

投资者情绪:理性与非理性*

胡昌生 池阳春

[摘要]本文对目前广泛使用的 Baker and Wurgler(2006)情绪指数构建进行了质疑,分析了 BW 情绪指数构建的前提条件,根据提取共同因子的思想提出了一种区分理性和非理性情绪的新思路。与现有研究不同的是,本文并没有仅将“理性”与基本因素相联系,而是从投资者追求财富最大化的角度对情绪指数中“理性”的范畴进行了重新界定;理性情绪反映的是“高明的投资”,对应地,非理性情绪反映的是“弱智的投资”。最后,我们对 15 种常用的单一情绪代理变量进行了比较分析,实证结果很好地支持了本文的理论观点。

关键词:情绪指数 理性情绪与非理性情绪 资产收益的可预测性

JEL 分类号:G11 G12 G14

一、引言

投资者情绪的度量是金融市场研究的基础性工作,现有文献中出现的情绪度量指标多达十几种,这些指标在丰富我们对投资者情绪特征理解的同时也给研究者如何选取一个统一的、具有代表性的情绪指数带来了困难。其中,主成分分析法由于能滤出与情绪无关的特质成分,得到更“纯”的投资者情绪而得到了最为广泛的应用。主成分分析法源于 Pearson(1901)和 Hotelling(1933)的工作,通过统计分析的方法实现数据降维,以减少研究变量个数而又能最大限度地保留原始变量中的信息。自从 Baker and Wurgler(2006)将这种方法运用到投资者情绪指数的构建之后,主成分分析法得到了广泛的认可,并成为此类研究的主流方法^①。从实证效果看,依据此法所构建的指数能较好地反映投资者的情绪特征,如对未来收益具有反向预测作用(Baker and Wurgler,2000; Frazzini and Lamont,2008),在横截面上对个体投资者偏好的股票、难以估值和套利的股票、有卖空限制的股票影响更为显著(Kumar and Lee,2006;Stambaugh et al.,2012)等。虽然Baker and Wurgler(2006)提到使用主成分分析法是因为各个情绪代理变量包含一个共同的情绪成分和与情绪无关的特质成分,主成分分析法的作用是“滤出特质噪音成分并捕捉到了变量中的共同成分”(Stambaugh et al.,2012),但是这些代理变量所包含的共同成分除了情绪成分还可能有其他成分。现有研究都忽略了主成分分析法可以被用来构建投资者情绪指数的前提条件,以至于很多学者在使用主成分分析法构建情绪指数时盲目地使用第一主成分。在很多情况下,这些研究都是用错误的方法得出了看似正确的结论,并且有很浓的数据挖掘色彩。本文首次明确提出可以使用主成分分析法构建投资者情绪指数的前提条件——投资者情绪成分必须要在所有的共同因素中占据的比例最大。但实际情况却可能由于所选取的代理变量不同而不能保证投资者情绪在原始变量中所

* 胡昌生,武汉大学经济与管理学院金融系,教授,博士生导师;池阳春,武汉大学经济与管理学院金融系,博士研究生。本文得到了国家自然科学基金“投资者情绪、资产估值与资产价格异常波动研究”(项目批准号:70971101)和教育部人文社科规划项目“投资者情绪与资产价格异常波动性研究”(项目批准号:09YJA790156)的资助。

^①事实上,在 Baker and Wurgler(2006)之前,Brown and Cliff(2004)已将这种方法运用到情绪指数构建中了,Baker and Wurgler(2006)则对这种方法进行了更为全面的讨论和检验。近期相关的文献包括 Glushkov(2006),Baker and Wurgler(2007),Ho and Hung(2009),Baker et al.(2012),Stambaugh et al.(2012)等。

占的信息比重最大。我们在后面的实证中给出了一个反例,说明片面使用第一主成分作为情绪指数并不可取,而是需要对各个成分进行仔细的检验。至于一些学者所使用的对各主成分加权获得情绪指数的方法,这从根本上混淆了降维与提取信息的关系。

Baker and Wurgler(2006), Kumar and Lee(2006)采用对宏观经济变量进行回归的方法剔除基本因素对投资者情绪的影响,以便能提炼到更“纯”的投资者情绪指数, Verma and Verma(2007)进一步把这种方法用来区分理性与非理性情绪。这实际上是把投资者情绪指数中能由基本因素解释的部分当作理性情绪,剩余部分当作是非理性情绪。Verma and Verma(2008), Verma and Soydemir(2009)也沿用了这一思路。我们认为这种定义理性与非理性情绪的方法有失偏颇。第一,投资者情绪作为投资者对未来收益的预期,具有前瞻性(Forward),而当期投资者情绪指标中被基本因素解释的部分更多的是后顾性的(Backward),或者说只是反映了当期的基本因素;第二,即使预先准确地知道了未来基本因素的信息,投资者就一定能对资产价格形成正确的预期吗?未来资产价格中的基本因素不一定能得到清算,即使要清算,还会有资本损益的问题。因此,如果要给投资者情绪冠以“理性”或“非理性”这样的词缀,我们认为理性不仅限于跟基本因素有关联,更重要的是能对未来收益有正确的预测,不可忽视投资者情绪之间的相互作用。本文基于提取共同因子的思想提出了一种区分理性和非理性情绪的新思路,这种思路不仅考虑到了理性情绪的前瞻性,而且还区分了投资者情绪与基本因素的关系。实证结果表明,理性情绪虽然不是当期基本因素的反映,但会受到前期基本因素的影响;而非理性情绪则与前期收益有关,并且不能由基本因素解释。

一个投资者情绪代理变量是否具有良好的代表性,以及如何对这些指数在作为投资者情绪代理变量的“好”与“坏”进行比较一直存有争议。根据 Baker and Wurgler(2007)总结的 12 种常见的投资者情绪指标和数据的可得性,本文对换手率、封闭式基金折价率、消费者信心指数、投资者新增开户数等 15 种常用单一情绪代理变量进行了比较分析^①。结果表明,没有一个所谓完全“好”的单一情绪代理变量,大多数单一代理变量都能在一定程度上反映投资者情绪,但都表现出不完整性。

本文余下部分结构安排如下:第二部分对相关文献进行述评;第三部分从理论上讨论了降维与提取信息的关系;第四部分构建了理性和非理性情绪指标并对 15 种常用的单一情绪代理变量进行了比较分析;最后一部分总结全文。

二、文献述评

情绪指数的构建是研究投资者行为特征的基础性工作。一般地,情绪指数可分为三类:显性指数、隐性指数和复合指数^②。前两类指数由于会受到指数本身特质的(Idiosyncratic)、与情绪无关的因素影响,而使得这些指数用作投资者情绪的代理变量并不一定具有代表性。如果我们能识别出指数中与情绪无关的特质因素并通过数据处理排除这些因素的影响,那么就能得到一个“纯”的投资者情绪代理变量。但是,现实中这些情绪代理变量本身的特质因素往往是难以识别的。即使我们能识别出其中某些特质因素,也常常是难以度量的。例如,投资者调查数据可能会受“知行不一”问题的困扰(Baker and Wurgler, 2007);共同基金现金流可能会存在由于基金业绩评价的偏差而导

^① Baker and Wurgler(2007)总结的 12 种情绪指标分别为:调查数据、情绪状态变量、个体投资者交易、共同基金现金流、成交量、股息溢价、封闭式基金折价率、期权隐含波动率、ipo 首日收益率、ipo 数量、股票相对发行量和内部交易。受数据来源和覆盖区间的限制,我们没有选取投资者情绪状态和期权隐含波动率。由于我国的股票和债券的发行受到了严格的控制,因此没有把 ipo 数量和股票相对发行量纳入复合指数的构建。

^② 显性指数是指通过直接调查得到的情绪指数;隐性指数是指通过市场中与投资者情绪有关的真实交易数据得到的情绪指数;复合指数是通过多个显性或隐性指数合成的情绪指数,也就是目前最常见的主成分分析法指标。

致的代理问题(Scharfstein and Stein,1990;Baker et al.,2011);换手率不仅反映投资者情绪,还有可能是投资者异质信念的代理变量(Boehme et al.,2006)。显然,单一情绪代理变量都存在着代表性不足的问题,基于此,Baker and Wurgler(2006)提出了用主成分分析法构建复合指数以提取各代理变量中的共同成分,达到获得“纯”情绪指数的目的。一般来说,各种投资者情绪代理变量的共同成分除投资者情绪外,还会反映基本因素。如果每个代理变量都包含三个层面的信息,即共同的基本因素、共同的情绪因素和特质因素,那么,先对每个代理变量进行排除基本因素的处理后,所提取的共同成分(第一主成分)就应该是投资者情绪了。这正是 Baker and Wurgler(2006)使用主成分分析构建复合情绪指数的思路。

