

## 《“以行定类”：基于持股偏好的投资者分类》附录

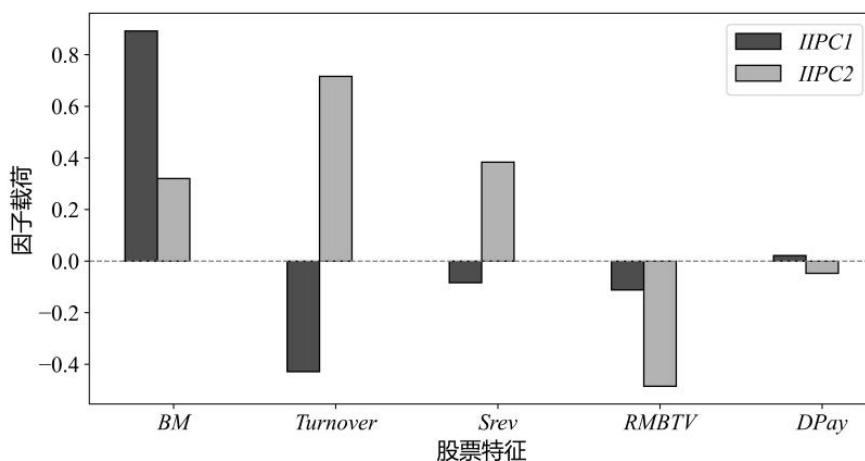
### 附录 1 基于基金持股偏好的分类结果

本文提出的分类方法同样适用于机构投资者数据。本附录关注的基金样本为投资于中国大陆市场的主动管理开放式股票型和偏股混合型基金，使用数据均来源于 CSMAR 数据库。本文对基金样本进行如下筛选：（1）基金类型必须为开放式普通股票型基金或偏股混合型基金；（2）剔除沪港通基金与投资海外的 QDII 基金；（3）剔除 ETF、指类型、非主动标识的基金；（4）剔除最近一次年报或半年报披露的中国市场股票持仓占总管理规模小于 50% 的基金。为了与个体投资者数据对齐，本文的基金样本研究同样选取 2016—2021 年的 2820 只基金进行分析，并主要使用其半年度披露的全部持仓数据。

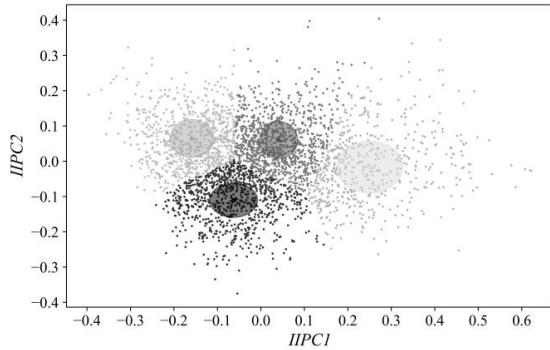
附表 1 汇报了基金投资者股票特征偏好的分歧度；附图 1 报告了经过主成分筛选后的基金两大偏好主成分；附图 2 报告了基金聚类在两大主成分下的结果；附图 3 报告了基金持仓在个体投资者集合下的分布。整体而言，基金持仓异质性小于个体投资者，主要以 *BM*（个体投资者财富加权中重要特征）为核心差异，且具有较强的行业属性。对比发现机构投资者持股偏好基本落入第三类个体投资者和部分第一类个体投资者中，说明该方法同样可以区分机构与个体样本。

