

散户偏好与股市博彩溢价——

基于中国股市的研究

摘要：投资者偏好具有博彩特性的股票，因此这类股票应该具有更低的期望收益率。但是，并不是所有投资者对股票的博彩特性持相同的偏好。相对于专业的机构和个人投资者，散户表现出更强的博彩偏好。本文基于机构投资者持股比例和一系列散户偏好的股票特征构建了散户偏好指标。基于该指标本文研究中国股市中散户偏好与博彩溢价之间的关系。（1）双变量分组和公司层面截面回归方法的结果表明，散户偏好越强的股票中博彩溢价越高。即使对于在全样本下不具有显著定价能力的部分博彩变量，在散户偏好最强的股票中该变量也能够表现出显著的定价能力。（2）基于子样本分析的方法，本文表明在剔除一部分散户偏好股票之后，剩余股票中博彩变量失去其定价能力。具有显著博彩溢价的股票所占的市值比重不大。（3）由于散户易受市场波动的影响，因此散户主导的博彩偏好溢价在股市表现良好时期更为明显。

一、引言与文献综述

Walker（1992）从生物学、心理学、宗教信仰和社会特征等方面证明人们存在“以小博大”的赌博倾向。Statman（2002）则指出博彩收益和股票投资收益存在众多共同点。有些股票具有在小概率下获得高收益率的博彩特性。这种特性被投资者所偏好，因此这类股票应该具有更低的期望收益率。一些学者通过构建理论模型对上述观点进行论证。Mitton和Vorkink（2007）将博彩偏好引入到效用函数中，而Brunnermeier和Parker（2005）、Brunnermeier et al.（2007）以及Barberis and Huang（2008）则通过博彩偏好影响概率判断的角度进行建模。尽管以上两类模型的建模思路有所不同，但是结论都表明在均衡市场结果中博彩偏好被定价，投资者愿意持有具有博彩特性股票，并为此接受较低的期望收益率。尽管使用不同的博彩测度方式，Zhang（2005）、Kapadia（2007）、Kumar（2009）、Boyer et al.（2010）以及Bali et al.（2011）的实证研究结果都支持以上的观点。在国内，孔东民等（2010）、郑振龙和孙清泉（2013）、郑振龙等（2013）、江曙霞和陈青（2013）以及李培馨等（2014）的研究表明博彩偏好在中国股市中被定价。

尽管前人在相关领域做出了重要的工作，但是本文需要进一步指出的是，股票的博彩特性并不是对所有的投资者具有相同的吸引力。不成熟、非专业的散户投资者具有更强的博彩偏好，具有博彩特性的股票有时被这类投资者视为彩票的替代品（Gao和Lin, 2015）。专业的机构和个人投资者的博彩偏好相对较弱。这类投资者更为关注投资的长期表现而非短期中的博彩收益。因此，博彩类股票在长期中的糟糕市场表现较大程度上促使他们规避持有这类股票。一系列基于经纪人账户数据（Kumar, 2009; Kumar et al., 2011）以及基于股市分笔数据（Han和Kumar, 2013; Lin和Liu, 2018）的研究支持上述观点。考虑到散户和专业投资者之间存在的显著区别，因此有必要专门研究散户偏好对股市造成的影响。但是，研究中国股市的散户偏好并非易事。一方面，经纪人账户数据并非公开数据，并不容易获得。另一方面，由于拆单行为的存在，导致无法辨别小额交易来源于机构投资者还是散户。导致基于分笔数据的研究方法有其重大缺陷所在，更何况中国股市分笔数据覆盖时间范围短。因此，如何通过可获得的股市数据构建合理的散户偏好指标成为研究中面临的一大问题。

一种方法是考察股票的散户持股比例。如果散户偏好某只股票，该股票的散户持股比例自然越高，而机构投资者持股比例则相应越低。通过机构投资者持股比例数据可以反推出散户持股比例。但是，这种方法也有其缺陷所在：第一，机构投资者持股比例数据会遗漏部分

小机构投资者持股以及专业的个人投资者持股；第二，该数据并未覆盖本文考察的完整样本时间范围；第三，该数据公布频率较低，一般以季度或者半年度为公布频率。为了应对以上问题，本文进一步利用文献中提及的以下八个散户所偏好的股票特征——低股价（Kumar, 2009）、高特质波动率（Kumar, 2009）、低市值（Barber和Odean, 2000; Gompers和Metrick, 2001; Gao和Lin, 2015）、低赢利能力（Gao和Lin, 2015）、高市场贝塔（Barber和Odean, 2000）、高账面市值比（Barber和Odean, 2000）、高换手率（Barber和Odean, 2000）以及不分发红利（Graham and Kumar, 2006）。借鉴Stambaugh et al.（2015）的方法，本文结合机构投资者持股比例数据以及散户偏好的股票特征数据构建了个股的散户偏好指标。

基于该指标，本文研究散户偏好与股市博彩溢价之间的关系。第一，基于双变量分组方法以及公司层面截面回归方法，本文证明散户偏好越强的股票中博彩溢价越强。即使对于部分在全样本下不具有显著定价能力的博彩变量，在其中散户偏好较强的股票中也能够表现出显著的定价能力。第二，基于子样本分析的方法，本文逐步剔除散户偏好最强的股票，剩余股票中博彩溢价随之逐步变弱。该研究结果表明即使对于在全样本下具有显著定价能力的博彩变量，其所对应的真正具有显著定价能力的股票所占的市值比重不大。第三，散户投资者比专业投资者更容易受到市场表现波动的影响。因此，如果市场中的博彩溢价主要集中在散户所偏好的股票中，那么在市场表现良好时期博彩溢价更为明显。这一观点得到了本文实证结果的有力支撑。

本文接下来的结构安排如下。第二部分提出散户偏好指标以及博彩变量构建方法，并且说明数据来源。第三部分和第四部分分别利用双变量分组方法以及公司层面截面回归方法探讨散户偏好与股市博彩溢价之间的关系。第五部分采用子样本分析方法进一步研究上述问题。第六部分从市场表现良好时期和市场表现糟糕时期博彩溢价的强弱差异进一步证明易受市场表现影响的散户在股市博彩溢价中所起到的重要作用。第七部分为稳健性检验。第八部分总结全文。

二、变量构建与数据来源

（一）散户偏好指标构建

衡量散户偏好的一种直接的测度方法是考察股票的散户持股比例。如果散户偏好某只股票，该股票的散户持股比例自然越高。与之相对应，专业的机构和个人投资者持股比例则相应越低。因此，利用机构投资者持股比例数据，我们可以大致反推出散户持股比例。

但是，这种做法有其缺陷所在：第一，机构投资者持股比例数据会遗漏小机构投资者持股，并且没有包含专业的个人投资者持股¹。第二，该数据起始时间为2001年6月，并未覆盖本文考察的完整样本期。第三，该数据公布频率较低。2004年6月之前该数据主要来源于中期报告和年度报告，因此一只股票一年内只包含两个机构投资者持股比例数据。2004年6月之后该数据来源还包括第一季报和第三季报，因此一年内包含四个数据。为了覆盖所有的月份，2004年6月之前本文将数据公布月份前面两个月以及后面三个月的该数据设定为等于数据公布月的数值；2004年6月之后本文将数据公布月份前后各一个月的该数据设定为等于数据公布月的数值。尽管这种做法使得该数据覆盖更多的月份，但是该数据的时变性依然不如每月公布一次的数据。

为了应对以上问题，本文进一步采用散户所偏好的股票特征来刻画散户偏好。Kumar（2009）指出低股价和高特质波动率是散户所偏好的两种股票特征。本文将股票在月内交易日收盘价的平均值定义为该股票在该月的价格。同Kumar（2009）一致，本文基于过去六个月的日度数据计算特质波动率。Barber和Odean（2000）、Gompers和Metrick（2001）以及Gao

¹ 为了应对未包含专业的个人投资者持股问题，本文在未汇报的结果中考察了使用前十大股东持股比例指标替代机构投资者持股比例指标的情形。两个指标下的结论非常一致。如有需要，可向作者索取。

和Lin（2015）等研究指出散户偏好低市值股票。本文将股票在月内每个交易日结束时的市值的平均值定义为该股票在该月的市值。Gao和Lin（2015）表明散户更偏好赢利能力较差的股票。为了获得稳健的股票赢利能力指标，本文考察了三种指标——每股收益、净资产收益率以及每股营业利润。本文分别按照这三个指标对截面股票进行排序并确定其所处的百分位位置。赢利指标的百分位位置越高，则意味该股票的赢利能力越强。然后本文将这三种排序结果的平均值定义为股票赢利能力指标。Barber和Odean（2000）指出散户偏好高市场贝塔和高账面市值比股票。本文基于过去60个月的月度数据估计市场贝塔，而账面市值比数据则由股票市净率数据的倒数获得。Barber和Odean（2000）的研究表明散户投资者的注意力容易被股票的高交易量所吸引从而参与其中。本文将高换手率指标定义为月内最高的日度换手率除以过去12个月内的平均日度换手率。Graham and Kumar（2006）指出散户总体来说更偏好无红利股票。本文将过去12个月内无分发红利行为的股票定义为无红利股票。