BW 的做法至少有两点值得商榷:第一,假设在排除基本因素的处理后,各代理变量确实只包含了共同的情绪与特质成分,能使用第一主成分作为情绪指数的前提条件是,在原始变量中共同成分所占的信息(用方差表示)比重最大。但是,如果每个代理变量中,当共同的情绪成分只占很小的比重,其余大部分为特质成分时,那么根据最大方差标准得到的第一主成分就不一定是共同情绪成分了。第二,各代理变量除了共同的基本因素和情绪成分外,还有没有其他共同成分?如果存在,即使在排除基本因素影响后,如果其他共同成分所占的信息比重比情绪成分大,那么,通过主成分分析得到的复合指数的代表性就值得怀疑了。因此,现有研究在没有仔细检验各个主成分经济含义的情况下就使用第一主成分作为复合情绪指数,虽然结果不一定错误,但却可能是通过错误的方法得到看似正确的结论,并且在选取代理变量时,还必须小心翼翼,以使得第一主成分的效果最好,这一做法带有很浓的数据挖掘色彩。虽然上述第一个问题——单一代理变量的代表性,研究者可以根据经验和直觉挑选代理变量来解决。但是,第二个问题——复合指数的代表性,难以通过代理变量的选取解决。BW 情绪指数可能遗漏了除基本因素和情绪因素之外的重要共同成分。这种遗漏的重要共同成分是什么?我们认为仍然是投资者情绪。

投资者情绪存在着多个维度,具体包括几个维度与投资者情绪的界定有关。从广义上说,投资者情绪可以界定为任何对投资者交易决策有影响的因素。这样,投资者情绪代理变量中基本因素可以视为投资者情绪的一个维度,投资者情绪中与当期基本因素无关的对未来收益正确预期的部分也可以视为一个维度,而投资者情绪中与当期基本因素无关的对未来收益错误预期的部分又是另外一个维度。为了明白这一点,我们来看看下面的例子。

Frankel and Froot(1988)在 80 年代中期研究了外汇预测与咨询机构关于美元汇率走势的看法与投资建议。当时美元虽然已经历一段时间的升值期,但是美国与其他国家的利差并未变大,同时美国的贸易赤字在增加。他们发现,在这段时间这些机构大多预期美元在接下来几个月仍然会升值,但是在一年内会由于潜在基本因素而贬值。与这种预期一致,预测机构做出的投资建议是尽管美元相对于其基本价值是高估了,但是仍然买入。

投资者明知道资产价格已严重超出了其基本价值仍然做出买入决策的原因是什么?明显地,他们认为其他投资者也会做出买入决策。即,投资者情绪中与基本因素无关的部分不仅反映了投资者对未来收益的与基本因素无关的预期,还存在一个信念推理的过程。正如华尔街传奇人物 Bernard Baruch 说的那样:“市场波动中重要的不是事件本身,而是人们对这些事件的反应”。凯恩斯(1936)认为,“这批职业投资者与投机者之精力与才干,大都用在其他方面。事实上,这批人最关切者,不在比常人高出一筹,预测某一投资品在其寿命中所产之收益如何,而在比一般群众稍微早一些,预测决定市价之成规本身会有什么改变。从社会观点看,要使得投资高明,只有战胜时间和无知之神秘力量,增加我们对于未来之了解;但从私人观点,所谓最高明的投资,乃是先发制人,智夺群众,把坏东西让给别人”。这里,凯恩斯所言的“高明的投资(Smart Money)”我们的理解应该是指对未来收益的正确预期,反映的是投资者作为“经济人”追求财富最大化的过程。我们把投资者

情绪中对未来收益正确预期、反映“高明投资”的部分定义为理性情绪。理性情绪的特征有三:第一,要成为值得我们研究的对象,必须要对资产价格有系统性的影响,而不是市场中能相互抵消的噪音;第二,作为市场中“高明投资”的体现,理性情绪能正确预测未来收益(Schmeling, 2007);第三,虽然难以战胜“时间和无知之神秘力量”,增加对于未来的了解,但是基本因素仍是资产长期收益的重要决定因素(Fama and French, 1988; Campell, 1999),因此,理性情绪会受到前期基本因素的影响。与之对应的,则是“弱智的投资”(Dumb Money),是指对未来收益的错误预期,反映的是投资者的“动物精神”,我们把投资者情绪中对未来收益不那么“高明”的部分定义为非理性情绪。非理性情绪的特征为:第一,与理性情绪一样,要对资产价格有系统性的影响;第二,由于非理性情绪是投资者对未来收益的错误预期,必然会有一个均值反转的过程,因此,在长期是未来收益的一个反向指标,但是,在短期,由于价格压力效应(De Long et al., 1990a),非理性情绪对资产收益有正向的影响;第三,由于非理性投资者具有盲目跟风 and 追涨杀跌等正反馈交易特征(De Long et al., 1990b),非理性情绪与基本因素无关,而是会受到前期收益的影响。

三、降维与提取信息

复合指数的构建通常采用主成分分析法来实现,理论上其目的一是降维,二是提取信息。具体地,对于变量 $X_1 \cdots X_n$,如果他们之间的共同成分占比较大,那么,通过提取少数几个主成分就能包含原始变量中的大部分信息,从而达到降维和提取信息的目的。首先考虑简单的完全共线情形,对于两变量 X_1 和 X_2 ,如果二者完全是共线的,即 $X_2 = aX_1 (a \neq 0)$, X_1 的方差为 σ_1^2 ,那么 X_1 和 X_2 的协方差矩阵为:

$$\begin{bmatrix} \sigma_1^2 & a\sigma_1^2 \\ a\sigma_1^2 & a^2\sigma_1^2 \end{bmatrix} \cdot \begin{bmatrix} 1 & a \\ 1 & a \end{bmatrix} \quad (1)$$

其对应的特征多项式为:

$$\varphi(\lambda) = \lambda^2 - (a+1)\lambda = 0 \quad (2)$$

因此,必有一特征根为 0,其对应的主成分的解釋比例也为 0,是一个冗余变量,应该舍去。变量之间完全共线的情形下,使用主成分分析可以在不损失任何信息的条件下达到降维的目的。以上的结论很容易推广到多变量情形,即如果有一个变量能用其他变量线性表示,那么必有一个特征根为零。

对于 n 个变量 $X_1 \cdots X_n$,如果它们相互之间是完全独立的,相应的协方差矩阵为:

$$\begin{bmatrix} \sigma_1^2 & & \\ & \dots & \\ & & \sigma_n^2 \end{bmatrix} \quad (3)$$

这是一个对角矩阵。相应地,特征根为对角矩阵上的方差,第 i 个特征向量中第 i 个元素为 1 其余元素为零。因此,此种情况下主成分分析并没有产生任何有意义的线性变换。

实际情况常常介于两者之间,为了简单起见,我们考虑两变量情形。对于两个变量 X_1 和 X_2 ,如果它们分别包含一个共同成分和一个特质成分,即:

$$\begin{aligned} X_1 &= f + \varepsilon_1 \\ X_2 &= f + \varepsilon_2 \end{aligned} \quad (4)$$

其中 f 为共同成分, ε 为特质成分,为了简单起见,此处假设各个变量在共同成分上的载荷为 1。相应的协方差矩阵为:

$$\begin{bmatrix} \sigma_f^2 + \sigma_{\varepsilon_1}^2 & \sigma_f^2 \\ \sigma_f^2 & \sigma_f^2 + \sigma_{\varepsilon_2}^2 \end{bmatrix} \cdot \begin{bmatrix} 1+r_1 & 1 \\ 1 & 1+r_2 \end{bmatrix} \quad (5)$$

