图 1 可见,随着  $\lambda$  值增加,在这个四个变量中仅沪深 300 收益率有一定解释能力,其他变量的解释能力则沿不同路径收敛至零。其次,附录图 2 显示了在 LASSO 模型下利用股票多头产品月度收益对表 4 中所列解释因子进行回归时,不同  $\lambda$  所对应的交叉验证后的 MSE 的范围。可以看到随着遍历  $\lambda$ , 在  $\lambda$  对应的模型选择参数数量为 1 时(附录图 2 顶部横轴所示),其平均平方误差(Mean Squared Error, MSE)最小,此时的 MSE = 0.004,且即使其交叉验证后的最大值也远远小于总均方差 0.032,此时的解释度值为 0.87,故判定模型有效。附录图 2 中被选择解释变量即为沪深 300 全收益回报率。同时,从附录图 2 曲线上可以看到如果增加其他的变量或者移除这一变量,都会使得优化对象的值上升。综合上述观察,在对股票多头策略指数进行业绩归因检验时,将沪深 300 全收益回报率作为主要的解释因子是 LASSO 框架下的最优模型选择。

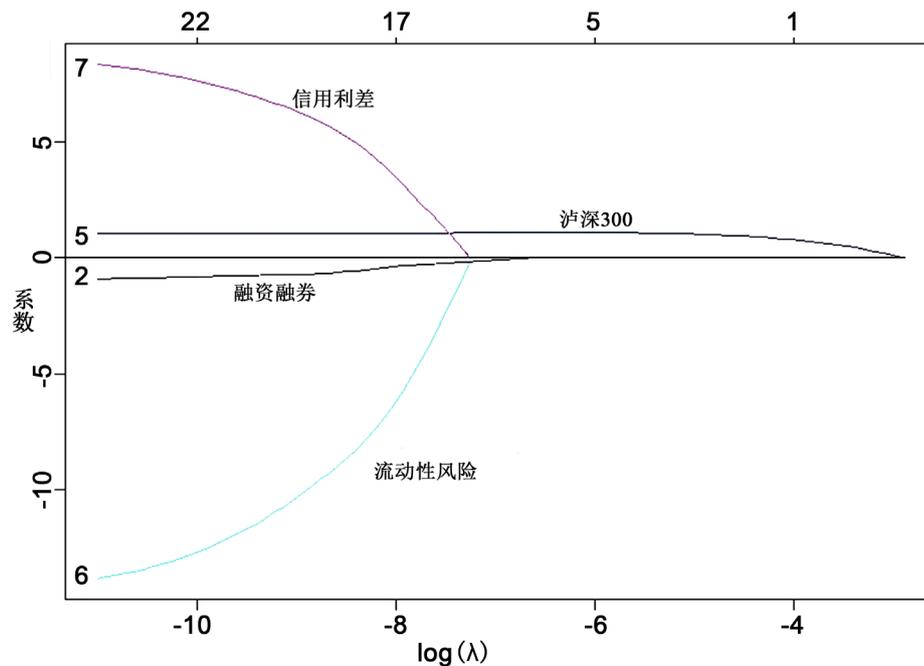
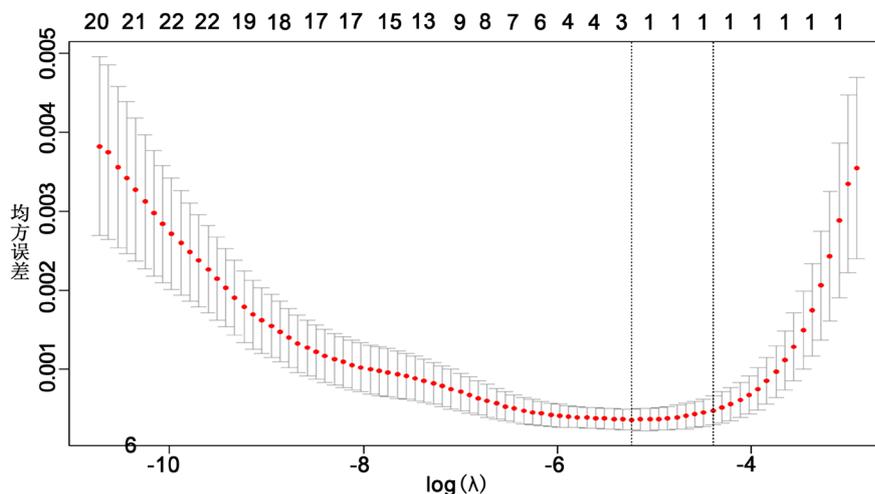


Figure A1. Selection of explanatory variables with different  $\lambda$  values

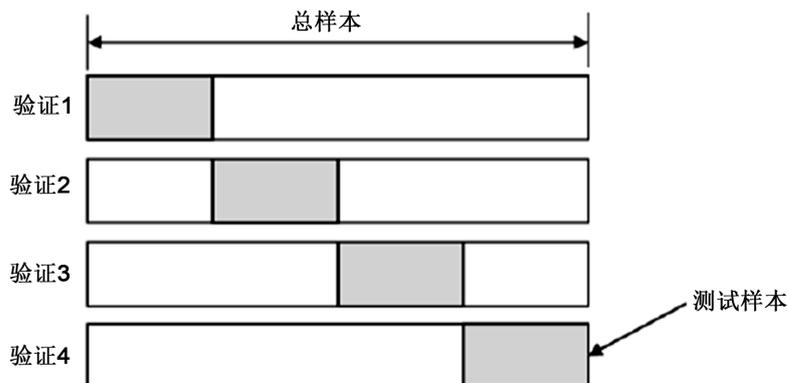
附录图 1. 不同  $\lambda$  下的解释变量选取



**Figure A2.** MSE with LASSO model with different  $\lambda$  values  
**附录图 2.** 不同  $\lambda$  下的 LASSO 模型均方误差

### LASSO 回归交叉验证

为了增加模型的可靠性以及确认参数，需要对 LASSO 进行交叉验证。本文中的交叉验证是通过对 LASSO 计算中不同  $\lambda$  所确定的模型进行交叉验证。如下附录图 3 所示，每一轮交叉验证都将样本历史数据分为  $K$  段(本文中  $K=4$ )，做  $K$  次样本内和样本外测试，每次测试选取其中 1 段保留作为测试样本(即样本外检验)，其他  $K-1$  段作为样本内数据确定模型参数。这样对 LASSO 计算的每个  $\lambda$  所确定的模型都有  $K$  个误差值，最后选取产生最小平均误差的  $\lambda$  值所对应的模型，并随后确定参数。原则上  $k$  取值越大，检验次数越多则可靠性越好，但同时考虑到本文所检验的私募证券投资基金指数的数据历史最小值为 36 个月，因此本文中  $k$  取值为 4。



**Figure A3.** Demonstration of cross-validation test in LASSO model  
**附录图 3.** 交叉验证基本思路

### 模型解释度

由于本文的研究假设不要求误差项满足正态分布， $e \sim N(0, \sigma^2 I)$ ，故参数不符合任何统计分布，无法以是否统计上显著来衡量参数的优劣。本文评定模型的优劣采用回归平方和(总平方和与残差平方和之差)与总平方和之比，即模型能够解释的被解释变量的变化部分，将其定义为解释度(也称为可决系数)。该值越接近与 1 则表示解释度越高。

**知网检索的两种方式：**

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-0967，即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：[fin@hanspub.org](mailto:fin@hanspub.org)