基于九个反映散户偏好的指标——低机构投资者持股比例、低股价、高特质波动率、低市值、低赢利能力、高市场贝塔、高账面市值比、高换手率以及不分发红利，本文借鉴Stambaugh et al.（2015）的方法构建散户偏好指标。具体来说，本文分别按照这九个指标对截面股票进行排序并确定其所处的百分位位置。本文要求排序后的百分位位置越高，对应散户投资者越偏好该股票。最后，本文将这九种排序结果的平均值定义为散户偏好指标。需要指出的是，在构建个股的散户偏好指标时，本文要求以上九个指标中至少有六个指标存在数据。但是，本文未汇报的结果表明，放松这一假设并不会对本文的结论造成影响。此外，由于担忧以上九个变量中的某一个变量会对本文结论造成决定性的影响。因此，本文逐一考察去掉任意一个变量然后构建散户偏好指标的情形。相应的结果表明去掉任意一个变量都不会改变本文的结论。

（二）博彩变量构建

Bali et al.（2011）将一个月內最高若干个日度收益率的平均值作为博彩偏好的反映指标（天数可取值为1天到5天）。参照这种做法，本文中MAX的构建方法如下：对于任意一只股票，找出该股票在该月的日度收益率最高的3个交易日，然后对这3个交易日的日度收益率取平均。另一类博彩指标为收益率偏度。借鉴Kumar（2009）以及Bali et al.（2011）的做法，本文构建总偏度和特质偏度。基于过去12个月的日度数据，本文将股票超额收益率的标准差定义为总偏度。类似地，基于过去12个月的日度数据，本文使用市场超额收益率对个股超额收益率进行回归，并将残差的标准差定义为该股票的特质偏度²。本文使用MAX、总波动率和特质波动率这三个博彩变量的原因有以下两点：第一，这三个变量在文献中最为常用；第二，这三个变量本身的股票定价能力强弱有所差异。因此，如果本文的结论对于这三个存在定价能力差异的变量都成立，这一点能够很大程度上支持本文的观点。

（三）数据来源

本文中股票日度数据来源于锐思数据库。本文考察的时间区间为2000年到2017年。由于本文需要使用过去五年的数据来估计市场贝塔，所以数据的实际起始年份为1995年。本文稳健性检验部分构建已实现偏度所利用的高频数据来源于国泰安高频数据库。由于该数据起始年份为2007年，所以该变量的样本时间范围为2007年到2017年。本文高频数据的抽样频率为五分钟。而本文数据抽样频率为五分钟的选择是基于满足收敛要求与减少市场微观噪音之间的平衡。同时，这种抽样频率也是该领域文献中最为常用的选择（Hansen 和 Lunde，2006；Bandi 和 Russell，2005）。

² 在未汇报的结果中，本文同样考虑了如下两种情形：解释变量为市场超额收益率以及其平方项（Harvey 和 Siddique，2000）以及解释变量为Fama-French三因子（Ang et al., 2006）。在这两种情形下，本文得到了较为一致的结论。如有需要，可向作者索取。

三、双变量分组

在分析散户偏好与股市博彩溢价之间的关系之前，本文先通过单变量分组的方法检验博彩变量的股票定价能力。该方法依据博彩变量将股票分成 5 个组合，然后观察这些组合未来一个月收益率表现。组合 1 对应博彩特征最弱的股票组合，组合 5 对应博彩特征最强的股票组合。博彩变量为 MAX、总偏度和特质偏度对应的结果见表 1 面板 A、面板 B 以及面板 C 左右部分的第 1 列。其中左侧部分对应等权重结果，右侧部分对应市值加权结果。

表 1：非独立的双变量分组结果

面板 A：MAX												
	等权重						市值加权					
	全样本	1	2	3	4	5	全样本	1	2	3	4	5
1	68.93	40.74	57.82	81.78	93.61	98.06	31.48	15.15	19.67	52.56	63.47	56.39
2	86.52	69.58	77.81	77.57	87.94	81.14	34.05	53.31	23.16	20.82	49.57	54.10
3	54.80	81.08	48.37	42.90	47.34	28.00	9.40	41.36	15.19	1.94	4.98	4.73
4	10.33	52.75	26.07	6.55	-5.40	-53.37	-2.96	32.41	-4.51	-22.09	-32.71	-81.87
5	-88.06	4.38	-43.96	-76.00	-103.62	-152.78	-96.98	-20.24	-53.48	-96.98	-114.52	-171.22
5-1	-156.99	-36.35	-101.78	-157.77	-197.23	-250.84	-128.46	-35.38	-73.15	-149.53	-177.99	-227.61
	(-7.62)	(-1.05)	(-4.11)	(-7.11)	(-8.13)	(-11.79)	(-4.41)	(-0.85)	(-2.11)	(-5.37)	(-5.96)	(-9.13)
三因子	-147.56	-10.33	-71.19	-135.43	-172.67	-232.81	-128.82	-12.32	-58.39	-138.90	-161.57	-213.41
	(-7.18)	(-0.29)	(-3.36)	(-6.66)	(-6.55)	(-8.88)	(-4.24)	(-0.28)	(-1.68)	(-5.52)	(-5.34)	(-7.69)
四因子	-148.91	-12.10	-72.79	-136.51	-173.58	-233.43	-130.31	-13.05	-60.06	-140.98	-163.55	-213.65
	(-7.28)	(-0.35)	(-3.43)	(-6.68)	(-6.58)	(-9.09)	(-4.44)	(-0.31)	(-1.85)	(-5.67)	(-5.76)	(-7.91)
五因子	-149.94	-17.31	-71.58	-129.78	-173.19	-231.62	-118.80	-8.72	-51.78	-130.43	-158.01	-209.57
	(-7.19)	(-0.47)	(-3.40)	(-6.02)	(-6.07)	(-8.87)	(-3.76)	(-0.19)	(-1.48)	(-5.08)	(-5.25)	(-7.26)

面板 B：总波动率												
	等权重						市值加权					
	全样本	1	2	3	4	5	全样本	1	2	3	4	5
1	40.59	56.62	44.98	49.45	38.10	15.41	2.24	12.93	9.99	23.03	9.45	-3.49
2	25.18	47.79	36.65	33.50	13.91	3.48	0.86	11.60	-6.01	-2.90	-30.07	-21.34
3	12.74	45.36	26.86	6.38	11.84	-32.66	-25.58	-0.04	-9.29	-28.48	-13.84	-54.85
4	-4.01	49.27	11.75	-2.02	-14.75	-34.79	-17.40	32.20	-24.76	-24.06	-58.12	-63.68
5	-16.82	24.31	-7.94	-25.67	-38.25	-69.18	-10.77	23.09	-10.86	-52.59	-54.14	-88.36
5-1	-57.41	-32.31	-52.92	-75.12	-76.35	-84.59	-13.00	10.16	-20.85	-75.62	-63.59	-84.88
	(-3.11)	(-1.54)	(-2.92)	(-3.65)	(-3.27)	(-4.58)	(-0.62)	(0.36)	(-1.45)	(-2.11)	(-2.19)	(-3.47)
三因子	-36.91	-19.46	-37.12	-62.11	-61.66	-76.87	19.39	28.86	0.76	-59.36	-46.58	-77.69
	(-2.94)	(-0.96)	(-2.07)	(-4.89)	(-3.55)	(-4.13)	(1.24)	(1.16)	(0.03)	(-1.92)	(-2.16)	(-3.74)
四因子	-37.10	-19.27	-36.71	-62.32	-62.23	-77.09	19.13	29.81	-0.07	-59.28	-48.01	-77.82
	(-2.89)	(-0.96)	(-1.95)	(-4.89)	(-3.44)	(-4.15)	(1.22)	(1.20)	(0.00)	(-1.91)	(-2.28)	(-3.70)
五因子	-36.74	-22.45	-32.84	-56.11	-59.51	-79.71	10.63	25.81	-8.15	-54.87	-47.89	-77.24
	(-2.90)	(-1.11)	(-1.84)	(-4.17)	(-3.60)	(-4.12)	(0.69)	(1.02)	(-0.38)	(-1.85)	(-2.23)	(-3.44)

面板 C：特质波动率												
	等权重						市值加权					
	全样本	1	2	3	4	5	全样本	1	2	3	4	5
1	23.35	37.87	31.21	20.31	12.45	6.59	-19.24	4.89	-15.53	-38.31	-19.25	-28.46