其中, $r_1 = \frac{\sigma_{\varepsilon_1}^2}{\sigma_f^2}$, $r_2 = \frac{\sigma_{\varepsilon_2}^2}{\sigma_f^2}$ 。其最大的特征值所占的方差解释比例为:

$$[(2+r_1+r_2)/2 + \sqrt{4+(r_1-r_2)^2}/(2+r_1+r_2)] \quad (6)$$

如果特质成分占比很小,如 $r_1=r_2=0$,那么所提取的第一主成分是共同成分。但是,如果特质成分占比很大,如 r_1 或 $r_2 \rightarrow \infty$,那么所提取的第一主成分就不是共同成分了。因此,能把第一主成分当作复合指数的前提条件是特质成分的方差比例不能太大。

更一般的情况是, n 个变量 $X_1 \cdots X_n$, 有 m 个共同因子 $f_1 \cdots f_m$, 不同的变量在共同因子上的载荷也不一定相等, 即

$$\begin{aligned} X_1 &= a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + \varepsilon_1 \\ &\cdots \\ X_n &= a_{n1}f_1 + a_{n2}f_2 + \cdots + a_{nm}f_m + \varepsilon_n \end{aligned} \quad (7)$$

从提取信息的角度来说,运用因子分析法更准确,也让我们更容易区分降维与提取信息关系。以上讨论可知,某一成分是否为共同情绪成分与所占信息比重(方差)的大小并无必然联系,其具体经济含义需要经过仔细检验才能确定。而一些学者将主成分进行加权以使信息比重更大的方法实际上是将完全独立的因子强行施加一个参数限制,混淆了降维与提取信息的关系。如前所述,我们假设投资者情绪代理变量中存在三个共同因子:基本因素、理性情绪和非理性情绪。由于主成分分析法所得到的主成分可以作为共同因子的近似,为保持与现有研究一致并便于直观比较,我们后面的分析仍然沿用主成分分析法^①。

四、实证分析

(一)初步分析

根据数据的可得性,我们选取了具有代表性的投资者情绪代理变量作为主成分分析的指标。这些指标分别为:换手率(turn)、封闭式基金折价率(cefd)、投资者新增开户数(open)和消费者信心指数(cci),分析的期间为2003年1月至2011年12月。我们对所有变量进行了标准化处理。在进行主成分分析之前,本文根据Baker and Wurgler(2006)的方法,确定各个代理变量的提前滞后关系,结果为封闭式基金折价率和投资者新增开户数采用滞后一期的数据,换手率和消费者信心指数采用当期数据^②。为了剔除基本因素的影响,我们使用Baker and Wurgler(2006)的方法,将情绪指标对宏观经济变量进行回归,选取相应的残差作为主成分分析的变量。选取的宏观经济变量为同比消费价格指数(cpi)、宏观经济景气指数(mci)和工业增加值增速比率(iavr)。主成分分析的结果如表1所示。

从表1中可以看到,第一主成分与第二主成分的解释比例分别为38.5%和32.1%,差别并不大。如果仅仅因为这6.4%的差别就将第一主成分作为复合情绪指数,显然难以令人信服。再者,第一主成分在换手率上的系数为负,在封闭式基金折价率上的系数为正,这与直觉不符。我们再来观察第二主成分,其在所有变量上的系数都符合投资者情绪的特征。从系数大小来看,第一主成分上系数最大的为封闭式基金折价率和消费者信心指数,除了符号与预期的不一致以外,封闭式基金

① 在附录中我们对主成分分析与因子分析的结果进行了比较,表明二者是非常类似的。

② 在附录中我们给出了具体的分析过程。

表 1 主成分分析结果

主成分	特征值	解释比例	累计解释比例	turn	L.cefd	cci	L.open
pc1	1.5400	0.3850	0.3850	-0.1152	0.6836	0.7056	0.1469
pc2	1.2840	0.3210	0.7060	0.7019	-0.0714	0.0364	0.7078
pc3	0.7185	0.1796	0.8856	0.6809	0.3522	-0.0980	-0.6346
pc4	0.4576	0.1144	1.0000	-0.1746	0.6352	-0.7009	0.2734

折价率的代理问题是有争议的^①, 消费者信心指数在很大程度上也可能会反映投资者对未来基本因素的预期。而第二主成分中系数最大的是换手率与投资者新增开户数。换手率作为情绪指数的代理变量有坚实的理论基础, Baker and Stein(2004)证明了在卖空限制下换手率是一个良好的情绪指标, 很少存在争议。投资者新增开户数作为情绪指标也是非常直观的, 而封闭式基金折价率和消费者信心指数的载荷并不大。由于第一主成分和第二主成分的解释比例十分接近, 如果以信息含量来评价两者的代表性, 差别并不大。但是这两个主成分哪一个是投资者情绪的代理变量? 由于我们已经剔除了当期基本因素的影响, 另一个在信息上具有同等重要性的主成分的经济含义是什么? 这些都需要我们作进一步的分析。

为了便于理解, 我们采用 Baker and Wurgler(2006)的历史情景解读法(Anecdotal History)进行一个直观的分析。图 1 依次给出了上证指数(pt)、第一主成分(pc1)和第二主成分(pc2)的时间序列。从图中可以明显地看到, 第一主成分与上证指数并没有清晰的提前滞后关系, 而第二主成分与上证指数的提前滞后关系非常明显。因此, 与第一主成分相比, 第二主成分更符合非理性情绪特征, 这说明了简单使用方差原则的不可靠性。

就与上证指数的关系而言, 虽然第一主成分看上去没有明显的提前滞后关系, 但是, 通过仔细观察可以发现, 在 2003 年底的低点和 2004 年 3 月的高点, 第一主成分都提前做出了反应, 在 2004 年底至 2006 年初这一段时间, 虽然市场走低, 但是第一主成分却在不断上行, 在 2006 年 1 月至 2007 年 10 月市场上升最快的时期, 第一主成分也提前做出了反应。与第二主成分不同的是, 第一主成分在这段时间波动十分剧烈, 市场指数每一次轻微的调整都带来了第一主成分剧烈的波动。如果我们把第一主成分看作是市场中“高明的投资”, 很明显, 这些“高明的投资”在市场非理性上升期间充分认识到了其中的风险, 一旦市场有任何风吹草动, 他们就能很快调整交易决策。然而, 他们的交易并不是简单的反向策略, 在确定市场还可能继续走高后, 他们又重新加入市场。尤其值得注意的是, 在 2008 年 10 月市场达到阶段性低点的时候, 第一主成分达到了 2007 年 10 月之后的阶段性高点。而在 2008 年 10 月以后市场的几个阶段性高点处, 第一主成分都处在阶段性的低点。与之形成鲜明对比的是, 第二主成分在对市场指数的反应方向上显得单一和呆滞。因此, 情景分析表明第一主成分更多地反映了市场上“高明的投资”, 符合我们前面所定义的理性情绪的特征。而第二主成分反映的是市场上“弱智的投资”, 更符合非理性情绪特征。在样本的最近一段时期, 虽然市场仍在不断下行, 但是第一主成分已经开始走高, 而第二主成分仍在低点。结合 2012 年上半年市场的走势, 我们对这两个指标的辨识得到了进一步的支持。

接下来, 我们再来检验第三和第四主成分是否为对市场收益有系统性影响的因素。表 2 给出了上证指数收益率(rt)与四个主成分滞后一期值的相关系数。由于这些主成分本身构造的关系,

^① 黄少安和刘达(2005)、伍燕然和韩立岩(2007)认为由于封闭式基金大都由个体投资者持有, 故该指标可用来衡量个体投资者情绪。然而, Chenet al.(1993)则质疑封闭式基金折价与小公司股票收益之间的正相关关系。Brown and Cliff(2004)通过分别对封闭式基金折价与小盘股和大盘股的收益进行回归, 结果发现前者并不能预测股票未来收益。张俊生等(2001)发现投资者情绪假说不能解释我国封闭式基金折价问题。

他们之间的几乎没有任何相关性。收益率与滞后一期的第一主成分和第二主成分之间的相关系数较大,与第三和第四主成分则几乎没有任何相关性,这初步说明第三和第四主成分对资产收益不具有系统性的影响。

