



中央财经大学

Central University of Finance and Economics

论文标题: Investor Sentiment Aligned: A Powerful Predictor of
Stock Returns 复刻报告

任课教师: 姜富伟

学年学期: 2023-2024 学年第一学期

课程名称: 金融计量学

小组成员: 董心诣 (2021310447) (组长)

郭玖俐 (2021310360)

谢妤晴 (2021310341)

邓羨韵 (2021310352)

祝奕忱 (2021310309)

课程分数:

目录

摘要	4
1. Econometric Methodology	5
2. Data.....	5
3. Empirical Results.....	7
3.1. Forecasting the market.....	7
3.2. Comparison with economic predictors	8
3.3. Out-of-sample forecasts.....	11
3.4. Predictability with longer horizons.....	14
3.5. Asset allocation implications.....	15
3.6. Forecasting characteristics portfolios	16
4. Economic Explanations	18
4.1. Cash flow and discount rate predictability	18
4.2. The cross-section of cash-flow channel.....	19
4.3. Market-volatility risk.....	20
参考文献:	22

图表目录

表 2-1: 变量名称及含义	6
表 2-2: 变量的描述性统计	7
表 3-1: 回归结果	8
表 3-2: 12 个经济预测因子	8
表 3-3: 12 个经济预测因子描述性统计	9
表 3-4: 单变量回归结果	9
表 3-5: 双变量回归结果	10
表 3-6: 第一步预测评估期的结果	12
表 3-7: 第二步预测评估期的结果	12
表 3-8: 长期预测结果	14
表 3-9: $\gamma=1$ 时,四个指数的预测结果	15
表 3-10: $\gamma=3$ 时,四个指数的预测结果	16
表 3-11: $\gamma=5$ 时,四个指数的预测结果	16
表 3-12: 在不同分组下对回报的预测结果	17
表 4-1: 渠道检验结果	19
表 4-2: S^{PLS} 对现金流量横截面的预测能力	19
表 4-3: 对于 10 个规模投资组合的回归结果	20
表 4-4: S^{PLS} 对市场波动的预测能力	21
图 2-1: 两种投资者情绪指数对比	7

摘要

本篇论文提出了一个新的投资者情绪指数，它与预测股票市场目的相一致。通过消除情感代理中常见的噪声成分，新的指数比现有的情感指数在样本内和样本外都有更大的预测能力，并且可预测性变得既具有统计意义又具有经济意义。此外，它优于公认的宏观经济变量，也可以预测按行业、规模、价值和动量排序的横截面股票收益。预测能力的驱动力来源于投资者对未来现金流的有偏信念。

1. Econometric Methodology

假设投资者情绪解释的提前一期预期超额股票收益率服从标准线性关系。

$$E_t(R_{t+1}) = \alpha + \beta S_t$$

其中 S_t 是真实但不可观测的投资者情绪，对预测资产收益很重要。那么已实现的股票收益等于其条件期望加上一个不可预测的冲击。

$$R_{t+1} = E_t(R_{t+1}) + \varepsilon_{t+1} = \alpha + \beta S_t + \varepsilon_{t+1}$$

令 $x_t = (x_{1,t}, \dots, x_{N,t})'$ 表示 t 期个体投资者情绪代理变量($t = 1, \dots, T$)的一个 $N \times 1$ 向量。在 Baker 和 Wurgler (2006 年、2007 年)中， x_t 是封闭式基金折价率、股票换手率、首次公开发行股票(Initial Public Offerings, IPOs)数量、IPO 首日回报率、股息溢价和新股权益份额。我们假设 $x_{i,t}$ ($i = 1, \dots, N$)具有因子结构。

采用偏最小二乘(PLS)方法有效地提取了 S_t ，并过滤掉了不相关的成分 E_t ，而主成分方法无法保证这样做。PLS 等价于两步 OLS 回归，得到情绪指标：

$$x_{i,t-1} = \pi_{i,0} + \pi_i R_t + \mu_{i,t-1}, \quad t = 1, \dots, T.$$

$$x_{i,t} = c_t + S_t^{\text{PLS}} \hat{\pi}_i + v_{i,t}, \quad i = 1, \dots, N,$$