2	18.08	45.37	18.40	20.23	-9.14	-24.79	-1.72	17.68	2.81	-8.45	-50.92	-50.89
3	10.07	44.93	27.67	-4.96	1.74	-26.28	-10.72	6.62	12.37	-44.76	-42.15	-59.71
4	7.29	65.54	10.50	17.64	9.50	-29.44	-14.69	58.37	-16.20	-1.82	-26.34	-32.98
5	-1.10	29.19	23.84	8.29	-4.21	-43.96	-11.61	-1.46	-12.15	-28.39	-23.90	-66.34
5-1	-24.45	-8.68	-7.37	-12.01	-16.66	-50.56	7.63	-6.35	3.38	9.91	-4.65	-37.88
	(-1.55)	(-0.49)	(-0.38)	(-0.75)	(-0.79)	(-2.71)	(0.37)	(-0.25)	(0.17)	(0.35)	(-0.19)	(-1.71)
三因子	-42.47	-28.22	-18.45	-18.86	-24.67	-59.51	-22.56	-36.56	-1.32	-0.11	-8.39	-52.00
	(-3.32)	(-1.39)	(-1.08)	(-1.08)	(-1.43)	(-3.93)	(-1.12)	(-1.29)	(-0.05)	(0.00)	(-0.47)	(-2.98)
四因子	-42.72	-28.56	-18.06	-18.87	-24.84	-59.70	-22.82	-36.02	-1.70	0.09	-8.40	-53.01
	(-3.36)	(-1.42)	(-1.03)	(-1.12)	(-1.44)	(-3.92)	(-1.12)	(-1.29)	(-0.07)	(0.00)	(-0.46)	(-3.04)
五因子	-38.14	-16.21	-6.67	-14.93	-22.61	-54.48	-13.94	-20.71	-2.01	1.72	-13.33	-47.45
	(-3.01)	(-0.76)	(-0.39)	(-0.84)	(-1.28)	(-3.37)	(-0.66)	(-0.68)	(-0.08)	(0.05)	(-0.67)	(-3.06)

注：括号中显示的是 Newey-West t 统计量。下同。

三个博彩变量中 MAX 的股票定价能力最为显著。无论使用等权重还是市值加权，组合 1 到组合 5 的收益率都总体上呈现出下降趋势。高减低组合收益率（第 5 行减去第 1 行）都在 1%水平上显著。经过 Fama-French 三因子（Fama 和 French, 1992; Fama 和 French, 1993; 杨旸和陈展辉, 2003）、Carhart 四因子（Jegadeesh 和 Titman, 1993; Carhart, 1997; 周琳杰, 2002; 朱战宇等, 2003; 王志强等, 2006）和进一步包含流动性因子的五因子（Amihud, 2002; Pástor 和 Stambaugh 2003; 吴文锋等, 2003）调整后，MAX 依然表现出显著的股票定价能力。在全样本下，总偏度的股票定价能力仅存在于等权重情形。在该情形下组合 1 到组合 5 的收益率呈现出单调递减趋势，并且高减低组合收益率在 1%水平上显著。但是，该变量的股票定价能力在市值加权情形下消失。特质偏度的股票定价能力在三者之中最为糟糕。在等权重下，基于特质偏度分组得到的高减低组合收益率在统计上不显著。尽管经过三因子、四因子和五因子调整后，相应的高减低组合收益率在统计上出现显著性，但是这种定价能力再次在市值加权情形下消失。

在以上研究的基础上，本文进一步使用双变量分组方法研究散户偏好与股市博彩溢价之间的关系。本文先将股票按散户偏好指标（控制变量）分成 5 组，然后再将这 5 个股票组合按博彩变量（关注变量）各自分成 5 组，最终得到 5×5 个组合。博彩变量分别为 MAX、总偏度和特质偏度的双变量分组结果分别见表 1 面板 A、面板 B 以及面板 C 左右两侧第 2 到 6 列。其中第 2 列对应散户偏好最弱的股票组合，而第 6 列对应散户偏好最强的股票组合。

首先考察博彩变量为 MAX 的情形。在等权重情形下，第 2 列到第 6 列的高减低组合收益率绝对值呈现出单调递增的趋势，并且组合之间存在着明显的差异。散户偏好最弱的股票中（第 2 列）高减低组合收益率为-36.35 基点，而散户偏好最强的股票中（第 6 列）高减低组合收益率则为-250.84 基点（对应年化收益率高达-30.10%）。后者数值约为前者 7 倍。两者之间的巨大差异同时在 Newey-West t 统计量上明显体现。前者仅为-1.05，意味着散户偏好最弱的股票组合中没有显著的博彩溢价。与之相对应的是，散户偏好最强的股票组合对应的 Newey-West t 统计量高达-11.79。此外，经过三因子、四因子和五因子调整情形后，散户偏好水平不同所导致的博彩溢价差异依然非常明显。同时，这一现象也保留在市值加权情形中。在市值加权下的散户偏好最弱的股票组合中，高减低组合收益率为-35.38 基点，且在统计上不显著；而散户偏好最强的股票组合中，高减低组合收益率为-227.61 基点，对应的 Newey-West t 统计量远超 1%显著水平的临界值。

接着研究的博彩变量为总偏度。在等权重情形下，高减低组合收益率绝对值同样随着散户偏好的上升而上升。散户偏好最弱的股票中高减低组合收益率为-32.31 基点，且不在统计

上显著：散户偏好最强的股票中高减低组合收益率为-84.59 基点，且在 1%水平上显著。因此两者之间在经济上和统计上都存在显著差异。这种差异经过因子调整后依然存在。更为有趣的现象是，尽管在全样本中，总偏度变量在市值加权情形下并未体现出显著的股票定价能力，然而在散户偏好较强的三个股票组合中，该变量却具有显著的股票定价能力。其中散户偏好最强的股票组合中博彩溢价最高，高减低组合收益率为-84.88 基点，对应的 Newey-West t 统计量-3.47。经过因子调整之后，该类股票高减低组合收益率依然在 1%水平上显著。

最后关注博彩变量为特质偏度的情形。上文单变量分组结果表明，特质偏度在全样本下并不具有显著的股票定价能力。但是，基于双变量分组的结果进一步表明，这种缺乏博彩溢价的现象并不是对所有股票都成立。在等权重的情形下，尽管其他的股票组合中特质偏度都不具有显著的定价能力，但是在散户偏好最强的股票组合中，高减低组合收益率为-50.56，且在 1%水平上显著。这一现象经过三因子、四因子和五因子调整后依然得到完整保留。在市值加权的情形下，散户偏好最强的股票组合中特质偏度依然始终具有较高的且显著的博彩溢价。

以上部分所使用的是非独立的双变量分组方法。这种方法的一个缺陷在于，如果关注变量和控制变量之间存在较高的相关性，那么在基于关注变量分组得到的组合中，控制变量水平仍可能存在一定差异。为了应对这个问题，本文进一步采用独立的双变量分组方法。这种方法独立地将控制变量和关注变量各自分成 5 组，最终形成 5×5 个组合。这种方法也有其缺陷所在——如果控制变量和关注变量之间相关性较高，部分组合中所包含的股票个数可能非常少。但是，如果两种方法下的结论一致，那么这将进一步加强我们对本文结论的信心。博彩变量为 MAX、总波动率以及特质波动率的独立的双变量分组结果分别见表 2 面板 A、面板 B 以及面板 C。

表 2：独立的双变量分组结果
面板 A：MAX

	等权重						市值加权					
	全样本	1	2	3	4	5	全样本	1	2	3	4	5
1	68.93	54.46	57.07	74.97	92.03	94.05	31.48	33.83	18.13	50.89	57.89	58.37
2	86.52	74.50	79.11	80.50	97.05	99.74	34.05	39.95	34.80	7.69	48.97	70.11
3	54.80	72.67	42.14	41.12	63.15	56.24	9.40	26.12	-1.20	8.56	27.79	25.21
4	10.33	30.16	16.97	13.03	5.26	-6.38	-2.96	6.41	-7.00	-13.95	-23.22	-38.26
5	-88.06	-21.19	-50.14	-84.34	-100.82	-124.93	-96.98	-46.98	-59.88	-105.59	-114.52	-143.03
5-1	-156.99	-75.65	-107.21	-159.31	-192.85	-218.98	-128.46	-80.82	-78.00	-156.48	-172.41	-201.40
	(-7.62)	(-2.28)	(-4.14)	(-7.35)	(-7.95)	(-10.05)	(-4.41)	(-1.81)	(-2.13)	(-5.70)	(-6.04)	(-7.97)
三因子	-147.56	-47.44	-77.36	-135.64	-169.02	-204.26	-128.82	-55.41	-63.71	-139.69	-158.60	-189.18
	(-7.18)	(-1.33)	(-3.34)	(-6.85)	(-6.59)	(-8.13)	(-4.24)	(-1.24)	(-1.75)	(-5.24)	(-5.27)	(-6.89)
四因子	-148.91	-48.50	-79.13	-137.02	-170.00	-203.78	-130.31	-55.62	-66.10	-142.11	-160.38	-188.04
	(-7.28)	(-1.36)	(-3.41)	(-6.66)	(-6.64)	(-8.36)	(-4.44)	(-1.29)	(-1.81)	(-5.38)	(-5.34)	(-6.92)
五因子	-149.94	-50.10	-78.84	-130.25	-168.59	-204.70	-118.80	-46.97	-58.04	-134.01	-153.92	-184.83
	(-7.19)	(-1.26)	(-3.35)	(-6.29)	(-6.36)	(-8.06)	(-3.76)	(-1.05)	(-1.51)	(-4.74)	(-5.03)	(-7.21)

