

低分散化是信息优势还是心理偏差因素^①

——基于信息配置的理论解释与实证分析

陈 炜¹, 袁子甲²

(1. 厦门大学中国资本市场研究中心, 厦门 361005; 2. 国信证券(香港)资产管理, 中国香港 999077)

摘要: 采用熵(entropy)度量以随机变量代表的新信息所包含的信息量,建立了基于主动学习的投资组合选择模型。模型刻画了投资者的信息配置影响最优组合选择的机理,表明在投资者主动进行信息配置条件下,信息优势和心理偏差都会使投资组合偏离传统投资组合理论描述的充分分散组合,导致低分散化现象。在理论解释基础上,以2002年至2007年所有基金在深市的数据检验低分散化现象的主导因素。结果一方面证实了本文的理论模型,即基金存在低分散化现象,另一方面证实了低分散化基金的业绩显著低于高分散化基金的业绩,表明中国股市基金低分散化现象源于信息优势多于心理偏差因素^②。

关键词: 低分散化; 贝叶斯学习; 投资组合选择; 心理偏差; 信息不对称

中图分类号: F830.9 **文献标识码:** A **文章编号:** 1007-9807(2015)10-0069-11

0 引 言

低分散化投资组合是指投资者通常只投资于少数几个金融资产,而不是按照标准投资组合理论的建议进行充分分散化的投资。Goetzmann 和 Kumar^[1], Polkovnichenko^[2]等实证研究都显示投资者普遍持有低分散化的投资组合, Barber 和 Odean^[3]也发现美国个人投资者平均持有 4.3 只股票。根据传统投资组合理论,分散化程度不足的投资组合中包含的非系统性风险没有风险补偿,因此理性投资者的组合应该是充分分散的。

至今为止,金融经济理论界对低分散化之谜尚无定论,目前还是学术界研究热点之一。另一方面,“分散投资还是集中投资”这个问题也一直是实务界争论的焦点问题之一,监管部门和学术

界推崇的所谓充分分散投资方法受到以著名投资家巴菲特为代表业内专家的强烈反对,巴菲特形象地将自身的投资要诀概括为“百鸟在林,不如一鸟在手”。关于低分散化现象,目前学术界主要有四种解释。第一种投资水平假说认为低分散化是因为投资者的投资水平低和不理解分散化组合的优势。第二种信息因素假说认为低分散化是因为投资者没有足够资源和能力去收集信息,或者投资者过度信赖私有信息质量或者缺乏能力去解释所获信息^[3,5]。第三种心理偏差假说认为低分散化是投资者存在强烈的心理偏差所致^[1],例如投资者过度自信、趋势交易行为、喜欢偏斜度较高的股票^[4]。同时,损失厌恶(loss aversion)、短视(narrow framing)、模糊厌恶(ambiguity aversion)、忠诚(loyalty)、过度乐观(optimism)、幼稚式分散

① 收稿日期: 2013-01-10; 修订日期: 2013-05-28.

基金项目: 国家自然科学基金资助项目(71072007).

作者简介: 陈 炜(1973—),男,浙江永康人,博士,研究员. Email: dchenwei@sina.com

② 本文仅代表作者个人学术观点,不构成投资建议,也不代表作者所在机构的观点.

化 (naive diversification)^③ 都可能导致低分散化现象。过度自信的投资者高估了其更新信息的能力,从而过高估计所收集信息的精度,从而导致过度交易^[5,6]。发现过度自信和感觉寻求 (sensation-seeking) 的投资者交易更加频繁。追求风险 (risk-seeking) 理论认为如果股票正偏 (positively skewed) 投资者可能为了获得高收益冒险选择低分散化组合^[2,4]。第四种假说是私有信息假说,与心理偏差解释不同,其解释是投资者试图利用优势私有信息。投资者如果拥有相关信息,将使得他们更加频繁交易。如果投资者高度风险厌恶,其也存在动力尽快利用私有信息^[7,8]。

现实中投资者选择低分散化组合 (持有集中组合) 的背后机制如何,无论从理论还是实践角度均值得研究。精明老练的投资者善于利用优势信息,并且具有低风险厌恶和较高的耐心,因而选择低分散化组合,笨拙的投资者选择低分散化组合,则可能是被心理偏差左右^[9]。国内学术研究中,许多文献也从行为金融角度探讨了心理偏差对资产定价及对投资收益的影响^[10,11,12],但是还缺少深入分析低分散化现象的文章。

此外,金融学领域中关于信息收集的研究始自 Grossman 和 Stiglitz^[13]、Verrecchia^[14],近期的研究包括 Peress^[15]、Peng^[16]、Huang 和 Liu^[17]等。这些研究均假设投资者只可以对一个风险资产进行投资,因此无法分析投资者可能面临的信息配置问题,即如果投资者只购买有限的信息量,那么应如何在不同的资产之间进行分配。

本文的理论模型主要基于信息优势与心理偏差两种因素建立动态学习模型分析信息配置如何影响投资组合选择,从而解释低分散化现象形成的深层次机理,这也是本文的主要创新之处。相比已有研究,本文基于动态学习的模型的假设更为宽松,即假定存在 n 种风险资产与一种无风险资产,并采用熵度量以随机变量代表的新信息所包含的信息量。同时,本文利用基金在深市的数据证实了中国股市基金这种机构投资者的低分散化现象,并检验了信息配置因素与心理因素

哪个更可能是我国股市基金低分散化的主导因素。

1 基于信息配置和心理偏差的动态组合选择模型

贝叶斯学习的投资组合选择模型可以分为被动学习和主动学习两种类型。传统的被动学习 (passive learning) 模型中,投资者通过观察价格或者收益率过程对信念进行更新,从而无需对所观察的信息进行选择。主动学习则是指有选择地收集信息以及观察到信息后进行的信念更新。