附表 1 基金投资者的股票特征偏好分歧度

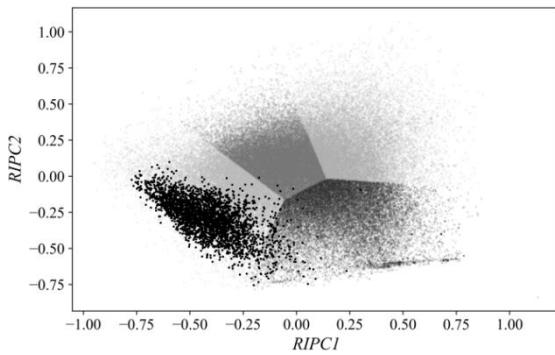
指标排名	等权		持仓金额加权		指标排名	等权		持仓金额加权	
	Mean	Std	Mean	Std		Mean	Std	Mean	Std
(1) <i>BM</i>	0.0272	0.0049	0.0244	0.0039	(11) <i>RMBTV</i>	0.0096	0.0033	0.0085	0.0037
(2) <i>Vol</i>	0.0189	0.0042	0.0174	0.0048	(12) <i>SOE</i>	0.0094	0.0009	0.0103	0.0012
(3) <i>Beta</i>	0.0172	0.0042	0.0158	0.0052	(13) <i>Illiq</i>	0.0094	0.0042	0.008	0.0044
(4) <i>Turnover</i>	0.0167	0.0025	0.0163	0.0029	(14) <i>ROE</i>	0.0092	0.0017	0.0089	0.0016
(5) <i>EP</i>	0.0137	0.0017	0.0124	0.0019	(15) <i>ME</i>	0.0076	0.0045	0.0066	0.0052
(6) <i>Srev</i>	0.0126	0.0036	0.012	0.0039	(16) <i>Skew</i>	0.0073	0.0025	0.0072	0.0029
(7) <i>Maxret</i>	0.0125	0.0026	0.0103	0.0027	(17) <i>LA</i>	0.0065	0.0008	0.0063	0.0011
(8) <i>Price</i>	0.0115	0.0023	0.0104	0.0028	(18) <i>AbT</i>	0.0063	0.0027	0.0057	0.0029
(9) <i>Return</i>	0.0108	0.0031	0.0095	0.0032	(19) <i>NEDMF</i>	0.0048	0.0021	0.0038	0.0026
(10) <i>Age</i>	0.0098	0.0017	0.0098	0.0016	(20) <i>DPay</i>	0.0047	0.0018	0.0040	0.0016



附图 1 基金偏好股票因子载荷



附图 2 基金聚类结果



附图 3 基金偏好在个体投资者分类下的对比

## 附录 2 详细特征筛选执行过程

第一步，剔除持股偏好度较低与随机异质性无明显差异的特征。具体方法为每期保持投资者持仓不变（包括持仓股票与金额占比），构建一个随机特征  $C_s^{rand}$ ，并计算该特征的偏好分歧  $Var_{rand}$ 。随后，重复该过程 10000 次，即可得到一个关于随机偏好分歧的抽样分布。提取随机抽样分布的平均值作为最终当期投资者随机的持股偏好并同样计算全部时期的平均值  $\bar{Var}_{rand}$ ，并删去所有时序平均小于该值的特征。

第二步，剔除相关性较高的特征。其伪代码如附表 2 所示。

附表 2 相关性筛选过程

---

### 算法 基于相关性的选股特征筛选

---

输入：特征持仓分歧度排名表  $Var$ ，相关系数矩阵  $Corr$ ，相关系数阈值  $\theta$ 。

输出：经过筛选后的特征持仓分歧度排名表  $Var^*$

初始化： $i = 1$

1: While  $i < \text{len}(Var)$  do

2:     在选股特征偏好度排名  $Var$  中选取第  $i$  个特征  $Var_i$

3:     在相关系数矩阵  $Corr$  中获取与特征  $Var_i$  相关系数绝对值大于  $\theta$  的特征  $Var_{select}$

4:     在相关系数矩阵  $Corr$  删除特征  $Var_{select}$  所在的行和列

5:     在选股特征偏好度排名  $Var$  删除特征  $Var_{select}$  所在的行

6:      $i = i + 1$

7: endwhile

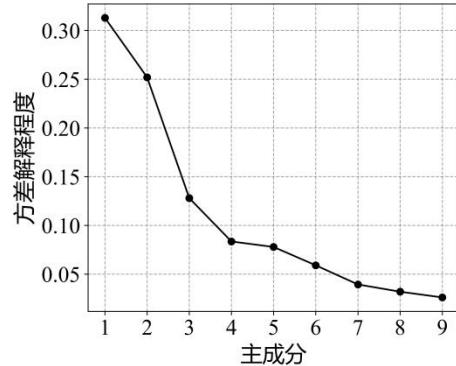
8: 返回经过特征筛选后的选股特征偏好度结果  $Var^*$

---

以此类推最终可求得经过特征筛选后的相关系数矩阵，如附图 4 所示。附图 5 汇报了经过特征筛选后，提取主成分方差解释度的大小。

	Price	EP	Vol	Return	Illiq	Skew	AbT	SOE	Srev
Price		0.12	0.35	0.34	-0.22	-0.12	0.13	-0.21	-0.00
EP	0.12		-0.46	0.05	-0.45	0.03	-0.05	0.12	-0.00
Vol	0.35	-0.46		0.46	0.26	-0.00	0.18	-0.28	0.01
Return	-0.34	0.05	0.46		-0.12	0.06	0.44	-0.07	0.32
Illiq	-0.22	-0.45	0.26	-0.12		-0.03	-0.00	-0.17	0.02
Skew	-0.12	0.03	-0.00	0.06	-0.03		0.02	0.03	0.08
AbT	-0.13	-0.05	0.18	0.44	-0.00	0.02		-0.06	0.37
SOE	-0.21	0.12	-0.28	-0.07	-0.17	0.03	-0.06		0.02
Srev	-0.00	-0.00	0.01	0.32	0.02	0.08	0.37	0.02	