面板 B：总波动率												
	等权重					市值加权						
	全样本	1	2	3	4	5	全样本	1	2	3	4	5
1	40.59	44.43	42.96	52.77	35.46	20.13	2.24	4.66	7.74	25.52	5.52	5.51
2	25.18	69.14	37.04	31.13	10.23	-6.78	0.86	49.21	-0.61	-13.57	-29.62	-34.72
3	12.74	53.08	29.04	7.58	10.39	-25.17	-25.58	-8.33	-5.08	-29.69	-22.24	-46.15

4	-4.01	47.73	14.64	-16.06	-19.70	-52.49	-17.40	15.39	-17.19	-36.24	-54.89	-91.23
5	-16.82	32.81	-8.63	-31.48	-34.08	-71.36	-10.77	32.19	-22.63	-53.22	-32.46	-97.84
5-1	-57.41	-11.62	-51.59	-84.25	-69.54	-91.49	-13.00	27.53	-30.37	-78.74	-37.98	-103.35
	(-3.11)	(-0.49)	(-2.51)	(-3.84)	(-3.09)	(-5.32)	(-0.62)	(0.95)	(-1.49)	(-2.48)	(-1.46)	(-4.62)
三因子	-36.91	1.45	-38.76	-69.13	-56.06	-82.37	19.39	47.08	-11.67	-59.43	-24.67	-95.10
	(-2.94)	(0.07)	(-2.27)	(-5.01)	(-3.10)	(-4.82)	(1.24)	(1.71)	(-0.48)	(-2.30)	(-1.28)	(-4.75)
四因子	-37.10	1.96	-38.08	-69.93	-56.33	-82.54	19.13	47.89	-11.84	-59.88	-25.79	-95.35
	(-2.89)	(0.09)	(-2.23)	(-4.94)	(-3.11)	(-4.75)	(1.22)	(1.73)	(-0.54)	(-2.32)	(-1.29)	(-4.66)
五因子	-36.74	1.24	-34.47	-65.16	-53.01	-84.29	10.63	42.01	-16.69	-56.65	-26.86	-95.88
	(-2.90)	(0.05)	(-2.06)	(-4.63)	(-3.17)	(-4.50)	(0.69)	(1.42)	(-0.78)	(-2.25)	(-1.25)	(-4.29)

面板 C：特质波动率

	等权重						市值加权					
	全样本	1	2	3	4	5	全样本	1	2	3	4	5
1	23.35	41.81	33.84	21.25	11.86	15.09	-19.24	-5.23	-13.28	-40.60	-31.30	-23.83
2	18.08	49.14	20.27	21.66	3.98	-20.05	-1.72	16.74	17.71	-20.95	-20.13	-48.21
3	10.07	51.53	19.20	2.96	-3.04	-32.78	-10.72	20.25	-9.40	-31.43	-47.56	-63.12
4	7.29	53.39	12.35	9.58	5.78	-29.06	-14.69	24.94	-31.35	-17.23	-26.25	-52.15
5	-1.10	28.83	28.42	6.74	-0.51	-45.98	-11.61	23.12	-7.85	-18.30	-24.55	-63.06
5-1	-24.45	-12.98	-5.43	-14.50	-12.36	-61.07	7.63	28.35	5.43	22.30	6.75	-39.23
	(-1.55)	(-0.76)	(-0.28)	(-0.85)	(-0.63)	(-3.43)	(0.37)	(0.89)	(0.25)	(1.12)	(0.28)	(-1.79)
三因子	-42.47	-32.60	-17.27	-20.13	-21.58	-68.91	-22.56	-5.17	-1.41	11.87	0.95	-50.13
	(-3.32)	(-1.73)	(-0.96)	(-1.12)	(-1.22)	(-4.70)	(-1.12)	(-0.18)	(-0.05)	(0.44)	(0.05)	(-2.77)
四因子	-42.72	-32.95	-16.81	-19.84	-21.55	-69.23	-22.82	-5.38	-1.36	11.95	1.17	-50.79
	(-3.36)	(-1.75)	(-0.90)	(-1.17)	(-1.25)	(-4.81)	(-1.12)	(-0.18)	(-0.05)	(0.43)	(0.06)	(-3.02)
五因子	-38.14	-19.07	-7.52	-17.35	-18.81	-64.07	-13.94	12.14	-3.16	8.86	-1.91	-47.89
	(-3.01)	(-0.96)	(-0.40)	(-0.95)	(-1.09)	(-4.77)	(-0.66)	(0.38)	(-0.13)	(0.32)	(-0.10)	(-3.07)

对于 MAX 和总波动率，博彩溢价随着散户偏好的上升而上升，且散户偏好最弱的股票组合和散户偏好最强的股票组合之间在经济上和统计上有着明显的差异。对于全样本下不具有显著定价能力特质波动率，在散户偏好最强的股票组合中该变量表现出显著的博彩溢价。以上证据证明散户偏好对股市博彩溢价的重要影响。

四、公司层面截面回归

上一部分使用双变量分组方法探讨散户偏好与股市博彩溢价之间的关系。这种方法属于非参数方法，其优势在于不要求被解释变量和解释变量符合某个特定的（线性或者非线性的）函数关系。此外，双变量分组方法每次只能控制一个其他变量。为了应对这个问题，这部分进一步引入公司层面截面回归方法。这种方法实质上是由一系列 Fama-MacBeth 截面回归（Fama 和 MacBeth, 1973）组成。

为了体现不同散户偏好股票组合中的博彩溢价，本文构建五个哑变量 $I(j)$ 。其定义如下：

$$I(j) = \begin{cases} 1, & \text{股票 } i \in \text{组合 } j \\ 0, & \text{股票 } i \notin \text{组合 } j \end{cases} \quad (1)$$

当股票处于根据散户偏好排序得到的第 j 个组合时，该哑变量取值为 1，在其余情形下该哑变量取值为 0。与上部分分成 5 个组合相对应，这部分 j 的取值为 1 到 5。通过博彩变量与

$I(j)$ 的交乘项可以考察不同散户偏好水平下的股票是否具有不同的博彩溢价。

对于样本中的每一月，本文对于该月交易的股票进行以下截面回归：

$$r_{i,t+1}^e = \alpha_t + \sum_j \beta_{j,t} X_{i,t} \cdot I(j) + \sum_k \lambda_{k,t} Z_{i,k,t} + \varepsilon_{i,t+1} \quad (2)$$

其中 $r_{i,t+1}^e$ 表示股票 i 在 $t+1$ 月的超额收益率， $X_{i,t}$ 表示在 t 月股票 i 的博彩变量观测值， $Z_{i,k,t}$ 表示在 t 月股票 i 的第 k 个控制变量观测值。本文通过以上回归得到每个月各个变量系数估计值，然后进一步对系数估计值的时间序列进行平均并计算相应的 Newey-West t 统计量。博彩变量分别为 MAX、总波动率和特质波动率的结果见表 3 第 1 到 4 列、第 5 到 8 列以及第 9 到 12 列。