本文建模针对投资者的主动学习行为 (active learning),同时考虑收集信息需要支付成本,构建了基于主动学习的动态投资组合模型。该模型可以用来分析投资者如何选择与配置信息:一是如何确定投资者的最优信息收集量;二是投资者如何在多种资产中选择何种信息 (与何种资产相关) 进行收集和学习;三是信息配置和心理偏差如何影响投资者的投资组合选择。值得说明的是,本文主要针对股票市场进行建模和实证,但是由于本文只需资产收益率分布遵从正态分布这一假设,并不需要假设市场有效,因此本文理论模型应用不限于股票市场本身。

1.1 基本模型设置

本文建立一个三阶段模型。在时刻 1 投资者选择信息的精度,精度越高的信息相应需要支付更高的成本。在时刻 2 投资者观察到一个信号并构建最优投资组合,在时刻 3 投资者获得支付并实现效用。

1.1.1 信息与投资者的学习过程

假设市场中存在一种无风险资产和 n 种风险资产。投资者关于风险资产的支付 z 具有如下先验分布:均值向量为 μ 、方差协方差矩阵为 Σ 的正态分布。在时刻 1 投资者会付出精力或者财富去收集与 z 相关的信息,比如购买投资研究报告、

③ 美国市场发现许多个人在选择投资组合采用简单的 $1/n$ 策略,即将投资平均分配在 n 种股票。

阅读财经报纸、浏览财经网页或者雇用理财顾问等。在时刻 2 投资者将观察到一个关于风险资产支付的信号

$$s = z + \delta \quad (1)$$

其中 δ 与 z 独立, 并且 $\delta \sim (0, \Sigma_\delta)$ 。

投资者通过贝叶斯法则 (Bayesian rule) 来更新其信念。收益率关于信息的条件分布为

$$z | s \sim (\hat{\mu}, \hat{\Sigma}) \quad (2)$$

其中 $\hat{\mu} = E[z | s] = (\Sigma^{-1} + \Sigma_\delta^{-1})^{-1} \times$

$$(\Sigma^{-1}\mu + \Sigma_\delta^{-1}s) \quad (3)$$

$$\hat{\Sigma} = \text{Var}[z | s] = (\Sigma^{-1} + \Sigma_\delta^{-1})^{-1} \quad (4)$$

投资者将根据对信息学习之后得到的条件概率分布来进行投资组合选择。如果信号 s 所包含的信息量较大或者精度较高, 那么投资者对 z 的预测也更为准确, 进而可以提升投资效果。信息理论中熵的概念可以用来度量一个新信息所包含的信息量^[18]。熵可以用于度量随机变量的不确定性程度, 由服从正态分布随机变量的熵的计算方法可以得出 z , $z | s$ 的熵分别为: $H(z) = \ln[(2\pi e)^n |\Sigma|]$; $H(z | s) = \ln[(2\pi e)^n |\hat{\Sigma}|]$ 。

那么相对于投资者原有的信息, 信号 s 所包含的信息量可以定义为观察到信号 s 之后使得 z 的熵所下降的量

$$I(z | s) = H(z) - H(z | s) = \ln\left(\frac{|\Sigma|}{|\hat{\Sigma}|}\right) \quad (5)$$

因此 s 所包含的信息量取决于 $k = |\Sigma|/|\hat{\Sigma}|$ 。 $|\hat{\Sigma}|$ 越小 (k 越大) 意味着 s 所包含的信息量越大, 也就是信号 s 的精度越高。 $|\hat{\Sigma}|$ 最大可以等于 $|\Sigma|$ 。在这种情形下 s 没有包含任何新信息, 因此 k 的取值范围为 $[1, +\infty)$ 。投资者购买信号 s 需要支付成本 $C(k)$, 假设成本函数 $C(k)$ 满足如下条件: $C(1) = 0$, $C'(k) > 0$, $C''(k) > 0$ 。这一假设表明没有新增信息时成本为 0, 信息成本随着信息量的增大而增大, 并且增加新信息的边际成本也随着信息量的增大而不断增大。举例来看, $C(k) = (k - 1)^b$, $b > 1$ 便是满足上述条件的函数形式。

行为金融学认为投资者在更新信息的过程中存在各种类型的心理偏差。本文的模型加入最典型的过度自信。过度自信 (overconfidence) 是指投资者高估了其更新信息的能力, 从而过高估计所收集信息的精度。过度自信已被用于解释众多金融现象并取得了满意的效果, 包括市场过度波动^[19]、股价长期反转^[20] 和过度交易^[5] 等。

本文采用 Peng 和 Xiong^[21] 中关于过度自信的定义。假设投资者支付信息成本 $C(k)$ 购买信号 s , 但主观认为信号 s 所包含的信息量不是 k , 而是 γk , 其中 $\gamma \geq 1$ 。此时投资者过高地估计了信号 s 的精度。参数 γ 可以度量投资者过度自信的程度。 $\gamma = 1$ 对应的是理性投资者的情形; 而当 $\gamma > 1$ 时, 投资者对其处理信息和学习的能力估计过高。

1.1.2 价值函数构造

在以上模型设置之下, 存在心理偏差的投资者面临如何信息配置, 并构造投资组合。投资决策问题应分为两个阶段: 信息配置和投资组合选择。在时刻 2 投资者已经观察到信号 s 和风险资产的价格向量 p 。假设 α 代表投资者所构建的投资组合, 其第 i 个分量 α_i 表示投资于第 i 个资产的份数, r_f 为无风险资产的回报。使用均值-方差模型的投资者将面临如下投资组合选择问题

$$\max_{\alpha} E[W | s] - \frac{\rho}{2} \text{Var}[W | s] \text{ s. t. } W = W_0 r_f + \alpha'(z - r_f p) - C(k) \quad (6)$$