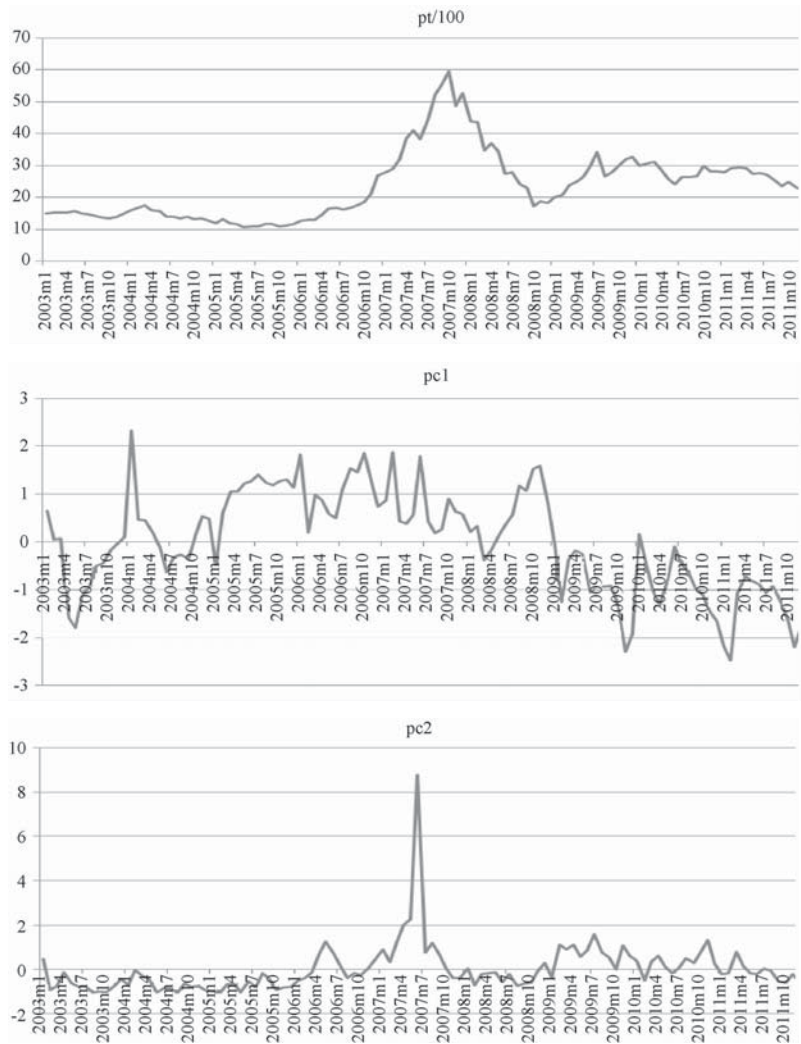


图1 上证指数、第一主成分和第二主成分时间序列图

表2 相关系数表

	rt	L.pc1	L.pc2	L.pc3	L.pc4
rt	1.0000				
L.pc1	0.1742	1.0000			
L.pc2	0.2856	-0.0043	1.0000		
L.pc3	0.0717	-0.0596	0.0056	1.0000	
L.pc4	0.0048	0.0701	0.0246	-0.0817	1.0000

表3给出了第一主成分和第二主成分的描述性统计量。从中可以看到,尽管根据在主成分分析中第一主成分所占的解释比例比第二主成分要大一些,但是,就这两个变量本身的方差而言,第二主成分的方差更大。从偏度来看,第一主成分略向左偏,而第二主成分严重地右偏。从峰度来看,第二主成分的峰度高达35.8263,其尾部很厚。这说明第一主成分所体现的投资者情绪悲观的时候

会比乐观的时候略多,而第二主成分所体现的投资者情绪在更多的时候是乐观的,并且其情绪很容易出现极端乐观和极端悲观的情形。因此,从描述性统计量来看,第一主成分更符合理性情绪特征,第二主成分更符合非理性情绪特征。从 jarque-bera 统计量来看,第一主成分具有良好的正态性,而第二主成分则不服从正态分布。ADF 统计量说明二者皆为单位根平稳的时间序列。

表 3 描述性统计

	mean	variance	skew	kurt	jarque-bera	adf
pc1	0.0000	1.0671	-0.1902	2.3653	2.4640	-3.077**
pc2	0.0000	1.1226	4.6221	35.8263	5233.61***	-5.838***

(二) 回归分析

为了进一步检验第三和第四主成分中是否存在未被发现的共同因子,我们对这两个主成分和上证指数收益率进行回归分析。表 4 给出了回归分析的结果,从中可以清楚地看到,第三和第四主成分对上证指数收益率几乎没有任何影响,也就是说这二者很大程度上是白噪音,而不是我们所忽略的共同因素。因此,只保留前两个主成分是合理的,我们后面不再把后两个主成分纳入分析范围。

表 4 第三第四主成分对收益的影响

滞后阶数	pc3		pc4	
	(1)	(2)	(3)	(4)
L.1	0.0082 (0.74)	0.0024 (0.19)	0.0008 (0.05)	-0.0098 (-0.49)
L.2		-0.0064 (-0.48)		0.0230 (1.06)
L.3		0.0194 (1.47)		0.0007 (0.03)
L.4		0.0177 (1.43)		-0.0203 (-1.06)
adj. R-sq	-0.0043	0.0282	-0.0095	-0.0149

注: *、**、和 *** 分别表示在 10%、5% 和 1% 的水平上显著。括号中给出的是相应变量的 t 统计量。其中(1)和(2)为第三主成分对上证指数收益率进行回归的结果,(3)和(4)为第四主成分对上证指数收益率进行回归的结果。

为了检验理性情绪(pc1)与非理性情绪(pc2)是否符合前面所定义的一些特征,我们用当期的理性和非理性情绪指数对上证指数未来 1-18 个月的收益分别进行回归^①,结果如表 5 所示。

表 5 中第 2、4 两列给出的是没有考虑宏观经济变量影响的结果,第 6、8 两列给出的是将宏观经济变量 mic、cpi、iavr 作为控制变量的结果,为了节省篇幅,我们没有在表中给出这些宏观经济变量的系数。每一系数估计值后面是相应回归得到的调整的 R²。从表中可以看到,第一和第二主成分对上证指数收益率有显著影响,是值得我们研究的系统性因素。无论是否控制宏观经济变量,第一主成分(pc1)在所有提前月份的系数都为正,虽然在第 14 至 18 个月不再具有显著性,但这也说明第一主成分对未来收益的预测至少是非负的。这有力地证明了第一主成分是一个能正确预测市场、引导投资者盈利的理性情绪指标。无论是否控制宏观经济变量,第二主成分只对未来 1 至 3 个月有显著的正向预测作用,并且其大小与显著性都在不断减弱,而在经过 5 个月(未来 4 至 8 个月)的不显著期后,对收益的预测变为负值,这与我们前面所定义的非理性情绪的第二条特征吻合。通过对两者的比较可以发现,第一主成分对未来收益的正向预测期间较长,而第二主成分对未来收益的正向预测期间较短,并且显著性在不断减弱。如果把这两个指标作为交易决策的依据,那么,以第一主成分作为交易决策依据的投资者对市场的反应更为敏锐。这再次印证了第一主成

^① Brown and Cliff(2005)也运用了类似的方法研究投资者情绪的预测性问题,不同之处在于他们的被解释变量为未来一段时间的平均收益(示来 6 个月、12 个月、24 个月和 36 个月)。为了更直观地体现理性与非理性情绪对未来收益的动态影响,本文不是对未来多期的平均收益,而是对未来单期收益进行回归。在未给出的结果中,我们也用 Brown and Cliff(2005)的方法以 3 个月为期间水平用投资者情绪对未来收益(3-36 个月)进行了回归,结果与表 5 类似,如有需要可向作者索取。