附图 4 个体投资者筛选后特征相关系数



附图 5 不同主成分方差解释程度

### 附录 3 关注度、壳价值以及相关性检验

关注度 (*Attention*) : 采用月度异常搜索量 ( $Attention_t = \log(SVI_t) - \log[Med(SVI_{t-1}, \dots, SVI_{t-36})]$ ) 作为关注度指标; 其中  $\log(SVI_t)$  为第  $t$  月搜索量的对数,  $Med(\dots)$  为中位数计算方法。

壳价值 (*ShellValue*) : 参考屈源育等 (2018) 使用  $SV \times Pr/ME$  衡量上市公司的壳价值含量, 其中  $SV$  代表如果上市公司参与借壳交易能够实现的壳价值,  $Pr$  为该上市公司参加借壳交易的概率。 $SV$  是一个根据借壳上市交易数据构建的壳价值指数:

$$SV_{it} = \exp(-4.421 + 3.118 \times \ln ME_{it-1} - 0.187 \times (\ln ME_{it-1})^2 - 0.592 \times SOE_{it-1}) \quad (A1)$$

$Pr$  是通过 Logit 模型估算出的一个上市公司在下一期成为“壳”的概率:

$$\begin{aligned} Pr_{it} = & (1 + \exp(-(-21.7 - 1.85 \times Rsize_{it-1} - 11.40 \times ROA_{it-1} - 0.93 \times sales\_g_{it-1} \\ & + 0.73 \times ST_{it-1} + 0.97 \times IPO\_rej_{it-1} + 1.35 \times insider_{it-1} + 0.57 \times ret_{it-1}))^{-1} \quad (A2) \end{aligned}$$

其中,  $Rsize$  为经过沪深 300 总市值调整的上市公司市值,  $ROA$  为总资产收益率,  $sales\_g$  为主营业务收入增长率,  $IPO\_rej$  为 IPO 拒绝率,  $insider$  为内部人持股比例,  $ret$  为前 12 个月股票相对于市场的超额收益率。

附表 3 个体投资者主成分相关性检验

	RIPC1	RIPC2	RIPC1_noprice	Attention	ShellValue
RIPC1	1				
RIPC2	0.0882	1			
RIPC1_noprice	0.8731	-0.0806	1		
Attention	<b>-0.4341</b>	-0.0194	<b>-0.4942</b>	1	
ShellValue	0.2133	<b>0.7011</b>	0.0979	-0.0577	1

### 附录 4 稳健性检验

附表 4 因子正交化后的个体投资者特征偏好分歧度

指标排名	等权		持仓金额加权		指标排名	等权		持仓金额加权	
	Mean	Std	Mean	Std		Mean	Std	Mean	Std
(1)Price	0.0631	0.0051	0.0602	0.0052	(11)Return	0.0512	0.0029	0.0483	0.0055
(2)EP	0.0601	0.0059	0.0479	0.0038	(12)Beta	0.0509	0.0027	0.0474	0.0047
(3)Vol	0.0598	0.0059	0.0523	0.0047	(13)Maxret	0.0507	0.0043	0.0457	0.0052
(4)SOE	0.0584	0.0034	0.0448	0.0051	(14)NEDMF	0.0506	0.0050	0.0453	0.0059
(5)ME	0.0538	0.0041	0.0521	0.0056	(15)DPay	0.0505	0.0025	0.0392	0.0032
(6)BM	0.0526	0.0018	0.0453	0.0050	(16)AbT	0.0492	0.0032	0.0454	0.0058
(7)Skew	0.0525	0.0026	0.0468	0.0041	(17)RMBTV	0.0486	0.0027	0.0480	0.0058
(8)ROE	0.0523	0.0026	0.0454	0.0045	(18)Srev	0.0485	0.0031	0.0468	0.0053
(9)Age	0.0513	0.0019	0.0444	0.0053	(19)IA	0.0477	0.0022	0.0441	0.0036
(10)Turnover	0.0513	0.0032	0.0473	0.0060	(20)Illiq	0.0462	0.0028	0.0421	0.0057