其中 ρ 表示投资者的风险规避程度。上述优化问题中的约束为投资者的预算约束, 此时信息成本 $C(k)$ 是给定的。记 α_s^* 为投资者选择的最优投资组合, $v(s)$ 表示上述投资组合选择问题的价值函数

$$v(s) = E[W_0 r_f + \alpha_s^*(z - r_f p) | s] - \frac{\rho}{2} \text{Var}[W_0 r_f + \alpha_s^*(z - r_f p) | s] - C(K) \quad (7)$$

在时刻 1 投资者首先确定所要购买信号的精度, 也就是信息量 k 。然后, 通过选择信号的方差协方差矩阵 Σ_δ 来确定信息在不同资产之间的配置, 从而使价值函数 $v(s)$ 的期望达到最大。 Σ_δ 与 $\hat{\Sigma}$ 是一

一对应的,所以选择最优的 Σ_s 等价于选择最优的 $\hat{\Sigma}$. 这里将直接对 $\hat{\Sigma}$ 进行优化. 因此在时刻 1 过度自信投资者面临的信息配置问题为

$$\max_{\kappa, \hat{\Sigma}} E_s [v(s)]; \text{ s. t. } \frac{|\hat{\Sigma}|}{|\Sigma|} \leq \gamma K; \hat{\sigma}_i^2 \leq \sigma_i^2, i = 1, \dots, n \quad (8)$$

其中 σ_i^2 和 $\hat{\sigma}_i^2$ 分别为 Σ 和 $\hat{\Sigma}$ 第 i 个对角上的元素, 上述问题中的第一个约束表示信号所含的信息量不能超过投资者主观认为其拥有的信息量 γk , 第二个约束表示不允许出现投资者购买信号之后反而对第 i 资产的了解程度下降的情形.

1.1.3 最优信息配置和投资组合选择

在时刻 2 给定信号 s 和 $\hat{\Sigma}$ 投资者可以根据式 (3) 和 (4) 计算出 $\hat{\mu}$. 标准均值 方差模型给出了 α_s^* 的表达式为

$$\alpha_s^* = \frac{1}{\rho} \hat{\Sigma}^{-1} (\hat{\mu} - r_f \rho) \quad (9)$$

下面求解信息配置问题 将 α_s^* 的表达式代入式 (7) 得到

$$v(s) = \frac{1}{2\rho} (\hat{\mu} - r_f \rho) \hat{\Sigma}^{-1} (\hat{\mu} - r_f \rho) - C(K) + W_0 r_f \quad (10)$$

这里只考虑不同风险资产的收益率相互之间独立并且与不同风险资产相关的信号相互之间也是独立的情形^④. Σ 和 Σ_s 在这一假设下均为对角矩阵, 从而 $\hat{\Sigma}$ 也是对角矩阵. 这样式 (10) 可以写成

$$v(s) = r_f + \frac{1}{2\rho} \sum_i \frac{(\hat{\mu}_i - r_f \rho_i)^2}{\hat{\sigma}_i^2} - C(K) + W_0 r_f \quad (11)$$

由于 $E_s [\hat{\mu}_i] = \mu_i, \text{Var}_s [\hat{\mu}_i] = \sigma_i^2 - \hat{\sigma}_i^2$, 并且式 (11) 中只有 $\hat{\mu}_i$ 是随机项, 因此信息配置问题的目标函数可以写成

$$E_s [v(s)] = r_f + \frac{1}{2\rho} \sum_{i=1}^n \frac{\sigma_i^2 - \hat{\sigma}_i^2 + (\mu_i - r_f \rho_i)^2}{\hat{\sigma}_i^2} - C(K) + W_0 r_f = r_f - \frac{N}{2\rho} + \frac{1}{2\rho} \sum_{i=1}^n \frac{\sigma_i^2}{\hat{\sigma}_i^2} \left[1 + \frac{(\mu_i - r_f \rho_i)^2}{\sigma_i^2} \right] -$$

$$C(K) + W_0 r_f \quad (12)$$

这样信息配置问题可以转化成

$$\max_{\{\hat{\sigma}_i^2\}} \frac{1}{2\rho} \sum_{i=1}^n \frac{\sigma_i^2}{\hat{\sigma}_i^2} \times \left[1 + \frac{(\mu_i - r_f \rho_i)^2}{\sigma_i^2} \right] - C(K) \text{ s. t. } \prod_{i=1}^n \frac{\sigma_i^2}{\hat{\sigma}_i^2} \leq \gamma K; \hat{\sigma}_i^2 \leq \sigma_i^2, i = 1, \dots, n.$$

从上述最优化问题可以看出, 信息配置问题在独立假设下已经转变成为一个精度配置问题. 假设 $C(K) = (K - 1)^b$, 求解上述问题会得到一个角点解 (corner solution) 并得到定理 1.

定理 1 假设满足以下条件:

- 1) 市场中存在一种无风险资产和 n 种风险资产. 投资者关于风险资产的支付 z 先验分布为均值为 μ 、方差协方差矩阵为 Σ 的正态分布;
- 2) 投资者选择信息的精度, 精度越高的信息相应需要支付更高的成本;
- 3) 投资者观察到信号 s 并进行学习之后, 根据条件概率分布构建最优投资组合;
- 4) 投资者购买信号 s 所需支付成本的函数 $C(k)$ 满足: $C(1) = 0, C'(k) > 0, C''(k) > 0$;
- 5) 投资者通过均值方差方法构建最优投资组合: $\max_{\alpha} E[W | s] - \frac{\rho}{2} \text{Var}[W | s]; \text{ s. t. } W = W_0 r_f + \alpha(z - r_f \rho) - C(k)$; 其中 ρ 表示投资者的风险规避程度.