表 5 情绪指数对未来收益的预测分析

提前月数	pc1	adj. R ²	pc2	adj. R ²	pc1(adj.)	adj. R ²	pc2(adj.)	adj. R ²
1	0.0155* (1.81)	0.021	0.0239*** (3.05)	0.073	0.0153* (1.83)	0.062	0.0239*** (3.12)	0.116
2	0.0267*** (3.12)	0.077	0.0177** (2.21)	0.036	0.0264*** (3.19)	0.137	0.0177** (2.27)	0.096
3	0.0186** (2.09)	0.031	0.0153* (1.89)	0.024	0.0187** (2.15)	0.081	0.0153* (1.94)	0.073
4	0.0193** (2.15)	0.034	0.0084 (1.02)	0.000	0.0194** (2.19)	0.069	0.0085 (1.04)	0.035
5	0.0221** (2.45)	0.047	-0.0107 (-1.29)	0.007	0.0220** (2.45)	0.057	-0.0107 (-1.30)	0.016
6	0.0336*** (3.85)	0.121	0.0049 (0.58)	-0.007	0.0326*** (3.80)	0.156	0.0047 (0.57)	0.034
7	0.0226** (2.47)	0.049	-0.0036 (-0.43)	-0.008	0.0222** (2.44)	0.065	-0.0037 (-0.45)	0.009
8	0.0171* (1.83)	0.023	0.0022 (0.26)	-0.009	0.0165* (1.80)	0.081	0.0021 (0.26)	0.050
9	0.0139 (1.47)	0.012	-0.0177** (-2.14)	0.035	0.0144 (1.53)	0.037	-0.0176** (-2.16)	0.060
10	0.0191** (2.02)	0.031	-0.0065 (-0.77)	-0.004	0.0202** (2.12)	0.039	-0.0064 (-0.75)	-0.002
11	0.0183* (1.85)	0.025	-0.0116 (-1.37)	0.009	0.0205** (2.02)	0.009	-0.0115 (-1.34)	-0.015
12	0.0176* (1.72)	0.020	-0.0248*** (-3.02)	0.079	0.0193* (1.82)	-0.006	-0.0248*** (-2.97)	0.050
13	0.0232** (2.24)	0.041	-0.0159* (-1.87)	0.026	0.0260** (2.39)	0.023	-0.0158* (-1.84)	-0.001
14	0.0128 (1.19)	0.004	-0.0237*** (-2.83)	0.070	0.0186 (1.65)	0.010	-0.0237*** (-2.82)	0.063
15	0.0125 (1.15)	0.003	-0.0145* (-1.67)	0.019	0.0158 (1.35)	-0.008	-0.0145 (-1.66)	0.002
16	0.0069 (0.62)	-0.007	-0.0290*** (-3.51)	0.111	0.0136 (1.14)	-0.006	-0.0292*** (-3.53)	0.106
17	0.0145 (1.31)	0.008	-0.0069 (-0.78)	-0.004	0.0217* (1.80)	0.012	-0.0070 (-0.79)	-0.018
18	0.0103 (0.92)	-0.002	-0.0084 (-0.96)	-0.001	0.0191 (1.58)	0.015	-0.0087 (-0.99)	-0.002

注：*、**、和 *** 分别表示在 10%、5%和 1%的水平上显著。括号中给出的是对应变量的 t 统计量。

分反映的是市场中“高明的投资”的理性情绪，而第二主成分反映的是市场中“弱智的投资”的非理性情绪。

通过观察两指标提前 1-2 个月的系数大小及其显著性我们可以发现，在未来第一个月，不论是否控制宏观经济变量，第一主成分的系数显著性只有 10%，而第二主成份的系数显著性为 1%。在未来第二个月，第一主成分系数的显著性突然上升至 1%，系数也是增大的，而第二主成分系数

的显著性下降到 5%,并在第三个月变为 10%,系数变小。这说明与第一主成分相比,第二主成分对未来收益有显著的价格压力效应。根据第二主成分进行交易的投资者具有盲目跟风、追涨杀跌的非理性特征。由于第一主成分比第二主成分能更早地预测市场,也就能有更多的市场择时机会。

(三)因果检验

根据前面的定义,理性情绪与基本因素有关,与前期收益无关;非理性情绪与基本因素无关,与前期收益有关。为了检验这一点,我们对市场收益率、理性情绪(pc1)和非理性情绪(pc2)进行 Granger 因果检验,并把宏观经济变量作为外生变量,根据 SBIC 信息准则,所选取的滞后阶数为 1,检验结果如表 6 所示。与前面的分析一致,在短期内,理性与非理性情绪都对市场收益有正向的影响,但是非理性情绪对市场影响的显著性更强。同时,市场收益除了受投资者情绪影响以外,也确实会受到基本因素的影响。理性情绪不能由前期收益 Granger 引起,而是会受到基本因素的影响。非理性情绪则是由前期收益 Granger 引起的,基本因素不构成非理性情绪的 Granger 原因。这些结果都符合我们前面对理性与非理性情绪的定义。因此,用主成分分析法提取不同情绪代理变量中的共同的、正交的、对资产价格有系统性影响的因素是区分理性与非理性情绪的一个良好思路。但是,要确定哪一个主成分是理性情绪与非理性情绪,需要仔细地进行检验。

表 6 Granger 因果检验结果

	L.rt	L.pc1	L.pc2	L.mci	L.cpi	Liavr
rt	0.0213 (0.22)	0.0153** (1.96)	0.0235*** (3.10)	-0.0245** (-2.42)	0.0030 (0.34)	0.0006*** (0.06)
pc1	-0.0663 (-0.10)	0.8263*** (15.92)	-0.0422 (-0.84)	-0.1009 (-1.50)	-0.0185 (-0.32)	0.2824*** (4.42)
pc2	4.2650*** (4.28)	-0.0521 (-0.64)	0.4238*** (5.37)	0.1131 (1.07)	-0.0826 (-0.91)	-0.1216 (-1.21)

注:*,**、和*** 分别表示在 10%、5%和 1%的水平上显著。括号中给出的是对应变量的 t 统计量。

(四)比较分析

根据前面的研究,由于所构建的复合情绪指数具有良好的代表性,我们以此为基准,对 15 个单一情绪代理变量(4 个用于构建复合指数的单一情绪代理变量和其他 11 个常用的未纳入主成分分析的单一情绪代理变量)的代表性进行比较分析,考察单一代理变量相对于复合指数的显著性和能被复合指数解释的程度。我们首先分析复合指数和前 4 个单一情绪代理变量的关系,结果如表 7 所示。

从表 7 中可以看到,换手率和投资者新增开户数主要反映的是非理性情绪,从拟合优度来看,非理性情绪对换手率回归的拟合优度达到了 61.6%,这说明换手率作为非理性情绪的代理变量所包含的其他噪音信息更小。封闭式基金折价率和消费者信心指数主要体现的是理性情绪,二者的拟合优度分别为 50.4%和 52.5%,当市场中封闭式基金折价率越高时,理性情绪也越高涨,这说明市场中真正理性的“高明的投资”正确认识到了资产价格存在一个均值反转的过程。马尔基尔在他的畅销书《漫步华尔街》中也建议投资者投资于折价率高的封闭式基金。我们这里的结果表明封闭式基金折价率是一个典型的理性情绪指标,如果用作非理性情绪代理变量不可避免地会存在问题,这也是封闭式基金折价率存在争议的重要原因。投资者新增开户数虽然反映的是非理性情绪,但是拟合优度并不高,仅为 7.2%,这说明投资者新增开户数除受非理性情绪影响外,还存在其他一些影响因素。由于我们前面构建复合指数方法的关系,宏观经济变量与理性和非理性情绪完全无关。在前面排除基本因素的影响的过程中,我们没有把无风险利率(rf)包括在内,这里我们用

无风险利率对理性和非理性情绪进行回归就不会遇到其他宏观经济变量的问题了。从表中可以看出,无风险利率与情绪因素没有任何关系,这再次说明了我们所构建的理性与非理性情绪指数并不包含有当期基本因素的信息,是比较“纯”的投资者情绪指标,尽管理性情绪能被前期基本因素解释。

基于数据的可得性,接下来,我们对后面 11 个单一情绪代理变量进行比较分析,具体包括:好淡指数(hd)、IPO 首日收益率(iport)、IPO 数量(iponm)、内部交易(insidetrd)、股息溢价(divprm)、个体投资者活跃度(indtrd)、股票型基金持

股比例(stkshare)、债券型基金持股比例(bndshare)、股票型基金现金流(stkflow)、债券型基金现金流(bndshare)、股票型基金与债券型基金现金流之差(flowdiff)。在附录中,我们对这些指标进行了详细的说明,分析的结果如表 8 所示。