由于 π_i 是未知的，第一阶段的回归斜率提供了 $x_{i,t}$ 如何依赖于 S_t^{PLS} 的初步估计。也就是说，PLS 使用时间 $t + 1$ 的股票收益来约束降维，以提取与预测相关的 S_t ，并抛弃与预测无关的共同和特殊成分，如 E_t 和 $e_{i,t}$ 。

2. Data

主要变量及其含义如下表：

表 2-1：变量名称及含义

变量名称	含义
市场收益率	S&P500 对数超额收益率
Close-end fund discount rate (CEFD)	流通市值加权封闭式股票型共同基金份额的净资产 价值与其市场价格的平均差额
Number of IPOs (NIPO)	首次公开发行股票月数
First-day returns of IPOs (RIPO)	首次公开发行的月度平均首日收益率
Dividend premium (PDND)	股利支付者和非股利支付者的流通市值加权平均市 账比对数之差
Equity share in new issues (EQTI)	每月股本总额除以每月股本总额加发债额
BW 指数	对 CEFD、NIPO、RIPO、PDND 和 EQTI 5 个情绪 代理变量进行主成分提取得到

在数据来源方面：

(1) 无风险收益率和 S&P 500 月度指数数据来自 Robert J. Shiller 主页，并按照如下的公式计算得到月度超额收益率：

$$R = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

$$R_m = R - R_f$$

(2) 5 个原始情绪代理变量数据来自 Wurgler 主页。数据区间：1965 年 7 月至 2022 年 6 月（共 683 条）。值得注意的是，在 Baker 和 Wurgle 的文章中使用了 6 个原始变量，比这里多 1 个变量：TURN，股票的换手率。但是由于机构高频交易导致换手率的含义出现了变化，Wurgler 将其舍弃。

在投资者情绪指标构建方面，考虑构建以下指数：BW 指数 S^{BW} 按照 Baker 和 Wurgle 的论文由主成分分析得到；一致的投资者情绪指数 S^{PLS} ，参考(Huang 等, 2015)的论文，由偏最小二乘法（PLS），即如下的两步最小二乘法公式回归得到：

$$x_{i,t-1} = \pi_{i,0} + \pi_i R_t + u_{i,t-1} \quad t = 1, \dots, T$$

$$x_{i,t} = c_t + S_t^{PLS} \hat{\pi}_i + v_{i,t}, \quad i = 1, \dots, N$$

PLS 回归的结果如下：

$$S^{PLS} = -0.24CEFD + 0.47NIPO + 0.76RIPO - 0.44PDND + 0.54EQTI$$

在对原始变量进行预处理之后，我们进行了全样本的 PLS 回归。这里的预处理遵循文章（平滑 6 期，对宏观变量正交，标准化）。灰色区域表示经济衰退期。两个指标对比和描述性统计如下所示：