那么: 令 $i^* = \arg \max_i \left\{ \frac{(\mu_i - r_f \rho_i)^2}{\hat{\sigma}_i^2} \right\}$, 投资者的最优信息配置为

$$K^* = 1 + \left[\frac{\gamma}{2b\rho} \left(1 + \frac{(\mu_i - r_f \rho_i)^2}{\sigma_i^2} \right) \right]^{\frac{1}{b-1}} \hat{\sigma}_i^{*2} = \frac{\alpha_i^2}{\gamma K^*} \hat{\sigma}_i^2 = \sigma_i^2, i \neq i^*$$

定理 1 可以得出以下几点结论:

第一, 投资者将所购买的信息全部用于降低夏普比例最高的那个资产的不确定性程度. 这一结论反映了投资组合选择对信息配置的影响, 由于投资者预期在夏普比例最高的风险资产上的投资比例较高, 因此将所收集的信息全部分配到该资产上可以获得最高的边际效用.

④ 不同资产收益率和信号均相互独立的假设主要是为了求解的便利.

第二,行为金融中典型的心理偏差(过度自信)会导致投资者增加对信息的投入($\frac{\partial K^*}{\partial \gamma} > 0$).这与 Guiso 和 Jappelli^[22]的实证研究相吻合.过度自信投资者由于高估了信息的精度从而高估了信息所带来的投资效果的提高,因此会加大对信息的投入.

第三,风险规避程度 ρ 的增大会导致对信息投入的减小($\frac{\partial K^*}{\partial \rho} < 0$).投资者的风险规避程度越高,在风险资产上的投资越少,从而信息投入所带来的边际效用下降.因此风险规避程度较高的投资者对信息的投入较少.

1.2 模型对投资组合低分散化现象的解释

以上模型表明信息配置和过度自信如何导致低分散化现象.首先考虑没有引入贝叶斯学习行为的情形,投资者将根据先验分布选择如下投资组合

$$\bar{\alpha} = \frac{1}{\rho} \Sigma^{-1} (\mu - r_f p) \quad (13)$$

这里将 $\bar{\alpha}$ 作为基准投资组合.上文已经求解了考虑贝叶斯学习时投资者所面临的信息配置和投资组合选择问题,并给出了考虑信息收集和过度自信时的最优投资组合

$$\alpha_s^* = \frac{1}{\rho} \hat{\Sigma}^{-1} (\hat{\mu} - r_f p) \quad (14)$$

$\hat{\mu}$ 在时刻 1 是一个随机向量.为了分析信息收集和过度自信等因素对投资组合的影响,这里应使用最优投资组合的期望值 $E_s[\alpha_s^*]$ 与基准投资组合 $\bar{\alpha}$ 进行比较.由于 $E[\hat{\mu}] = \mu$,从而学习所带来的投资组合的变化可以表示为

$$\begin{aligned} \Delta \alpha &= E_s[\alpha_s^*] - \bar{\alpha} \\ &= \frac{1}{\rho} (\hat{\Sigma}^{-1} - \Sigma^{-1}) (\mu - r_f p) \end{aligned} \quad (15)$$

$\Delta \alpha$ 的第 i 个分量可以具体写成

$$\Delta \alpha_i = \begin{cases} (\gamma K^* - 1) \bar{\alpha}_i & i = i^* \\ 0 & i \neq i^* \end{cases} \quad (16)$$

从上式可以看出,投资者对先验分布下夏普比例最高的资产 i^* 所投资的份数将有所增加(因为 $\gamma K^* > 1$),而在其他风险资产上所投资的份数保持不变,这将导致投资组合的分散化程度下降.

现在可以分析过度自信程度 γ 和风险规避程度 ρ 对投资组合的影响.由于 $\frac{\partial K^*}{\partial \gamma} > 0$,因此

$\frac{\partial \Delta \alpha_i^*}{\partial \gamma} > 0$.这说明过度自信程度越高的投资者,对资产 i^* 的份数更大,与此同时对其他风险资产投资的份数保持不变,从而导致投资组合的分散化程度降低.事实上,随着投资者过度自信程度的不断增大:一方面,投资者会投入更多信息成本去购买包含信息量更大的信号,另一方面,投资者会高估信号所包含的信息量.而全部新增信息仍然被用于降低资产 i^* 的不确定性,因此在投资者的后验信念中,资产 i^* 的不确定性 $\hat{\alpha}_i^2$ 随着过度自信程度 γ 的上升不断下降,导致投资者提高对资产 i 投资的份数,从而使得投资组合更加集中.

此外,风险规避程度越高的投资者将选择分散化程度较高的投资组合,这是因为 $\frac{\gamma K^* (\gamma)}{\partial \rho} >$

0 , 所以有 $\frac{\partial \Delta \alpha_i^*}{\partial \rho} < 0$, 并且当 $\rho \rightarrow +\infty$ 时, $E_s[\alpha_s^*] \rightarrow$

$\bar{\alpha}$ 这一结论同样与 Goetzmann 和 Kumar^[1]的实证检验结果相吻合的.但是 Goetzmann 和 Kumar^[1]指出,风险规避程度较高的投资者选择分散化投资组合是由于他们在金融投资活动中的表现更为成熟,了解对风险进行分散的重要性.而在这里,本文给出了风险规避程度影响投资组合分散化程度的另一个传导机制:风险规避程度增大会导致投资者减少对风险资产的投资,此时投资者对信息的投入所带来的边际回报也随之下降,因此会减少对信息的收集.