从表 8 中可以看到,好淡指数(hd)既有理性情绪的成分,也有非理性情绪的成分,从拟合优度来看,理性情绪为 33.3%,非理性情绪为 9.3%,因此,好淡指数主要反映的是理性情绪。IPO 首日收益率(iport)反映的是非理性情绪,拟合优度为 10.5%。从符号上来看,IPO 首日收益率越高,所反映的非理性情绪也越高。虽然同样为 IPO 数据,IPO 数量(iponm)却比 IPO 首日收益率多包含了一个维度的信息,不仅反映非理性情绪,还反映了理性情绪。从符号上来看,IPO 数量越大,理性情绪越低落,非理性情绪越高涨,这说明 IPO 数量是理性情绪的反向指标。结合中国股票市场的圈钱属性和 IPO 行政审批这一事实可知,IPO 在一定程度上是当局利用投资者非理性情绪高涨过度融资,从而在长期会对市场收益产生负面的影响,成为理性情绪的反向指标。内部交易(insidetrd)反映的是理性情绪。由于我们是通过大宗交易折价率构建内部交易指标,因此,大宗交易的折价率越高,所反映出来的理性情绪越低,市场中“高明的投资”对未来收益的预期也就越悲观。股息溢价(divprm)反映的是理性情绪,市场上对高红利的股票支付的价格越高,理性情绪越低,与封闭式基金折价率一样,说明市场中“高明的投资”充分认识到了股息红利存在着一个均值反转的过程,而不是能够持续高涨。个体投资者活跃度(indtrd)反映的是市场中的非理性情绪,个体投资者交易越活跃,所反映出来的非理性情绪也就越高,从拟合优度来看,拟合优度达到了 33.7%,这说明个体投资者是市场中“弱智的投资者”的典型代表。市场中的开放式基金虽然是属于机构投资者范畴,应该表现出更多的理性特征,但是由于受到个体投资者申购和赎回的影响和基金本身的资产配置要求,使得其交易行为难免也带有非理性特征。这一点在股票型基金的持股比例(stkshare)上体现得尤为明显。其所反映的理性和非理性情绪拟合优度分别为 8.9%和 3.0%,理性成分高于非理性

表 7 用于构建复合指数的单一情绪代理变量的宏观经济变量比较结果

	pc1		pc2		adj. R-sq
turn	-1.1076	(-1.56)			0.013
			5.5296***	(13.15)	0.616
cefd	7.8158***	(10.47)			0.504
			-0.9657	(-0.96)	-0.001
cci	2.7837***	(10.91)			0.525
			0.1219	(0.35)	-0.008
open	0.0018	(0.80)			-0.004
			0.0062***	(3.04)	0.072
mci	-0.0000	(-0.00)			-0.009
			0.0000	(0.00)	-0.009
cpi	0.0000	(0.00)			-0.009
			-0.0000	(-0.00)	-0.009
iavr	0.0000	(0.00)			-0.009
			0.0000	(0.00)	-0.009
rf	-0.0000	(-0.22)			-0.009
			-0.0000	(-0.76)	-0.004

注:*,**、和 *** 分别表示在 10%、5%和 1%的水平上显著。括号中给出的是对应变量的 t 统计量。

表 8 未用于构建复合指数的单一情绪代理变量比较结果

	pc1		pc2		obs.	adj. R-sq
hd	7.2953***	(5.43)			58	0.333
			3.1558**	(2.61)	58	0.093
iport	8.7954	(1.49)			108	0.011
			19.5749***	(3.67)	108	0.105
iponm	-45.2066**	(-2.28)			108	0.038
			42.5512**	(2.26)	108	0.037
insidetrd	-0.0454***	(-3.63)			106	0.104
			-0.0156	(-1.24)	106	0.005
divprm	-0.0177***	(-2.79)			84	0.076
			0.0063	(1.01)	84	0.000
indtrd	-0.1246	(-1.22)			66	0.007
			0.4391***	(5.83)	66	0.337
stkshare	-1.1789***	(-2.27)			69	0.089
			0.7036*	(1.76)	69	0.030
bndshare	-1.0560***	(-3.95)			69	0.177
			0.0843	(0.31)	69	-0.013
stkflow	0.2945**	(2.52)			69	0.073
			0.3346***	(2.91)	69	0.099
bndflow	0.4144***	(3.73)			69	0.159
			0.0095	(0.08)	69	-0.014
flowdiff	0.2378**	(2.00)			69	0.042
			0.3443***	(3.00)	69	0.105

注：*、**、和 *** 分别表示在 10%、5% 和 1% 的水平上显著。括号中给出的是对应变量的 t 统计量。

成分。而债券型基金的持股比例(bndshare)只反映了理性情绪,拟合优度达到了 17.7%。不论是股票型基金还是债券型基金,其持股比例越高的时候,理性情绪越低,这说明开放式基金投资行为本身并不具有理性特点,相反,是市场中“高明的投资”的反向指标。当市场中开放式基金的持股比例达到一个高点后,市场上涨的动力难以为继。

与持股比例相比,基金的现金流更多地反映个体投资者行为。从表 8 中可以看到,对于股票型基金的现金流(stkflow),理性情绪与非理性情绪拟合优度分别为 7.3% 和 9.9%,更多地反映了非理性情绪^①。债券型基金现金流(bndflow)则只反映了理性情绪,与非理性投资者相比,理性投资者会有更强的意愿投资于债券型基金以满足避险需求。股票型基金和债券型基金的现金流之差(flowdiff)更直接地反映了个体投资者的资金在两种基金之间的相对流动。从表 8 中可以看到,这种相对现金流动更多地反映的是非理性情绪,投资者在股票型基金上投入的资金相对在债券型基金上的越多,所反映的非理性情绪越高。我们对基金持股和现金流的比较分析说明了市场中的开放式基金并不是“高明的投资”的良好代表,而是会由于个体投资者的申购与赎回对其投资行为产生影响,表现出非理性特征。

从这一部分的讨论可以看出,不存在一个“纯”的理性或非理性单一情绪指标。每一单一投资者情绪代理变量都可能因为其特质因素或者同时包括理性和非理性情绪成分而减弱其单独作为理性或非理性情绪代理变量的代表性。在实践中,由于数据的可得性以及研究的期间水平等问题,

① 用基金现金流进行回归时,我们对所有变量进行了标准化处理。

不一定能找到一系列的情绪代理变量进行主成分或因子分析以构建复合指数,只能单独使用某一个显性或隐性指标。应该选取哪一个指标,本文对此提供了一个思路。从这部分的比较分析中我们可以看到,最具有代表性的非理性情绪指标为换手率(turn)和个体投资者活跃度(indtrd),它们和理性情绪没有显著的相关关系,而非理性情绪对它们的解释比例分别为 61.6%和 33.7%。换手率也是上面所讨论的代理变量中可得性最好、受期间水平限制最小的指标。因此,在数据可得性的限制下,我们建议使用换手率作为非理性情绪的代理变量,特别是在进行周和日甚至是更高频研究的时候。对于理性情绪指标的选取,从这部分的比较分析中可以看到,最具有代表性的指标是消费者信心指数(cci)、封闭式基金折价率(cefd)、和好淡指数(hd),相应的理性情绪的解释比例分别达到了:52.4%、50.4%和 33.3%,研究者可以根据数据的可得性和研究的期间水平进行选择。

五、结 语

本文对 Baker and Wurgler(2006)以来所广泛使用的用于构建复合情绪指数的主成分分析法进行了质疑,提出了主成分分析法应用上具有可行性的前提条件,厘清了降维与提取信息的关系,为相关研究提供了有价值的参考。从投资者追求财富最大化的角度,我们对理性与非理性情绪进行了重新界定,并提出了使用主成分分析(因子分析)对二者进行区分的新思路,同时还对现有研究中可得的单一投资者情绪代理变量进行了一个综合比较。在文章结尾我们想强调以下几点:

第一、情绪仍然是情绪。虽然本文所构建的理性情绪指标对未来收益存在着很强的正向预测,但它仍然属于投资者情绪范畴。尽管能被前期基本因素解释,但是从后面的比较分析我们也可以看到它与当期基本因素没有相关性。在横截面上,我们推测其总体特征仍然符合Baker and Wurgler(2006)难以估值和套利的假设。在具体分析中,建议同时使用理性和非理性情绪指标,以检验其具体的差异。

第二、复合指数的基准问题。由于用来构建复合指数的代理变量选取的差异,所得到的理性和非理性情绪指标的主成分也会不同。事实上,研究者通过自己良好的直觉和经验研究选取构建复合指数的代理变量,一般是能得到一个良好的非理性情绪的代理变量的,但是存在方法的错误,并且忽略了其他重要的共同因子。因此,我们建议在构建复合指数时,共同因子的具体经济含义要通过仔细检验才能确定。