附表 5 主成分分析横向稳健性检验

	面板 A: 第一主成分							面板 B: 第二主成分							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)		(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) 20_all	1							1							
(2) 20_t5	0.9507	1						0.9377	1						
(3) 20_t4	0.8674	0.7996	1					0.8299	0.8264	1					
(4) 20_t6	0.9645	0.9526	0.8913	1				0.9430	0.9651	0.8590	1				
(5) 109_all	0.8448	0.7468	0.8705	0.8109	1			0.8980	0.7861	0.8278	0.8260	1			
(6) 109_t5	0.8899	0.9307	0.7378	0.9067	0.7913	1		0.9407	0.9752	0.7903	0.9344	0.7894	1		
(7) 109_t4	0.6475	0.7362	0.8349	0.6591	0.7624	0.7623	1	0.6512	0.7600	0.7174	0.6785	0.8421	0.8045	1	
(8) 109_t6	0.7391	0.8384	0.7462	0.7721	0.7925	0.9141	0.9070	0.7886	0.8683	0.6843	0.8531	0.8053	0.8859	0.9246	

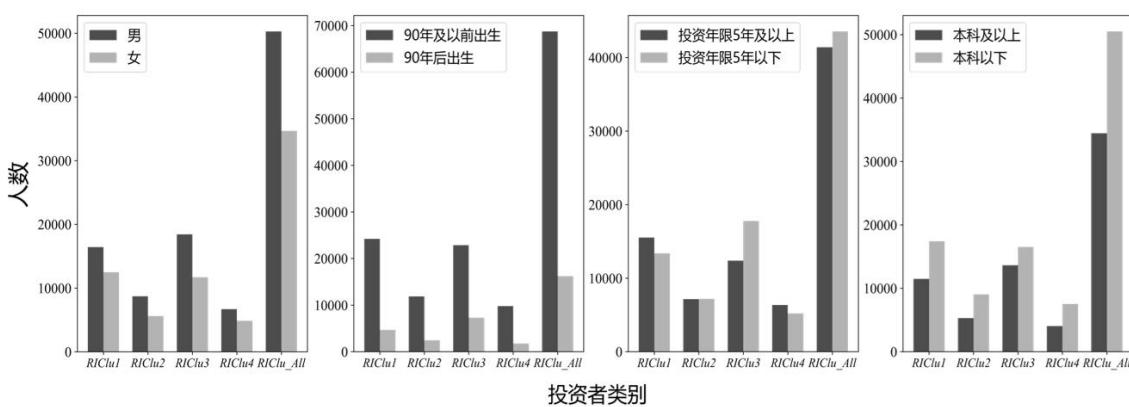
注：主成分因子载荷构成请详见网络附录。

附表 6 主成分分析时序稳健性检验

	面板 A: 第一主成分						面板 B: 第二主成分					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
(1) allyear	1						1					
(2) y16	0.9566	1					0.9147	1				
(3) y17	0.8427	0.9484	1				0.8178	0.9672	1			
(4) y18	0.7707	0.9122	0.9886	1			0.6932	0.8616	0.9542	1		
(5) y19	0.6053	0.8049	0.9096	0.9592	1		0.7947	0.5236	0.5022	0.3507	1	
(6) y20	0.9273	0.7824	0.6072	0.5239	0.3869	1	0.9110	0.7029	0.5251	0.4872	0.9723	1
(7) y21	0.8443	0.6538	0.5563	0.4134	0.4752	0.9786	0.7963	0.5399	0.4769	0.3766	0.9944	0.9741

附表 7 剔除持仓规模在 10 万以下的个体投资者特征偏好分歧度

指标排名	等权		持仓金额加权		指标排名	等权		持仓金额加权	
	Mean	Std	Mean	Std		Mean	Std	Mean	Std
(1) Price	0.0538	0.0034	0.0522	0.0061	(11) Illiq	0.0437	0.0037	0.0462	0.0068
(2) Vol	0.0521	0.0024	0.0491	0.0036	(12) RMBTV	0.0432	0.0034	0.0463	0.0059
(3) BM	0.0516	0.0031	0.0516	0.0043	(13) ROE	0.0432	0.0028	0.0466	0.0044
(4) Turnover	0.0484	0.0027	0.0460	0.0051	(14) Srev	0.0399	0.0032	0.0445	0.0053
(5) NEDMF	0.0480	0.0040	0.0490	0.0042	(15) Skew	0.0397	0.0026	0.0404	0.0041
(6) Return	0.0475	0.0034	0.0477	0.0045	(16) AbT	0.0394	0.0055	0.0414	0.0071
(7) EP	0.0470	0.0031	0.0439	0.0030	(17) SOE	0.0382	0.0024	0.0384	0.0029
(8) Maxret	0.0470	0.0032	0.0465	0.0054	(18) IA	0.0362	0.0025	0.0387	0.0051
(9) Beta	0.0460	0.0030	0.0456	0.0041	(19) Age	0.0333	0.0026	0.0352	0.0048
(10) ME	0.0438	0.0027	0.0421	0.0035	(20) DPay	0.0283	0.0021	0.0279	0.0047