值得指出的是,从式(16)看出,即使不考虑过度自信($\gamma = 1$),由于 $K^* > 1$,本文模型仍然可以得出投资者会选择低分散化投资组合的结论.这说明考虑投资者的主动学习行为时,低分散投资组合实际上是投资者的理性选择.

2 中国股市基金低分散化现象动因的实证检验

本文模型证明了信息优势和心理偏差将导致出现低分散化现象,但是两种因素的差异在于信

息优势带来的低分散化能够提高投资业绩,而心理偏差带来的低分散化会降低投资业绩.因此,本文以下进一步利用这种差别来建立实证模型检验中国市场低分散化的主要动因是信息优势因素还是心理偏差因素.具体而言,根据理论假设和模型,如果投资者选择低分散化的投资组合是因为信息优势,那么由于投资者会利用优势信息地位选择优先配置夏普比例较高的品种,应该那些持有低分散化组合投资者的业绩应该优于持有高分散化组合投资者的业绩.另一方面,如果持有低分散化组合的投资者是因为心理偏差选择低分散化组合,其业绩应该低于那些持有高分散化组合投资者的业绩.因此,可以通过检验分散化程度与投资业绩之间的关系来研究在中国市场,到底低分散化现象的主要原因是心理偏差还是信息优势因素.

中国股市中,基金是最典型的一类机构投资者,其流通市值规模和专业投资性质决定了基金是二级市场重要价格影响者,也是市值占比最高的专业投资者^⑤.以深市为例,2009年至2011年基金流通市值占比分别为19.19%、15.93%和13.02%,交易金额占比分别为7.04%、6.11%和6.38%.为了控制其他因素的影响,实证模型还选择了投资规模、组合股票平均规模、组合股票平均波动率、组合贝塔值、组合波动率、组合偏斜度这6个控制变量^⑥.在控制变量约束下,考察组合分散化程度与基金业绩之间的关系,从而验证基金的低分散化现象是否主要源于信息优势还是心理偏差因素.

2.1 实证检验模型

为了检验基金组合分散化程度与投资业绩之间的关系,进而分析低分散化现象的主要原因是心理偏差还是信息优势因素,本文建立以下多元线性回归模型,估计方法采用最小二乘法.

$$Performance_i = \beta_0 + \beta_1 Puidiv_i + \sum_{j=1}^n \beta_j Controlfactors_{ji} + \varepsilon_i \quad (17)$$

其中 $Performance_i$ 为投资者投资业绩衡量指标; $Puidiv$ 为检验变量^⑦, $Controlfactors$ 为控制变量.如果低分散化程度 $Puidiv$ 系数显著 β_1 不等于零,说明低分散化程度对投资业绩的影响显著.

2.2 数据来源和变量计算方法

本文采用2002年至2007年所有基金在深市股票上的相关数据进行实证研究,数据得到深圳证券交易所支持,所有数据用SAS9.1进行处理.基金的业绩度量主要采用以下步骤:

(1) 对于每只基金,计算2002年至2007年期间每个交易日投资组合的收益率,形成日收益率序列.具体方法是用明细持股数据计算每天每只基金在每只股票持有头寸上的收益率,该基金的当天收益率等于经过价值加权的组合中所有股票的收益率.

(2) 计算业绩度量指标.先计算每只基金的日收益率序列,然后经过三因素模型进行风险调整后得到基金业绩度量指标 α .三因素因子日序列来自清华锐思数据库系统.其中,三因素模型的估计公式如下

$$R_{it} - R_{ft} = \alpha_{it} + b_i(R_{mt} - R_{ft}) + s_iSMB_t + h_iHML_t + e_{it} \quad (18)$$

其中 R_{it} 为资产 i 在时间 t 的收益率; R_{ft} 为时间 t 的无风险收益率; $R_{it} - R_{ft}$ 为组合的超额平均收益; R_{mt} 是时间 t 的市场收益率; SMB_t 为规模因子; HML_t 为帐面市值比因子; e_{it} 为误差项,服从标准正态分布.

低分散化程度以基金组合与市场组合权重的差异来衡量.计算公式为 $Puidiv = \sum_{i=1}^n |w_i - W_i| \mu_i$ 为持有 i 种股票占其投资组合的比重, W_i 是市场中 i 种股票占总市值的比重.该指标越接近0,

⑤ 基金在专业机构中市值占比最高.基金、保险机构、社保基金、QFII、券商自营和券商资产管理这六类专业机构之外的非专业机构占比近40%,包括了开户的一般经营性公司、IPO或股改解禁的机构股东等等.

⑥ 值得说明的是,投资者特征、行为与投资业绩关系方面已有的研究较少,已有研究主要讨论了投资业绩与经验^[23-24]、年龄或认知能力^[9]、持股集中度^[25]、交易活跃程度^[3].

⑦ 源自“Portfolio underdiversification”的缩写.

表明组合越分散,反之,指标值越大(越接近 1), Goetzmann 等^[1]等文献均采用类似的方法来度量投资组合的低分散化程度,表明组合越呈现低分散化,偏离市场组合越远,量投资组合的低分散化程度.