第三、期间关系。构建复合指数的单一情绪代理变量所进行的排除基本因素的处理是针对当期的基本因素,这样做的目的是为了防止所得到的情绪指数仅仅是当期基本因素的反映,而在检验投资者情绪与基本因素的因果关系时,使用的是前期基本因素,以检验投资者情绪对前期基本因素的反应。这种期间关系的设定主要是基于我们对理性和非理性情绪的定义。未来收益的分布被凯恩斯称为“时间和无知之神秘力量”,当期基本因素和未来基本因素并不能在当期被直接观察到,再加上对其进行估计并不能完全确定未来收益,特别是短期收益。从心智的消耗来看,与此种“时间和无知之神秘力量”作斗争并不明智。而前期基本因素作为现成的历史信息不难获取。“高明的投资”与“弱智的投资”相比的一个重要区别在于前者的视界更宽,更不容易受表征性等非理性特征的影响(如,认为股票的好业绩会一直持续),而是会认知到资产的价格存在一个均值反转的过程,因此,会把前期基本因素这种容易获取而又被“弱智的投资”忽略的因素纳入考虑。这种期间上的提前滞后关系差异是理性和非理性情绪的重要区别。

第四、对资本市场的启示。本文的研究为投资者具体的交易决策提供了有益的参考。我们的分析结果给出了一系列理性情绪指标,投资者可以对这些指标进行综合考虑,同时,本文的非理性情绪指标也可以提醒投资者避免落入特定的非理性投资陷阱。

参考文献

- 黄少安、刘达(2005):《投资者情绪理论与中国封闭式基金折价》,《南开经济研究》,第4期。
- 凯恩斯(1936):《就业、利息和货币通论(中译本)》,商务印书馆1997年版。
- 马尔基尔(2010):《漫步华尔街(中译本)》,机械工业出版社。
- 伍燕然、韩立岩(2007):《不完全理性、投资者情绪与封闭式基金之谜》,《经济研究》,第3期。
- 张俊生、卢贤义、杨熠(2001):《噪声理论能解释我国封闭式基金折价交易现象吗?》,《财经研究》,第5期。
- Baker, M. and J. Wurgler (2000): "The Equity Share in New Issues and Aggregate Stock Returns", *Journal of Finance*, 55, 2219–2257.
- Baker, M. and J. Stein (2004): "Market Liquidity as A Sentiment Indicator", *Journal of Financial Markets*, 7, 271–299.
- Baker, M. and J. Wurgler (2006): "Investor Sentiment and the Cross-Section of Stock Returns", *Journal of Finance*, 61, 1645–1680.
- Baker, M. and J. Wurgler (2007): "Investor Sentiment in the Stock Market", *Journal of Economic Perspectives*, 21, 129–151.
- Baker, M., B. Bradley and J. Wurgler (2011): "Benchmarks as Limits to Arbitrage: Understanding the Low-volatility Anomaly", *Financial Analysts Journal*, 67, 40–54.
- Baker, M., J. Wurgler and Y. Yuan (2012): "Global, Local, and Contagious Investor Sentiment", *Journal of Financial Economics*, 104, 272–287.
- Boehme, R., B. Danielsen and S. Sorescu (2006): "Short-Sale Costs, Differences of Opinion and Over-valuation", *Journal of Financial and Quantitative Analysis*, 41, 455–487.
- Brown, G. and M. Cliff (2004): "Investor Sentiment and the Near-Term Stock Market", *Journal of Finance*, 11, 1–27.
- Brown, G. and M. Cliff (2005): "Investor Sentiment and Asset Valuation", *Journal of Business*, 78, 405–440.
- Campbell, J. (1999): "Asset Prices, Consumption, and the Business Cycle", in *Handbook of Macroeconomics*, edited by Taylor, J. and M. Woodford. Vol. 1, North-Holland, Amsterdam, 1231–1303.
- Chen, F., R. Kan and M. Miller (1993): "Are the Discount on Close-end Funds a Sentiment Index", *Journal of Finance*, 48, 795–800.
- De Long, J., A. Shleifer, L. Summers and R. Waldmann (1990a): "Noise Trader Risk in Financial Markets", *Journal of Political Economy*, 98, 703–738.
- De Long, J., A. Shleifer, L. Summers and R. Waldmann (1990b): "Positive Feedback Investment Strategies and Destabilizing Rational Speculation", *Journal of Finance*, 45, 375–395.
- Fama, E. and K. French (1988): "Dividend Yields and Expected Stock Returns", *Journal of Financial Economics*, 22, 3–25.
- Frankel, J. and K. Froot (1988): "Explaining the Demand for Dollars: International Rates of Return and the Expectation of Chartists and Fundamentalists", in *Agriculture, Macroeconomics and the Exchange Rate*, edited by Chambers, R. and P. Parilberg, Westfield Press, Boulder, Co.
- Frazzini, A. and O. Lamont (2008): "Dumb Doney: Mutual Fund Flows and the Cross-Section of Stock Returns", *Journal of Financial Economics*, 88, 299–322.
- Glushkov, D. (2006): "Sentiment Beta", *SSRN Working Paper*, Available at: <http://www.inquire-europe.org/seminars/2007/papers%20Brighton/Abstract%20Glushkov.pdf>
- Ho, C. and C. Hung (2009): "Investor Sentiment as Conditioning Information in Asset Pricing", *Journal of Banking & Finance*, 33, 892–903.
- Hotelling, H. (1933) "Analysis of A Complex of Statistical Variables into Principal Components", *Journal of Educational Psychology*, 24, 417–441, 498–520.
- Kumar, A. and C. Lee (2006): "Retail Investor Sentiment and Return Comovements", *Journal of Finance*, 61, 2451–2486.
- Pearson, K. (1901): "On Lines and Planes of Closest Fit to Systems of Points in Space", *Philosophical Magazine*, 62, 559–572.
- Scharfstein, D. and J. Stein (1990): "Herd Behavior and Investment", *American Economic Review*, 80, 465–479.
- Schmeling, M. (2007): "Institutional and Individual Sentiment: Smart Money and Noise Trader Risk", *International Journal of Forecasting*, 23, 127–145.
- Stambaugh, R., J. Yu and Y. Yuan (2012): "The Short of It: Investor Sentiment and Anomalies", *Journal of Financial Economics*, 104, 288–302.
- Verma, R. and G. Soydemir (2009): "The Impact of Individual and Institutional Investor Sentiment on the Market Price of Risk", *Quarterly Review of Economics and Finance*, 49, 1129–1145.
- Verma, R. and P. Verma (2007): "Noise Trading and Stock Market Volatility", *Journal of Multinational Financial Management*, 17, 231–243.
- Verma, R. and P. Verma (2008): "Are Survey Forecasts of Individual and Institutional Investor Sentiments Rational", *International Review of Financial Analysis*, 17, 1139–1155.

(责任编辑:周莉萍)

附录

(一)因子分析与主成分分析的比较

因子分析法是提取共同因子更为常见和直观的方法,由于使用主成分分析法得到的共同成分是因子分析的近似,同时也方便与现有研究进行比较,我们在正文中给出的是主成分分析的结果。表 9 给出了使用主成分因子法 (Principal Factors)(pf) 和迭代主成分法 (Iterated Principal Factors)(ipf)所得到的共同因子。从表 9 中可以看到,所有方法的第一个因子在换手率上的载荷都是负的,在封闭式基金折价率上的符号为正,与直觉不符,而第二个因子在情绪代理变量上的符号与预期的一致。这再次说明了以信息含量确定共同因子经济含义的方法不可取。表 10 给出了主成分分析法和这两种因子分析法所得到的第一第二因子(主成分)的相关系数表。从中可以看到,第一因子(主成分)之间、第二因子(主成分)之间的相关性很高,并且第一、第二主因子(主成分)之间相关性很低,因此,使用主成分分析法得到的第一、第二主成分是共同因子的很好近似。

表 9 因子分析结果

	pf1	pf2	ipf1	ipf2
turn	-0.0873	0.4282	-0.0929	0.5371
L.cefd	0.6139	-0.0548	0.6926	-0.0868
cci	0.6460	0.0329	0.7874	0.0360
L.open	0.1108	0.4496	0.1354	0.6029

表 10 相关系数表

	pc1	pf1	ipf1	pc2	pf2	ipf2
pc1	1.0000					
pf1	0.9997	1.0000				
ipf1	0.9994	0.9997	1.0000			
pc2	0.0037	0.0000	0.0130	1.0000		
pf2	0.0078	0.0038	0.0170	0.9998	1.0000	
ipf2	-0.0031	-0.0077	0.0056	0.9989	0.9996	1.0000