附图 6 基于性别、年龄、投资年限、学历的各类投资者比例

附录 5 注册制改革下个体投资者的壳价值偏好演化

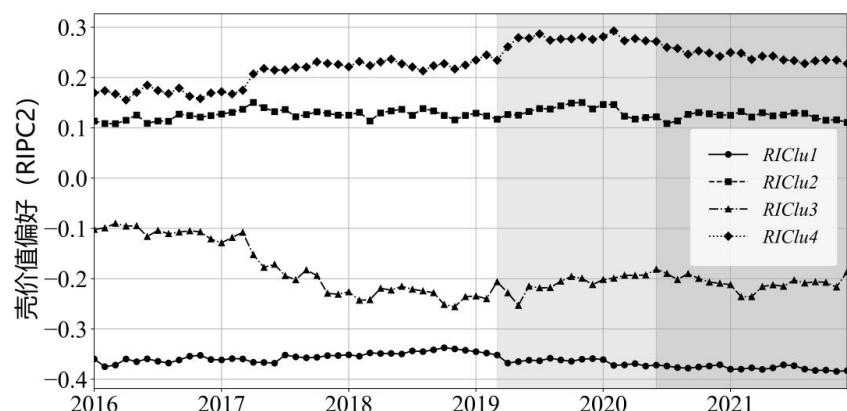
本文核心发现之一为个体投资者最大的股票类型偏好分歧是壳价值，然而近年来资本市场不断优化，推行以信息披露为核心的股票发行注册制改革，一个自然而然的问题是“在注册制改革相关政策逐步推行过程中，个体投资者的政策反馈如何？”。

分月度计算全部个体投资者的壳价值偏好 ( $RIPC2$ ) 均值, 结果如附图 7 所示。从时序演化的角度看, 个体投资者壳价值偏好的变化具有明显的政策反馈, 大致可分为三个阶段: 第一阶段为 2016 年 1 月至 2019 年 1 月, 投资者对于壳价值偏好整体处于小幅波动的状态, 但大体上较为稳定; 第二阶段为 2019 年 1 月至 2020 年 5 月, 《关于在上海证券交易所设立科创板并试点注册制的实施意见》与《科创板首次公开发行股票注册管理办法(试行)》相继出台, 明确了科创板注册制的具体规则, 在个股普遍上涨的乐观市场情绪中, 个体投资者对波动较大的壳价值型股票偏好有所上升; 但在第三阶段 2020 年 6 月后, 证监会发布创业板改革并试点注册制相关制度规则以及《科创板上市公司证券发行注册管理办法(试行)》, 明确了资本市场注册制改革的核心路线, 壳价值偏好开始迅速下降。总体来看, 投资者对政策出台有着积极反馈, 整体对于壳价值的偏好在注册制一系列文件的出台中有所降低。

进一步, 本文分别计算了四类个体投资者的月度壳价值偏好 ( $RIPC2$ ) 均值, 结果如附图 8 所示。四类投资者对壳价值型股票存在明显偏好分歧, 按照偏好程度排序中, 第四类“老练小资金型投资者” ( $RIClu4$ ) 和第二类“高换手风险追逐型投资者” ( $RIClu2$ ) 偏好于壳价值型股票。而第三类“新兴年轻型投资者” ( $RIClu3$ ), 投资者较为特殊, 在政策出台之前就开始逐渐调整持仓, 变更为偏向成熟稳健的投资组合, 壳价值偏好逐渐减弱。而第四类投资者本身就对于壳价值型股票具有极高偏好, 并且在政策讨论、试行的第一阶段和第二阶段对其偏好持续上升, 直到政策不断出台后, 对于壳价值的偏好才逐渐减小, 属于投资者教育需要重点关注的类别。偏好演化结果表明, 壳价值偏好分歧持续存在, 表明本文的偏好分歧识别与投资者类别划分具有稳健性。此外, 通过本文的技术路线, 采用行为决定定义的方式能够将投资者进行明确划分, 证明了本文分类方式的有效性。



附图 7 注册制改革下个体投资者的整体壳价值偏好演化



附图 8 注册制改革下不同类别个体投资者的壳价值偏好演化