表 3：公司层面截面回归结果

	1	2	3	4	5	6	7	8	9	10	11	12
截距	1.94** (2.55)	1.92** (2.47)	15.17*** (5.16)	-1.94 (-0.70)	-0.03 (-0.03)	-0.04 (-0.06)	6.50** (2.14)	-10.13*** (-3.36)	0.10 (0.13)	0.11 (0.13)	8.43*** (2.75)	-8.54*** (-2.83)
博彩变量	-41.98*** (-10.59)				-0.51*** (-2.81)				-0.03 (-0.28)			
博彩变量 * $I(1)$		-31.74*** (-3.39)	-12.35 (-1.57)	-1.13 (-0.14)		0.06 (0.27)	0.11 (0.56)	0.16 (0.85)		0.21 (1.64)	0.48*** (3.29)	0.36*** (2.78)
博彩变量 * $I(2)$		-37.62*** (-6.70)	-32.03*** (-6.38)	-18.22*** (-3.28)		-0.53* (-1.96)	-0.33 (-1.49)	-0.23 (-1.06)		0.03 (0.23)	0.03 (0.20)	0.03 (0.26)
博彩变量 * $I(3)$		-38.47*** (-7.63)	-38.42*** (-8.79)	-24.59*** (-4.37)		-0.64*** (-2.64)	-0.23 (-1.35)	-0.14 (-0.74)		0.05 (0.40)	-0.14 (-1.68)	-0.11 (-1.16)
博彩变量 * $I(4)$		-41.22*** (-9.73)	-44.24*** (-10.49)	-29.75*** (-5.31)		-0.94*** (-3.57)	-0.51*** (-2.98)	-0.36* (-1.94)		-0.08 (-0.64)	-0.31*** (-3.61)	-0.21*** (-2.61)
博彩变量 * $I(5)$		-45.46*** (-12.79)	-51.40*** (-15.09)	-37.01*** (-7.57)		-1.00*** (-4.35)	-0.54*** (-3.08)	-0.37* (-1.93)		-0.33** (-2.38)	-0.60*** (-5.17)	-0.48*** (-4.25)
市场贝塔			0.29 (0.91)	-0.02 (-0.06)			-0.68** (-2.29)	-0.63** (-2.00)			-0.46 (-1.52)	-0.50 (-1.61)
公司市值			-0.59*** (-4.59)	0.11 (0.86)			-0.22* (-1.71)	0.48*** (3.40)			-0.31** (-2.26)	0.41*** (2.92)
账面市值 比			0.80*** (5.02)	0.82*** (5.54)			0.73*** (4.61)	0.74*** (4.88)			0.82*** (5.55)	0.77*** (5.33)
动量因子				0.13 (0.49)				0.41 (1.33)				0.39 (1.27)
反转因子				-2.98*** (-2.69)				-6.27*** (-6.92)				-6.15*** (-7.05)
低流动性 指标				0.77*** (6.69)				0.90*** (7.56)				0.88*** (7.60)
协偏度				-0.35* (-1.73)				-0.33 (-1.52)				-0.31 (-1.44)
Adj. R^2	0.0202	0.0546	0.1004	0.1320	0.0104	0.0185	0.0817	0.1220	0.0069	0.0179	0.0807	0.1214

注：为了方便显示，表中的系数估计值放大了 100 倍。在回归中，本文对公司市值、账面市值比以及低流动性指标取了自然对数。*、**和***分别表示在 10%、5% 和 1% 显著水平上显著。

首先考察的博彩变量为 MAX。第 1 列对应只包含 MAX 的情形，其结果与上文单变量分组一致，即 MAX 在全样本中具有很强的股票定价能力。本文的目标是研究散户偏好对股

市博彩溢价的影响。因此，第 2 列考察五类不同散户偏好水平股票中 MAX 的定价能力。由此可以很清楚地观察到，MAX 系数估计值的绝对值随着散户偏好的上升而上升。散户偏好最弱的股票中博彩变量系数估计值为-0.3174，而散户偏好最强的股票中系数估计值则上升到-0.4546。两类股票的 Newey-West t 统计量的大小也存在着明显的区别，后者约为前者 4 倍。在控制 Fama-French 三因子情形下，两者之间的差异被进一步放大。相应的结果见第 3 列。散户偏好最弱的股票中博彩变量系数估计值下降到-0.1235，且不再在统计上显著。与之相对应，散户偏好最强的股票中系数估计值则上升到-0.5140，对应的 Newey-West t 统计量也随之上升到-15.09。第 4 列和第 5 列中分别考察进一步加入动量因子 (Jegadeesh 和 Titman, 1993; Carhart, 1997) 以及反转因子 (Lehmann, 1990; Jegadeesh, 1990)、低流动性因子 (Amihud, 2002) 和协偏度因子 (Harvey 和 Siddique, 2000) 的情形。在这两种情形下，以上结论依然得到完整保留。

博彩变量为总偏度的公司层面截面回归结果与基于 MAX 的结果比较相似 (见第 5 到 8 列)。略有不同的是，由于总偏度本身的股票定价能力弱于 MAX，因此可以观察到在更多的低散户偏好股票中总偏度系数估计值不显著。但是，最为关键的是，总偏度系数估计值及其 Newey-West t 统计量的绝对值依然随着散户偏好的上升而上升，从而体现出结论的一致性。

最后关注博彩变量为特质偏度的情形。与单变量分组中的结果一致，由第 9 列可以看到特质偏度在全样本下并不具有股票定价能力。第 10 列考察五类不同散户偏好水平股票中特质偏度的定价能力。其中只有在散户偏好最强的两个股票组合中特质偏度系数估计值为符合经济逻辑的负数，并且只有在散户偏好最强的股票中该系数估计值具有统计上的显著性。在控制 Fama-French 三因子的情形下，散户偏好最强的三个股票组合中特质偏度系数估计值为负数，且其中散户偏好最强的股票组合中该系数估计值在 1% 水平上显著。在进一步控制动量因子、反转因子、低流动性因子以及协偏度因子的情形下，本文亦得到了一致的结论。

以上公司层面截面回归结果再一次表明，散户越偏好的股票中，博彩变量的定价能力越强，散户偏好对股市博彩溢价造成明显的影响。

五、子样本分析

这部分借鉴 Avramov et al. (2007) 的方法，采取子样本分析的方式研究博彩变量在多大比重的股票中具有定价能力及其与散户偏好的关系。进行这一研究的目的是进一步体现散户偏好在博彩溢价中起到的重要作用。

由于散户偏好越强的股票中博彩溢价越显著，因此，如果从样本中逐步删去散户偏好最强的股票，剩余股票中博彩变量的定价能力将逐步削弱。当足够大比重的散户偏好股票被剔除时，剩余股票中博彩变量定价能力将会消失。需要指出的是，由于这种方法只能应用于在全样本下具有定价能力的变量，因此这部分仅考察博彩变量为 MAX 的情形。需要关注的重点是在剔除多大百分比的散户偏好股票之后，博彩变量将失去其股票定价能力。与此同时，本文还关注不具有显著博彩溢价的股票所占的市值比重。这部分使用 5% 的显著水平作为博彩变量是否具有定价能力的判断标准。相应的结果见表 4。

表 4：子样本分析结果

	等权重				市值加权				市值 比重
	5-1	三因子	四因子	五因子	5-1	三因子	四因子	五因子	
10%	-139.88 (-6.58)	-124.76 (-5.95)	-125.89 (-6.05)	-125.20 (-5.85)	-113.23 (-3.81)	-109.91 (-3.53)	-111.19 (-3.71)	-98.29 (-3.12)	96.02%
20%	-127.88 (-5.90)	-109.98 (-5.22)	-111.35 (-5.30)	-109.83 (-5.09)	-107.66 (-3.57)	-102.32 (-3.28)	-103.84 (-3.47)	-89.95 (-2.90)	91.37%

30%	-117.47 (-5.36)	-97.96 (-4.71)	-99.32 (-4.81)	-97.64 (-4.57)	-92.95 (-3.08)	-85.45 (-2.78)	-86.85 (-2.94)	-71.99 (-2.18)	86.28%
40%	-95.40 (-4.39)	-73.67 (-3.48)	-75.14 (-3.53)	-73.05 (-3.28)	-77.70 (-2.50)	-66.93 (-2.04)	-67.78 (-2.16)	-52.13 (-1.57)	80.69%
50%	-80.90 (-3.57)	-57.76 (-2.60)	-59.33 (-2.52)	-57.12 (-2.31)	-65.59 (-2.03)	-51.45 (-1.45)	-52.24 (-1.54)	-36.22 (-1.03)	74.44%
60%	-67.75 (-2.89)	-41.80 (-1.75)	-43.20 (-1.69)	-42.50 (-1.68)	-55.10 (-1.68)	-39.87 (-1.07)	-40.29 (-1.15)	-25.55 (-0.69)	66.79%
70%	-53.05 (-1.93)	-25.86 (-0.95)	-27.30 (-0.97)	-28.08 (-0.97)	-42.21 (-1.23)	-24.43 (-0.64)	-25.00 (-0.68)	-16.26 (-0.40)	57.36%
80%	-34.05 (-0.98)	-7.90 (-0.22)	-9.65 (-0.28)	-15.55 (-0.41)	-35.83 (-0.87)	-13.52 (-0.31)	-14.35 (-0.34)	-10.67 (-0.24)	44.76%
90%	-43.73 (-1.18)	-21.64 (-0.53)	-23.41 (-0.57)	-24.00 (-0.53)	-20.76 (-0.46)	-0.93 (-0.02)	-14.35 (-0.34)	1.71 (0.03)	26.89%