(二)提前滞后的选择

本文采取以下方法:首先,选择所有情绪代理变量的当期值和滞后期值,共 8 个指标,使用主分分析法,构造一个综合指数 COMP;第二,计算各投资者情绪因子当期值和滞后期值与第一步所构造综合指数的相关系数,相关系数较高者即表明是二者中蕴含投资者情绪因素相对较多的一个;最后,再用选择出的情绪因子或者其滞后值做主成分分析。主成分分析结果见表 11。

为了尽可能多地提取信息,我们选取主成分使得其累计方差解释率不低于 85%,从表 11 上可以看到,前 4 个主成分累计方差解释为 88.1%,因此,选取前四个主成分合成一个综合指标COMP,其合成方式是通过分别加总前 4 个主成分与其特征根之积,再除以累计解释方差得到。即: $COMP=(2.9081*cp1+2.0826*cp2+1.2027*cp3+0.8548*cp4)/0.8810$ 注意,我们此处的目的不是为了提取信息,而是要最大限度地保留原始变量中的信息,因此采用了对各个主成分进行加权的处理方式。

在得到综合指标 COMP 后,我们计算进行主成分分析所使用的指标及其滞后值与 COMP 的相关系数以确定各指标的当期值还是滞后值进入我们最后的情绪指标构建中,结果如表 12 所示。可以看到,换手率的当期值、封闭式基金折价率的滞后一期值、消费者信心指数的当期值和新增开户数的滞后一期值与 COMP 的相关性更高,即封闭式基金折价率和开户数两个指标提前反映投资者情绪。因此,我们选择这 4 个变量进行主成分分析以构建投资者情绪指数。

表 11 加入滞后项的主成分分析结果

主成分	cpc1	cpc2	cpc3	cpc4	cpc5	cpc6	cpc7	cpc8
特征根	2.9081	2.0826	1.2027	0.8548	0.5852	0.2339	0.1006	0.0321
方差解释度	0.3635	0.2603	0.1503	0.1068	0.0732	0.0292	0.0126	0.0040
累计解释率	0.3635	0.6238	0.7742	0.8810	0.9542	0.9834	0.9960	1.0000
turn	-0.1939	0.5645	0.1806	0.0185	-0.3644	-0.6838	0.0444	-0.0920
cefd	0.4792	0.0678	0.4960	0.0649	0.0930	-0.0904	0.0210	0.7058
cci	0.4831	0.1236	-0.3932	-0.0632	-0.2890	-0.0324	-0.7129	-0.0026
open	0.0490	0.4175	-0.1033	-0.7251	0.5356	-0.0003	0.0097	-0.0050
L.turn	-0.2077	0.5686	0.1490	0.0712	-0.2969	0.7106	-0.0430	0.1065
L.cefd	0.4703	0.0915	0.5012	0.1099	0.1248	0.1040	-0.0044	-0.6934
L.cci	0.4792	0.1060	-0.4033	-0.0737	-0.3168	0.0595	0.6978	-0.0168
L.open	0.0555	0.3799	-0.3465	0.6657	0.5330	-0.0622	0.0222	0.0293

表 12 各变量与综合指标的相关系数表

	turn	cefd	cci	open	L.turn	L.cefd	L.cci	L.open
COMP	0.4646	0.6332	0.6401	0.3026	0.3516	0.6587	0.5686	0.5218
观测值个数	108	108	108	108	108	108	108	108

(三)数据介绍及关于指标构建的说明

换手率(turn):换手率为全部 A 股的换手率,数据来源于锐思数据库。数据覆盖区间为 2003 年 1 月至 2011 年 12 月,共 108 个观测值。

封闭式基金折价率(cefd):本文通过对我国市场上 85 只封闭式基金净值的整理,算出其各自的每月的折价率,然后以其净值为其权重,用加权平均的方法计算了市场整体的封闭式基金折价率。封闭式基金月度单位净值和收盘价的数据来源于国泰安数据库。数据覆盖区间为 2003 年 1 月至 2011 年 12 月共 108 个观测值。第 i 只封闭式基金第 t 期的折价率为:

$$cefd_{it} = (nav_{it} - price_{it}) / nav_{it}$$

其中, $price_{it}$ 为基金的市场价格, nav_{it} 为基金的净值。

整个市场第 t 期的封闭式基金折价率为:

$$cefd_t = \frac{\sum_{i=1}^n nav_{it} * cefd_{it}}{\sum_{i=1}^n nav_{it}}$$

消费者信心指数(cci):消费者信心指数来源于中经网,数据覆盖区间为 2003 年 1 月至 2011 年 12 月共 108 个观测值。

投资者新增开户数(open):随着中国股票市场的发展,投资者的参与程度必然不断增加,从长期来看,投资者的新增开户数处于一个上升趋势,为了剔除这种影响,我们将投资者新增开户数用市场中的总户数进行调整,即:

$$open_t = t \text{ 期新增净开户数} / t \text{ 期累计总开户数}$$

数据覆盖区间为 2003 年 1 月至 2011 年 12 月共 108 个观测值。

好淡指数(hd):本文的好淡指数来源于《股市动态分析杂志》。好淡指数更新的频率为周,包括

短期(周)和中期(月)指标,本文使用的第 t 期的好淡指数为当月的好淡指数中期指标的平均值。数据覆盖区间为 2007 年 3 月至 2011 年 12 月,共 58 个观测值。

IPO 数据: IPO 数据来源于国泰安数据库,数据覆盖区间为 2003 年 1 月至 2011 年 12 月。由于融资是中国股市的最重要功能,中间我国几次暂停 IPO 的期间在很大程度上都是由于市场丧失了融资功能,因此,我们把 IPO 暂停期间的 IPO 首日收益率和 IPO 数量设定为 0,共计 108 个观测值。IPO 首日的收益率($iport$)是通过将当月每只 IPO 股票首日收益的平均后得到。IPO 数量($iponm$)为首发募集资金数额(亿元)。

内部交易($insidetrd$):内部交易指标通过大宗交易折价率计算得到。在每一个交易日,我们先计算出每一只股票的折价率:

$$disc_{it} = 1 - \frac{\sum_{j=1}^n volm_{ijt} \cdot prc_{ijt}}{\sum_{j=1}^n volm_{ijt} \cdot close_{it}}$$

其中 $volm_{ijt}$ 为第 i 只股票第 j 笔交易的成交量, prc_{ijt} 为相应的成交价格, $close_{it}$ 为该只股票的交易所收盘价,由于大宗交易平台的交易时间为每个交易日的 15:00-15:30,因此我们选取当天的该只股票的收盘价作为计算折价率的基准。在计算出每一只股票的折价率后,整个市场的折价水平通过将每一只股票的折价率平均后得到。大宗交易的数据来源于国泰安数据库,数据覆盖区间为 2003 年 1 月至 2011 年 12 月,中间有两个月的数据缺失,实际的观测值个数为 106 个。

股息溢价($divprm$):本文的股息溢价的计算公式为:

$$divprm_t = \log(hldex_t) - \log(pt_t)$$

其中 $hldex_t$ 为红利指数, pt 为上证指数。由于红利指数从 2005 年 1 月开始发布,因此,数据覆盖区间为 2005 年 1 月至 2011 年 12 月,共有 84 个观测值。

个体投资者活跃度($indtrd$):个体投资者活跃度数据来源于某证券公司两个营业部。为了更好地体现个体投资者特征,我们去掉了每月最高成交金额 200 万元或者平均成交金额大于 100 万元的账户的数据。经过初步处理后,从 2006 年 7 月至 2011 年 12 月,共 14033 个投资者账户,统计期内所有样本股票存在的有效交易记录共计 1139583 条。个体投资者账户的平均资产规模为 246636 元,平均交易规模为 17518 元。个体投资者活跃度指标是先通过计算每一个投资者当月的交易次数,然后再将所有个体投资者的交易次数平均后得到,观测值个数为 66 个。

基金数据:基金的数据来源于国泰安数据库,我们去掉了投资风格不明确的基金,由于基金的持股比例和申购赎回只有季度数据,因此,我们假设基金的持股在本季度不变,其现金流均值地发生在本季度的每一个月。基金持仓数据用的是样本内每一期基金持仓的平均值,基金现金流的数据用的是整个样本的总数据,数据覆盖区间为 2006 年 4 月至 2011 年 12 月,共有 69 个观测值。