表 1 变量的定义与度量标准

Table 1 The definition and metrics of variables

变量	变量名称	符号	度量标准
因变量	投资业绩	<i>Performance</i>	经三因素模型调整
检验变量	低分散化程度	<i>Pudiv</i>	用偏离市场组合的程度度量
控制变量	投资规模	<i>Insize</i>	组合价值的对数值
	组合股票平均规模	<i>Shr_Size</i>	组合中股票加权流通市值的对数值
	组合股票平均波动率	<i>Bdl</i>	组合股票加权波动率
	组合贝塔值	<i>Beta</i>	组合贝塔值
	组合波动率	<i>Var</i>	组合收益率序列方差
	组合偏斜度	<i>Skew</i>	组合收益率序列偏斜度

2.3 实证结果

表 2 为组合低分散化程度的描述性统计. 市场组合的低分散化程度指标为 0,完全偏离市场组合的低分散化程度指标为 1. 可以发现,多数基金低分散化程度(持股集中度)较高,其平均值 0.87,大大偏离标准的市场组合(市场组合的低分散化程度为 0,低分散化程度指标越高,持股集中度越高). 组合低分散化程度指标的 20% 分位数为 0.83,表明 80% 的基金组合分散度指标高于 0.83. 所有基金平均持有股票数量仅仅为 21 只,远远低于市场交易股票数量. 因此中国股市中,正如本文模型所刻画,基金的投资组合低分散化程度较高. 从表 3 各组基金样本的业绩指标来看,总体基金经过三因素模型调整之后的平均业绩大于零,其中最高组的风险调整后的收益为 1.558%,最低组的风险调整后的收益为 0.031%,表明总体基金业绩优于市场,但是业绩较低的四组风险调整后业绩仅仅略高于市场组合.

从表 3 的两组数据可以看出:低分散化程度

越高(持股集中度高)的组,业绩越好. 按业绩分组则看到,业绩越高的组,其低分散化程度越高(持股集中度高). 例如,业绩最好的一组,其低分散化程度指标为 0.906,远远偏离完全分散组合相应指标(0),其组合平均仅有 13.93 只股票,低于其它各组的平均股票只数.

模型调整后 R 平方等于 26.8%,低分散程度指标估计系数为 -0.0071,在 10% 的水平显著,表明基金组合的低分散化程度与投资者业绩水平呈现正比,即低分散化程度越高(持股集中),基金投资业绩越高. 反之低分散化程度越低(持股分散),基金投资业绩越低. 该结果表明中国股市基金的低分散化现象更多来源于信息优势因素.

当然,正如理论模型中所证实的,信息优势和心理偏差都是低分散化现象的动因. 在中国市场这两种因素都影响着基金管理人选择低分散化组合. 实证结果表明,信息优势因素多于心理偏差因素,因而结果体现为低分散化的基金组合业绩低于高分散化基金组合的业绩.

表 2 变量描述性统计

Table 2 Collections of statistics of paper

年份	2002	2003	2004	2005	2006	2007	总计
基金数目	77	101	141	184	254	290	1 047
组合分散度平均值	0.90	0.93	0.90	0.89	0.87	0.83	0.87
组合股票平均数量	33.06	16.38	21.33	18.21	18.94	23.03	21.06

表3 基金组合低分散程度分组统计

Table 3 Statistics of variable in different group

按低分散化程度分组	组合业绩(%) (Performance)	平均股票数量 (Shr_Num)	组合业绩 分组	低分散化程度 指标(Pudiv)	平均股票数量 (Shr_Num)
最高组	0.031	45.54	最低组	0.871	21.26
第2组	0.028	19.75	第2组	0.858	27.77
第3组	0.044	16.36	第3组	0.857	23.19
第4组	0.041	14.80	第4组	0.881	18.28
最低组	1.558	8.02	最高组	0.906	13.93

注:表中为根据低分散化程度指标和业绩指标按基金样本数量等分成五组。

表4 模型估计结果

Table 4 Collections of hypotheses test result

变量	变量名称	符号	估计系数	T值
截距项	α	Intercept	0.19	13.03
检验变量	组合低分散化程度	Pudiv	-0.0071	-1.70*
控制变量	投资规模	Invsiz	-0.0004	-1.21
	组合股票平均规模	Shr Size	-0.0086	-12.45**
	组合股票平均波动率	Bdl	6.14	7.16**
	组合贝塔值	Beta	-0.0128	-5.34**
	组合波动率	Var	26.25	11.34**
	组合偏斜度	Skew	-0.0015	-3.36**

注:*表示在0.1水平上显著,**表示在0.01水平上显著。

3 基金信息优势和心理偏差的进一步分析

基金组合低分散化程度越高,投资业绩越好,表明中国股市基金的低分散化主要动因还在于信息优势。由于基金管理人利用优势信息地位选择优先配置夏普比例较高的品种,使得低分散化组合的业绩优于高分散化组合的业绩。

基金等机构投资者在股票市场上占据了信息资源的优势:(1)拥有专业的研究队伍和投资决策人员,可以及时快速地收集信息;(2)基金公司作为上市公司的重要股东,比一般投资者更有条件与资源从上市公司获得信息;(3)作为证券公司的重要客户,基金根据券商提供研究服务和信息的质量与准确性分配仓位佣金,这种利益关系使得基金往往能够从券商获得其它投资者无法得到的信息或者优先获得信息,从而占据了信息优势。另外一种不公平的信息优势在于内幕信息或利益输送,基金等机构投资者(也包括知情人士、

证券从业人员等)利用特殊渠道获得的优势信息进行投资获利,这已经是业内普遍知晓的事实:①基金公司通过与上市公司实际控制方(例如股东、董事、监事、高级管理人员以及实际控制人)紧密的信息互通甚至合谋以影响股价走势为自己或相关利益主体谋利;②基金公司通过利益输送等方式提高旗下明星基金的业绩;③基金等机构投资者等利用信息优势进行交易获取超额收益。以上的情况中,都会使得基金表现为低分散化程度越高(持股集中),业绩越好。