表 4 左半部分对应等权重情形。第 1 列显示一定百分比的散户偏好股票被剔除后，未经其他股票定价因子调整的高减低组合收益率。第 2 列到第 4 列分别表示经过三因子、四因子和五因子调整后的结果。最后 1 列表示剩余股票所占的市值比重。当散户偏好最强的 70% 股票被删去时，未经因子调整的高减低组合收益率不再在统计上显著，即 MAX 的股票定价能力不再在统计上显著。剩余 30% 股票所占的市值比重为 57.36%，这意味着占市值一半以上的股票组合并不具有显著的博彩溢价。这一状况经过其他因子调整后变得更为明显。经过因子调整后，MAX 在剔除散户偏好最强的 60% 股票后变得不再具有显著定价能力，而剩余股票所占的是市值比重约为 2/3。在市值加权的情形下，无博彩溢价的股票占据更大的市值比重。由第 5 行可知在未经其他因子调整情形下，散户偏好较弱的 40% 股票不具有显著的博彩溢价；第 6 和 7 行表明经过三因子和四因子调整后，散户偏好较弱的 50% 股票中 MAX 不具有显著的定价能力；第 8 行表明经过五因子调整后，MAX 的定价能力在剔除前 60% 散户偏好的股票后消失，而这部分股票所占的市值比重高达 80.69%。若以此为判定标准，则意味着占市值 4/5 的股票不具有显著的博彩溢价。尽管从整体来说，MAX 表现出不错的定价能力，但是以上证据表明这种定价能力仅仅来源于占市值比重并不大的散户偏好股票。

六、博彩溢价与股市表现

相对于专业的机构和个人投资者，散户更容易受到市场表现波动的影响。在市场表现良好时期，散户的博彩冲动会被激发出来，而在市场表现糟糕时期，散户的博彩欲望会被其恐慌心理所压制。如果股市的博彩偏好由情绪化的散户主导，那么相对于市场表现糟糕时期，市场表现良好时期博彩变量应该具有更强的博彩溢价。同上部分一致，这部分考察在全样本下显著的博彩变量 MAX。相应的结果见表 5。

表 5：不同市场表现下的博彩溢价

	市场表现良好		市场表现糟糕	
	等权重	市值加权	等权重	市值加权
5-1	-235.40 (-7.72)	-167.49 (-3.83)	-77.84 (-2.35)	-89.06 (-2.06)
三因子	-244.12 (-9.03)	-203.08 (-4.84)	-48.02 (-2.38)	-47.25 (-1.69)
四因子	-243.83	-196.40	-48.55	-51.54

	(-8.66)	(-4.88)	(-2.53)	(-1.82)
五因子	-258.18	-202.14	-45.36	-36.04
	(-8.81)	(-4.61)	(-2.47)	(-1.36)

本文基于市场月度收益率将样本拆分为两个部分。在市场表现良好时期，在等权重（市值加权）情形下，高减低组合收益率为-235.40 基点（-167.49 基点），对应的 Newey-West t 统计量为-7.72（-3.83）。经过三因子、四因子和五因子调整后，高减低组合收益率进一步上升到-243 到-258 基点（-196 到-203 基点）。在市场表现糟糕时期，在等权重（市值加权）情形下，高减低组合收益率大幅度下降到为-77.84 基点（-89.06 基点），对应的 Newey-West t 统计量也随之下降到-2.35 和-2.06。经过三因子、四因子和五因子调整之后，高减低组合收益率进一步下降到-45 到-48 基点（-36 到-51 基点）。两个时期下的巨大差异给出积极证据支持散户在股市博彩溢价方面起到的重要作用。

七、稳健性检验

（一）其他博彩变量

本文正文中所采用的博彩变量为 MAX、总偏度和特质偏度。尽管基于这三个常用的博彩变量本文得到了一致的结论，但是出于稳健性的考虑，这部分进一步采用另外三个博彩变量——已实现偏度、宽度（Breadth）以及尾部偏度。本文借鉴 Amaya et al.（2015）基于过去一个月抽样频率为五分钟的高频数据构建已实现偏度指标。参考 Kapadia（2006）的方法并基于过去 12 个月的日度收益率数据，本文将均值和中位数之间的差值定义为宽度指标。均值与中位数之间的差异越大意味着越有可能出现正的极端值，即具有更强的博彩特性。借鉴 Zhang（2006）的研究，本文基于过去 12 个月的日度收益率数据定义尾部偏度。其表达式如下：

$$\text{尾部偏度} = \frac{(P_{95}-P_{50})-(P_{50}-P_5)}{(P_{95}-P_5)} \quad (3)$$

其中 P_{95} 、 P_{50} 以及 P_5 分别表示 95 分位点、50 分位点（即中位数）以及 5 分位点。该做法利用收益率尾部数据，因此能够较好地反映博彩收益的极端特性。基于这三个变量的结果分别见表 6 面板 A、面板 B 以及面板 C³。

表 6：其他博彩变量的双变量分组结果

面板 A：已实现偏度

	等权重					市值加权				
	1	2	3	4	5	1	2	3	4	5
1	86.83	102.02	107.07	124.74	105.58	53.90	41.50	25.41	70.58	60.61
2	90.27	90.21	89.19	101.98	103.14	58.21	48.71	49.84	67.70	43.53
3	60.84	38.11	62.16	75.17	59.30	11.88	-11.76	2.17	39.65	28.58
4	50.74	44.15	27.43	14.44	-2.46	-12.06	-42.81	-48.42	-38.92	-33.65
5	-15.39	-43.83	-34.69	-74.94	-113.99	-38.60	-75.60	-73.55	-123.29	-158.52
5-1	-102.21	-145.85	-141.76	-199.69	-219.56	-92.50	-117.10	-98.96	-193.87	-219.12
	(-3.67)	(-5.36)	(-5.76)	(-5.40)	(-7.41)	(-2.54)	(-2.85)	(-3.06)	(-5.31)	(-6.35)
三因子	-103.65	-146.99	-135.91	-192.37	-216.71	-95.78	-121.06	-96.91	-187.95	-219.77
	(-3.39)	(-5.75)	(-4.92)	(-4.89)	(-6.78)	(-2.61)	(-3.15)	(-3.07)	(-5.05)	(-6.79)
四因子	-102.38	-142.16	-135.02	-190.56	-216.72	-103.45	-116.09	-99.06	-186.13	-219.25
	(-3.13)	(-5.96)	(-4.94)	(-4.83)	(-6.90)	(-2.76)	(-3.02)	(-3.22)	(-5.07)	(-6.68)

³ 囿于篇幅，本文在稳健性检验部分仅汇报非独立的双变量分组结果。如有需要完整结果，可向作者索取。

五因子	-108.41 (-3.01)	-136.25 (-5.37)	-133.90 (-5.02)	-183.03 (-4.75)	-206.90 (-6.54)	-106.87 (-2.74)	-107.40 (-2.92)	-98.35 (-3.08)	-178.14 (-4.58)	-206.81 (-6.45)
面板 B：宽度										
	等权重					市值加权				
	1	2	3	4	5	1	2	3	4	5
1	61.43	56.11	64.55	66.13	29.53	31.82	24.93	25.90	34.48	7.70
2	57.20	40.66	52.61	51.71	44.39	41.86	10.04	37.96	22.48	23.92
3	41.90	48.38	24.11	34.01	3.50	20.51	38.42	4.24	11.41	-28.68
4	56.77	35.34	22.95	8.92	-16.60	23.30	8.03	-5.98	-19.45	-39.31
5	38.00	-10.15	-25.63	-40.56	-61.94	12.17	-47.48	-52.18	-61.50	-91.65
5-1	-23.43 (-0.89)	-66.26 (-3.57)	-90.17 (-4.92)	-106.69 (-4.67)	-91.47 (-5.43)	-19.64 (-0.65)	-72.41 (-2.69)	-78.08 (-2.81)	-95.97 (-4.04)	-99.35 (-5.17)
三因子	6.32 (0.29)	-44.12 (-2.55)	-61.35 (-5.43)	-84.53 (-4.70)	-67.58 (-3.56)	7.37 (0.30)	-46.91 (-1.71)	-49.00 (-2.35)	-76.84 (-3.40)	-80.91 (-3.95)
四因子	6.24 (0.29)	-45.30 (-2.64)	-61.92 (-5.40)	-84.78 (-4.73)	-67.59 (-3.57)	6.28 (0.25)	-47.96 (-1.76)	-50.51 (-2.51)	-77.04 (-3.65)	-79.95 (-3.87)
五因子	1.54 (0.07)	-47.94 (-2.90)	-67.11 (-5.54)	-86.39 (-4.65)	-72.80 (-3.58)	-2.31 (-0.08)	-58.36 (-2.31)	-61.14 (-2.89)	-78.87 (-3.62)	-86.89 (-3.78)
面板 C：尾部偏度										
	等权重					市值加权				
	1	2	3	4	5	1	2	3	4	5
1	68.02	72.30	69.21	83.07	68.64	48.71	39.55	29.86	46.51	41.72
2	63.41	47.59	56.60	48.43	26.65	47.64	19.68	39.00	18.44	3.78
3	43.96	24.05	24.74	23.38	-2.08	11.17	-1.75	5.88	-6.80	-32.55
4	50.61	31.21	12.80	0.67	-26.63	28.27	14.16	-22.20	-21.91	-53.12
5	29.21	-5.23	-24.79	-35.31	-67.96	6.32	-30.39	-43.08	-55.08	-84.14
5-1	-38.81 (-1.76)	-77.54 (-4.01)	-94.00 (-4.67)	-118.38 (-6.51)	-136.60 (-7.20)	-42.39 (-1.48)	-69.94 (-2.96)	-72.95 (-2.22)	-101.59 (-4.48)	-125.86 (-6.09)
三因子	-9.56 (-0.48)	-48.81 (-3.02)	-61.83 (-4.16)	-97.61 (-5.05)	-114.11 (-5.41)	-15.11 (-0.63)	-39.22 (-1.56)	-33.27 (-1.48)	-86.91 (-4.06)	-108.79 (-4.50)
四因子	-8.00 (-0.42)	-49.04 (-3.04)	-62.52 (-4.03)	-97.94 (-5.12)	-113.82 (-5.02)	-13.68 (-0.55)	-39.73 (-1.55)	-35.47 (-1.63)	-88.13 (-4.08)	-107.59 (-4.63)
五因子	-11.12 (-0.59)	-50.43 (-3.09)	-60.23 (-4.17)	-93.79 (-4.51)	-110.94 (-4.43)	-23.45 (-0.98)	-45.14 (-1.73)	-36.82 (-1.58)	-86.74 (-3.85)	-106.90 (-4.23)

由上表可见，已实现偏度、宽度和尾部偏度的股票定价能力随着散户偏好的上升而上升。散户偏好最弱的股票组合中博彩溢价亦最低，其中宽度和尾部偏度在该组合下甚至不具有显著的股票定价能力。与之相对应，散户偏好最强的股票组合中博彩溢价也最为显著。因此，这部分基于三个新的博彩变量得到的结论与前文一致，从而进一步支持本文的观点。

（二）博彩变量构建方法

前文将一个月内最高 3 个日度收益率平均值定义 MAX，并基于过去 12 月的日度数据计算总偏度和特质偏度。出于稳健性的考虑，这部分采取其他的变量构建方法。MAX 在这部分被定义为过去一个月内最高日度收益率或者最高 5 个日度收益率的平均值，而总偏度

和特质偏度指标则基于过去 6 个月的日度数据或者过去 5 年的月度数据获得。相应的结果见表 7 面板 A、面板 B 以及面板 C⁴。

表 7：博彩变量的其他构建方法

面板 A：MAX										
	MAX (1)					MAX (5)				
	1	2	3	4	5	1	2	3	4	5
1	43.27	56.90	69.86	88.50	86.01	40.48	53.84	71.48	94.07	93.73
2	77.62	66.43	80.62	78.15	77.23	63.67	78.73	85.65	86.60	90.12
3	71.47	45.38	43.85	41.76	9.58	82.45	51.35	55.91	57.99	30.62
4	46.55	26.49	0.02	-17.42	-61.23	60.69	24.22	7.63	-9.23	-54.76
5	9.34	-29.00	-61.47	-71.70	-111.18	1.26	-41.87	-87.42	-109.41	-158.52
5-1	-33.93	-85.90	-131.33	-160.20	-197.20	-39.22	-95.71	-158.90	-203.48	-252.25
	(-1.17)	(-3.99)	(-6.89)	(-7.47)	(-11.23)	(-1.09)	(-4.01)	(-6.83)	(-7.97)	(-11.35)
三因子	-16.82	-66.08	-113.18	-141.99	-190.66	-12.53	-63.94	-136.91	-180.02	-232.95
	(-0.57)	(-4.08)	(-6.64)	(-5.70)	(-11.18)	(-0.36)	(-2.85)	(-6.66)	(-5.72)	(-8.76)
四因子	-19.16	-67.78	-113.98	-142.86	-191.74	-14.22	-65.67	-138.38	-181.74	-234.21
	(-0.63)	(-4.03)	(-6.82)	(-5.69)	(-11.20)	(-0.41)	(-2.72)	(-6.69)	(-5.68)	(-8.81)
五因子	-17.54	-64.59	-107.44	-141.56	-191.61	-18.80	-63.74	-134.32	-182.85	-234.50
	(-0.54)	(-3.63)	(-6.13)	(-5.46)	(-10.64)	(-0.50)	(-2.76)	(-6.15)	(-5.18)	(-8.89)
面板 B：总偏度										
	基于 6 个月日度数据					基于 5 年月度数据				
	1	2	3	4	5	1	2	3	4	5
1	59.90	57.76	59.31	66.74	46.83	58.92	53.54	34.82	53.80	23.88
2	62.98	42.31	42.76	35.11	19.57	56.94	41.63	40.78	46.05	-4.11
3	43.25	41.71	20.69	24.31	-16.45	51.35	32.63	32.18	20.83	29.74
4	52.53	28.45	20.39	4.73	-16.83	13.48	8.45	33.84	17.58	-7.66
5	36.19	-0.15	-4.87	-11.26	-34.82	41.68	2.71	-10.66	-19.84	-29.51
5-1	-23.72	-57.91	-64.19	-78.00	-81.65	-17.24	-50.83	-45.48	-73.64	-53.39
	(-1.02)	(-2.73)	(-2.75)	(-3.96)	(-4.68)	(-0.84)	(-2.46)	(-1.84)	(-4.24)	(-3.21)
三因子	-4.21	-40.24	-41.83	-62.73	-68.67	-16.44	-43.30	-27.14	-60.82	-35.37
	(-0.19)	(-2.03)	(-2.26)	(-4.15)	(-3.92)	(-0.83)	(-2.30)	(-1.46)	(-3.79)	(-2.37)
四因子	-4.06	-41.29	-42.74	-63.53	-69.55	-12.76	-42.29	-26.13	-61.33	-34.99
	(-0.18)	(-2.13)	(-2.34)	(-4.01)	(-4.07)	(-0.66)	(-2.20)	(-1.40)	(-3.46)	(-2.35)
五因子	-6.41	-43.37	-42.83	-64.00	-68.39	-20.59	-35.05	-22.73	-54.16	-31.34
	(-0.29)	(-2.32)	(-2.37)	(-4.10)	(-4.11)	(-1.10)	(-1.79)	(-1.17)	(-2.97)	(-2.02)
面板 C：特质偏度										
	基于 6 个月日度数据					基于 5 年月度数据				
	1	2	3	4	5	1	2	3	4	5
1	40.67	34.35	27.23	25.53	19.61	37.87	31.21	20.31	12.45	6.59
2	56.03	35.34	17.93	12.22	-13.61	45.37	18.40	20.23	-9.14	-24.79
3	58.34	24.60	17.55	17.40	-25.40	44.93	27.67	-4.96	1.74	-26.28

⁴ 囿于篇幅，本部分只汇报等权重下的结果。未汇报的市值加权结果与等权重结果非常相似。如有需要，可向作者索取。

4	47.55	26.02	17.38	3.61	-16.07	65.54	10.50	17.64	9.50	-29.44
5	24.33	13.43	10.90	1.74	-30.16	29.19	23.84	8.29	-4.21	-43.96
5-1	-16.34	-20.91	-16.33	-23.79	-49.77	-8.68	-7.37	-12.01	-16.66	-50.56
	(-0.85)	(-1.01)	(-0.89)	(-1.19)	(-2.75)	(-0.49)	(-0.38)	(-0.75)	(-0.79)	(-2.71)
三因子	-34.01	-33.76	-21.14	-33.03	-54.83	-28.22	-18.45	-18.86	-24.67	-59.51
	(-2.02)	(-1.71)	(-1.24)	(-1.94)	(-3.73)	(-1.39)	(-1.08)	(-1.08)	(-1.43)	(-3.93)
四因子	-34.46	-33.20	-21.18	-33.66	-54.77	-28.56	-18.06	-18.87	-24.84	-59.70
	(-2.05)	(-1.72)	(-1.27)	(-1.98)	(-3.88)	(-1.42)	(-1.03)	(-1.12)	(-1.44)	(-3.92)
五因子	-24.98	-22.25	-18.59	-28.84	-49.22	-16.21	-6.67	-14.93	-22.61	-54.48
	(-1.30)	(-1.15)	(-1.08)	(-1.71)	(-3.59)	(-0.76)	(-0.39)	(-0.84)	(-1.28)	(-3.37)