当然,这不意味着中国股市的基金不存在心里偏差因素,有证据表明,中国股市中基金存在典型的行为特征。中国股市的基金群体投资策略是典型的趋势(惯性)交易,倾向于买入前期收益率较高的股票,卖出前期收益率较低的股票。为了说明基金的趋势交易行为特征,这里采用2006年2010年间所有基金在深市的交易数据进行计算,每个交易日基金强烈买入(卖出)的股票构成强烈买入(卖出)组合(根据净买卖指数由低到高分成10组,最低组为强烈卖出组,最高组为强烈买

入组), 然后计算这些组合在交易前后若干个交易日的平均累计超额收益率^⑧。基金是典型的趋势交易者, 倾向于买入前期收益率较高的股票, 卖出前期收益率较低的股票。基金强烈买入之后, 股价呈现上涨趋势, 反之, 基金强烈卖出时候股价下跌。表 5 的结果表明, 基金强烈买入组合在过去 20 个交易日的涨幅高达 4.76%。强烈买入组合在过去 20、15、10 和 5 个交易日中涨幅均显著高于其强烈卖出组合, 买入卖出组合事件之前 5 个交易日的累计收益率高于卖出组合 2.77%, 前 20

个交易日的累计超额收益率则高出 4.6%。

因此, 基金具有趋势交易的行为特征。但是在中国股市, 由于信息高度不对称, 信息优势使得基金管理人通过优先配置夏普比例较高的品种, 从而提高组合集中度并提升投资者业绩。与此对照的是, 虽然基金管理人也存在心里偏差, 但是因为心里偏差降低投资业绩的效果低于信息优势提高投资业绩的效果, 最终结果是高分散化组合的业绩组合低于低分散化组合的业绩。

表 5 基金的买入卖出组合的累计超额收益率(%)

Table 5 Collections of mutual fund buy-sell portfolios' CAR(%)

组合	$k = -20$	$k = -15$	$k = -10$	$k = -5$	$k = 0$	$k = 5$	$k = 10$	$k = 15$	$k = 20$
强烈买入	4.76	4.33	3.66	2.62	1.31	2.01	2.62	3.13	3.48
强烈卖出	0.16	0.10	-0.04	-0.15	-0.32	-0.32	-0.28	-0.05	0.06
买入 卖出	4.60	4.22	3.70	2.77	1.63	2.33	2.90	3.18	3.42

注: k 表示交易天数, 买入卖出组合指根据该类投资者强烈买入组合和强烈卖出组合构造的零投资组合。

4 结束语

本文构建了基于信息配置与心理偏差的动态组合选择模型, 该理论模型表明信息优势和心理偏差因素都会导致投资者选择低分散化组合而提高组合中股票的集中度。首先, 由于投资者会将获得的优势信息投资于先验分布中夏普比例最高的那个资产, 从而导致该资产在最优投资组合中所占的比例增大。其次, 心理偏差(过度自信)会加大投资组合低分散化的程度。此外, 风险规避较小等因素也会导致低分散化。因此, 信息优势带来的低分散化能够提高投资业绩, 而心理偏差导致的低分散化会降低投资业绩。

在理论模型基础之上, 本文通过检验基金组合低分散化程度与投资业绩之间的关系来验证中国股市基金低分散化现象的主要动因是信息优势因素还是心理偏差因素。本文实证结果表明, 对于基金这样的专业投资者, 低分散化组合的业绩要优于高分散化组合的业绩, 这表明中国股市中低分散化现象的主要原因是投资者的心理偏差因素多于信息优势因素。本文的实证结论与中国股

市的现状是吻合的, 基金投资者利用其自身研究队伍、与上市公司和券商的密切联系, 比一般投资者拥有更多的信息优势, 这种信息优势往往带来组合的低分散化和业绩的提高。进一步而言, 尽管基金也存在导致组合低分散化的各种心理或心里偏差, 但其对业绩的负面作用要小于信息优势因素带来的业绩提升, 从而体现为低分散化程度与投资业绩的正相关关系。

本文的研究结论表明传统投资组合和资产定价理论并不能很好的解释投资实践中投资组合构建的低分散化现象, 而行为金融和信息不对称是两种重要的理论解释。本文理论和实证模型的结论与国外文献的结论是基本一致的, 例如 Kumar 和 Korniotis^[9] 等研究都表明信息优势与行为因素都会导致低分散化现象, 本文的创新之处在于从信息配置角度建立假设条件更加宽松的模型解释了两种因素导致低分散化的机理, 并进一步通过实证证实了中国股市中基金低分散化现象的主导因素是信息优势。

本文的实证结果也表明, 就中国股市基金这类机构投资者而言, 低分散化组合业绩优于高分散化组合, 集中持股和分散持股的孰优孰劣不仅

^⑧ 这里的超额收益率是指实际收益率减去市场收益率, 本文使用深市成份指数收益率作为市场收益率。

是一个学术问题,也是一个需要今后继续通过理论研究和实践进行检验的问题,其背后的原理还需要理论界和实务界进一步深入研究。此外,进

一步的研究还可以集中于中国股票市场投资者的各种心理或行为因素如何影响投资者组合构造,并解释其影响投资业绩的内在机理。

参考文献:

- [1]Goetzmann, William N, Kumar A. Equity portfolio diversification[J]. *Review of Finance*, 2008, 12(3): 433 – 463.
- [2]Polkovnichenko V. Household portfolio diversification: A case for rank-dependent preferences[J]. *Review of Financial Studies*, 2005, 18(4): 1467 – 1502.
- [3]Barber, Brad M, Odean T. Trading is hazardous to your wealth: The common stock investment performance of individual investors[J]. *Journal of Finance*, 2000, 55(2): 773 – 806.
- [4]Mitton T, Vorkink K. Equilibrium under-diversification and the preference for skewness[J]. *Review of Financial Studies*, 2007, 20(4): 1255 – 1288.
- [5]Odean T. Do investors trade too much? [J]. *American Economic Review*, 1999, 89(5): 1279 – 1298.
- [6]Grinblatt M, Keloharju M. Sensation seeking, overconfidence, and trading activity[J]. *Journal of Finance*, 2009, 64(2): 549 – 578.
- [7]Holden, Craig W, Subrahmanyam A. Long-lived private information and imperfect competition[J]. *Journal of Finance*, 1992, 47(1): 247 – 270.
- [8]Back K, C Henry C, Gregory A, et al. Imperfect competition among in-formed traders[J]. *Journal of Finance*, 2000, 55(5): 2117 – 2155.
- [9]Kumar A, Kormiotis G. Do portfolio distortions reflect superior information or psychological biases? [J]. *Journal of Financial and Quantitative Analysis*, 2013, 48(1): 1 – 45.
- [10]张 维,赵帅特. 认知偏差、异质期望与资产定价[J]. *管理科学学报*, 2010, 13(1): 52 – 59.
Zhang Wei, Zhao Shuaite. Cognitive biases, heterogeneous expectation and asset pricing[J]. *Journal of Management Sciences in China*, 2010, 13(1): 52 – 59. (in Chinese)
- [11]池丽旭,庄新田. 投资者的非理性行为偏差与止损策略——处置效应、参考价格角度的实证研究[J]. *管理科学学报*, 2011, 14(10): 54 – 66.
Chi Lixu, Zhuang Xintian. Investors' behavioural biases and stop-loss strategy: Empirical study based on disposition effect and reference point[J]. *Journal of Management Sciences in China*, 2011, 14(10): 54 – 66. (in Chinese)
- [12]张永杰,张 维,熊 熊. 投资策略与投资收益: 基于计算实验金融的研究[J]. *管理科学学报*, 2010, 13(9): 107 – 118.
Zhang Yongjie, Zhang Wei, Xiong Xiong. Strategies and investment returns: Agent-based computational finance perspective[J]. *Journal of Management Sciences in China*, 2010, 13(9): 107 – 118. (in Chinese)
- [13]Grossman S, Stiglitz J. On the impossibility of informationally efficient markets[J]. *American Economic Review*, 1980, 70(3): 393 – 408.
- [14]Verrecchia R. Information acquisition in a noisy rational expectations economy[J]. *Econometrica*, 1982, 50(6): 1415 – 1430.
- [15]Peress J. Wealth, information acquisition and portfolio choice[J]. *Review of Financial Studies*, 2004, 17(3): 879 – 914.
- [16]Peng L. Learning with information capacity constraints[J]. *Journal of Financial and Quantitative Analysis*, 2004, 40(2): 1749 – 1790.
- [17]Huang L, Liu H. Rational inattention and portfolio selection[J]. *Journal of Finance*, 2007, 62(4): 1999 – 2040.
- [18]Sims C A. The implications of rational inattention[J]. *Journal of Monetary Economics*, 2003, 50(3): 665 – 690.
- [19]Odean T. Volume volatility, price, and profit when all traders are above average[J]. *Journal of Finance*, 1998, 53(6): 1887 – 1934.

- [20] Daniel K, Hirshleifer D, Subrahmanyam A. Investor psychology and security market under- and overreactions [J]. *Journal of Finance*, 1998, 53(6): 1839 – 1885.
- [21] Peng L, Xiong W. Investor attention, overconfidence and category learning [J]. *Journal of Financial Economics*, 2006, 80(3): 563 – 602.
- [22] Guiso L, Jappelli T. Information Acquisition, Overconfidence and Portfolio Performance [R]. CEPR Working Paper # 5901, 2006.
- [23] Linnainmaa J. Do limit orders alter inferences about investor performance and behavior? [J]. *Journal of Finance*, 2010, 65(4): 1473 – 1506.
- [24] Nicolosi G, Peng L, Zhu N. Do individual investors learn from their trading experience? [J]. *Journal of Financial Markets*, 2009, 12(2): 317 – 336.
- [25] Ivković Z, Sialm C, Weisbenner S. Portfolio concentration and the performance of individual investors [J]. *Journal of Financial and Quantitative Analysis*, 2008, 43(3): 613 – 656.

Is underdiversification caused by superior information or psychological bias: The theoretical explanation and empirical test based on information allocation

CHEN Wei¹, YUAN Zi-jia²

1. The China Center of Capital Market Research, Xiamen University, Xiamen 361005, China;

2. Guosen Securities(HK) Asset Management Company Limited, Hong Kong 999077, China

Abstract: This paper examines the portfolio selection problem when investors should pay for information acquisition and have psychological bias. We propose a model to analysis the interaction between information choice and investment choice, and find that the investor's optimal information acquisition strategy is using all the information to learn about the asset with the highest Sharpe ratio, which induce an under-diversification portfolio. The model also shows the psychological bias will cause an under-diversification portfolio. We use all mutual fund trade data between 2002 and 2007 in Shenzhen Stock Exchange to examine our hypothesis. The empirical evidence supports the hypothesis of asymmetric information hypothesis. Under-diversification portfolio is caused by superior information of mutual funds in Chinese Stock Markets.

Key words: underdiversification; Bayesian learning; portfolio selection; psychological bias; asymmetric information