面板 A 左侧和右侧分别对应使用月内最高日度收益率以及最高 5 个日度收益率的平均值定义 MAX 的情形。面板 B 和面板 C 左侧和右侧对应基于过去 6 个月日度数据以及基于过去 5 年月度数据计算总偏度和特质偏度的情形。将这三个表格中的结果与前文定义下的双变量分组结果相对比,不难发现变量不同变量构建方式下的结果非常一致。这表明本文的结论在不同的变量构建方式下依然稳健。

八、总结

本文基于机构投资者持股比例以及一系列散户偏好的股票特征构建散户偏好指标,并基于此研究散户偏好与股市博彩溢价之间的关系。本文的主要结论有:

(1) 本文基于双变量分组以及公司层面截面回归的结果表明,散户偏好越强的股票中博彩溢价越高。即使对于在全样本下不具有显著定价能力的博彩变量,在散户偏好最强的股票中该变量也能够表现出显著的定价能力。

(2) 本文基于子样本分析的方法研究具有博彩溢价的股票所占的市值比重。从样本中逐步剔除散户偏好股票,剩余样本中博彩溢价逐步下降直至消失。这部分研究结果表明,即使对于在全样本下具有定价能力的博彩变量,其所对应的具有博彩溢价的股票所占的市值比重不大。

(3) 散户投资者比专业投资者更容易受到市场表现波动的影响。因此,如果市场中的博彩偏好主要集中在散户上,那么在不同市场表现下博彩溢价应该会有明显差异。本文基于单变量分组的结果表明,市场表现良好时期博彩溢价远远高于市场表现糟糕时期。

参考文献

- [1] 江曙霞, 陈青. 赌博特征股票的收益预测及解释[J]. 财贸研究, 2013, 24(3): 99-107.
- [2] 孔东民, 代昀昊, 李捷瑜. 知情交易与中国股市博彩溢价[J]. 金融评论, 2010, 2(2): 61-72.
- [3] 李培馨, 刘悦, 王宝链. 中国股票市场的赌博行为研究[J]. 财贸经济, 2014, 35(3): 68-79.
- [4] 吴文锋, 芮萌, 陈工孟. 中国股票收益的非流动性补偿[J]. 世界经济, 2003(7): 54-60.
- [5] 王志强, 齐佩金, 孙刚. 动量效应的最新研究进展[J]. 世界经济, 2006(2): 82-92.
- [6] 杨炘, 陈展辉. 中国股市三因子资产定价模型实证研究[J]. 数量经济技术经济研究, 2003, 20(12): 137-141.
- [7] 周琳杰. 中国股票市场动量策略赢利性研究[J]. 世界经济, 2002(8): 60-64.
- [8] 朱战宇, 吴冲锋, 王承炜. 不同检验周期下中国股市价格动量的盈利性研究[J]. 世界经济, 2003(8): 62-67.
- [9] 郑振龙, 孙清泉. 彩票类股票交易行为分析:来自中国 A 股市场的证据[J]. 经济研究, 2013(5): 128-140.

- [10] 郑振龙, 王磊, 王路跖. 特质偏度是否被定价?[J]. 管理科学学报, 2013, 16(5): 1-12.
- [11] An L., Wang H., Wang J. and Yu J., Lottery-related anomalies: the role of reference-dependent preferences[J]. Globalization & Monetary Policy Institute Working Paper, 2015.
- [12] Ang A., Hodrick R. J., Xing Y, and Zhang X., The cross-section of volatility and expected returns[J]. The Journal of Finance, 2006, 61(1): 259-299.
- [13] Avramov D., Chordia T., Jostova G., and Philipov A., Momentum and credit rating[J]. The Journal of Finance, 2007, 62(5): 2503-2520.
- [14] Bali T. G., Cakici N., and Whitelaw R. F., Maxing out: stocks as lotteries and the cross-section of expected returns [J]. Journal of Financial Economics, 2008, 99(2):427-446.
- [15] Bandi F. M., Russell J. R., Microstructure noise, realized volatility, and optimal sampling[R]. Working paper, University of Chicago Graduate School of Business, 2005.
- [16] Barber B. M., and 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.
- [17] Baber B. M., and Odean T., All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors[J]. The Review of Financial Studies, 2008, 21(2):785-818.
- [18] Barberis N., and Huang M., Stocks as Lotteries: The implications of probability weighting for security prices[J]. American Economic Review, 2008, 98(5):2066-2100.
- [19] Boyer B., Mitton T., and Vorkink K., Expected idiosyncratic skewness[J]. Review of Financial Studies, 2010, 23(1):169-202.
- [20] Brunnermeier M. K., Christian G., and Parker J. A., Optimal beliefs, asset prices, and the preference for skewed returns[J]. The American Economic Review, 2007, 97(2):159-165.
- [21] Carhart M. M., On persistence in mutual fund performance[J]. The Journal of finance, 1997, 52(1): 57-82.
- [22] Fama E. F., and French K. R., Common risk factors in the returns on stocks and bonds[J]. Journal of financial economics, 1993, 33(1): 3-56.
- [23] Fama E. F., and French K. R., The cross-section of expected stock returns[J]. the Journal of Finance, 1992, 47(2): 427-465.
- [24] Fama E. F., and MacBeth J. D., Risk, return, and equilibrium: empirical tests[J]. Journal of Political Economy, 1973, 81(3): 607-636.
- [25] Gao X., and Lin T. C., Do individual investors treat trading as a fun and exciting gambling activity? evidence from repeated natural experiments[J]. The Review of Financial Studies, 2014, 28(7): 2128-2166.
- [26] Gompers P. A., and Metrick A., Institutional investors and equity prices[J]. The Quarterly Journal of Economics, 2001, 116(1): 229-259.
- [27] Graham J. R., and Kumar A., Do dividend clienteles exist? evidence on dividend preferences of retail investors[J]. Journal of Finance, 2006, 61(3):1305-1336.
- [28] Han B., and Kumar A., Speculative retail trading and asset prices[J]. Social Science Electronic Publishing, 2013, 48(2):377-404.
- [29] Hansen P. R., and Lunde A., Realized variance and market microstructure noise[J]. Journal of Business & Economic Statistics, 2006, 24(2): 127-161.
- [30] Harvey C. R., and Siddique A., Conditional skewness in asset pricing tests[J]. The Journal of Finance, 2000, 55(3): 1263-1295.

- [31] Jegadeesh N., and Titman S., Returns to buying winners and selling losers: implications for stock market efficiency[J]. *Journal of Finance*, 1993, 48(1):65-91.
- [32] Jegadeesh N., Evidence of predictable behavior of security returns[J]. *The Journal of finance*, 1990, 45(3): 881-898.
- [33] Kapadia N., *The next microsoft? skewness, idiosyncratic volatility, and expected returns*[J]. Social Science Electronic Publishing, 2007.
- [34] Kumar A., Page J. K., and Spalt O. G., Religious beliefs, gambling attitudes, and financial market outcomes[J]. *Journal of Financial Economics*, 2011, 102(3):671-708.
- [35] Kumar A., Retail investor sentiment and return comovements[J]. *Journal of Finance*, 2006, 61(5):2451-2486.
- [36] Kumar A., Who gambles in the stock market?[J]. Social Science Electronic Publishing, 2009, 64(4):1889-1933.
- [37] Lehmann B. N., Fads, martingales, and market efficiency[J]. *The Quarterly Journal of Economics*, 1990, 105(1): 1-28.
- [38] Lin T. C., and Liu X., Skewness, individual investor preference, and the cross-section of stock returns[J]. *Review of Finance*, 2018, 22 (5): 1841-1876.
- [39] Mitton T., and Vorkink K., Equilibrium underdiversification and the preference for skewness[J]. *Review of Financial Studies*, 2007, 20(4):1255-1288.
- [40] Nofsinger J. R., *The psychology of investing*[M]. 6th edition, Routledge, 2017.
- [41] Brunnermeier M. K., Gollier C., and Parker J. A., Optimal beliefs, asset prices, and the preference for skewed returns[J]. *The American Economic Review*, 2007, 97(2): 159-165.
- [42] Pástor L., and Stambaugh R. F., Liquidity risk and expected stock returns[J]. *Journal of Political Economy*, 2003, 111(3): 642-685.
- [43] Stambaugh R. F., Yu J., and Yuan Y., Arbitrage asymmetry and the idiosyncratic volatility puzzle[J]. *The Journal of Finance*, 2015, 70(5): 1903-1948.
- [44] Zhang Y., Individual skewness and the cross-section of average stock returns[J]. Working Paper, Yale University, 2005.