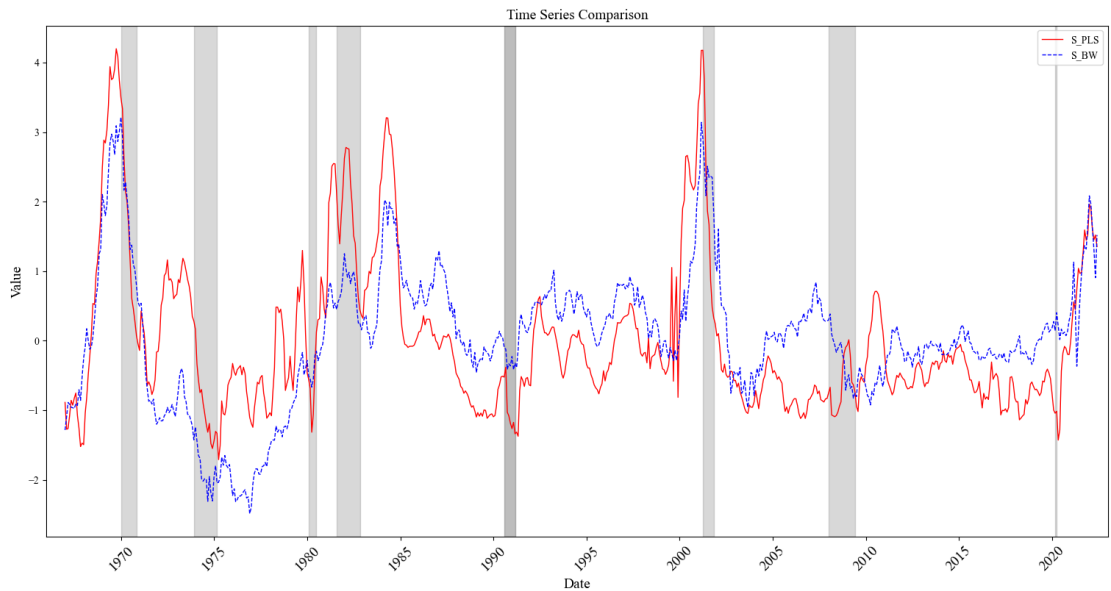


图 2-1：两种投资者情绪指数对比

表 2-2：变量的描述性统计

	Mean	Std	Skew	Kurt	Min	Max	$\rho(1)$	SR
R^m (%)	0.22	3.69	-1.22	5.00	-22.88	11.34	0.24	0.06
R^f (%)	0.37	0.28	0.61	0.42	0.00	1.35	0.98	
S^{PLS}	0.00	1.00	1.58	2.36	-1.71	4.20	0.97	
S^{BW}	0.00	1.00	0.19	1.13	-2.49	3.21	0.98	

其中，SR 表示夏普比率。 S^{BW} 和 S^{PLS} 的 Pearson 相关系数为 0.66。

3. Empirical Results

3.1. Forecasting the market

为检验情绪指标是否对下一期超额收益率有预测能力，我们进行了按照如下公式作为基准回归进行检验：

$$R_{t+1}^m = \alpha + \beta S_t^k + \epsilon_{t+1}, \quad k = PLS, BW, EW$$

回归结果如下：

表 3-1：回归结果

	β	t-stat	$R^2(\%)$	$R_{up}^2(\%)$	$R_{down}^2(\%)$	$R_{high}^2(\%)$	$R_{low}^2(\%)$
S_{BW}	-0.24	-1.65	0.40	0.45	0.30	0.83	0.14
S_{EW}	-0.58*	-1.71	0.40	0.67	-0.06	0.51	0.40
S_{PLS}	-0.44***	-3.47	1.80	1.59	2.19	3.25	0.87
$S_{PLS(bias-free)}$	-0.43***	-2.71	1.40	1.42	1.32	1.65	1.24
ripo	0.06	0.31	0.00	-0.13	0.34	-0.06	0.06
nipo	-0.05	-0.30	0.00	0.01	0.03	0.09	-0.03
cefd	0.14	0.97	0.00	-0.09	0.65	0.30	0.04
Kitchen Sink			0.02	1.86	4.98	3.89	2.18

其中， S^{EW} 表示对 5 个原始代理变量进行等权重加权指标， $S^{PLS(bias-free)}$ 表示通过迭代回归去除数据窥探问题后的指标，Kitchen Sink 代表把所有原始代理变量对收益率进行多元线性回归。结果显示， S^{BW} 的回归系数不显著，这与 Baker 和 Wurgler 的论文结果一致。经过改进后的 S^{PLS} 和 $S^{PLS(bias-free)}$ 回归系数显著，说明投资者情绪指标对股票超额收益率具有预测能力。

3.2. Comparison with economic predictors

本小节中我们挑选了 12 个经济预测因子与 S^{PLS} 进行预测能力的比较。我们作出如下假设：

H1: S^{PLS} 的单变量预测能力优于大多数现有经济指标。

H2: 在控制了其他经济指标之后， S^{PLS} 仍然具有增量预测能力。

12 个经济预测因子的数据来源于 Welch 和 Goyal (2008)，选取 1967 年 1 月至 2022 年 5 月的月度数据。各经济预测因子的变量名、统计量含义以及描述性统计如下表 1 和表 2 所示。

表 3-2：12 个经济预测因子

变量名	统计量
DP	Dividend-price ratio
EP	Earnings-price ratio
DE	Dividend-payout ratio
SAVR(%)	Stock return variance
BM	Book-to-market ratio
NTIS	Net equity expansion
TBL(%)	Treasury bill rate
LTY(%)	Long-term yield
LTR(%)	Long-term return
TMS(%)	Term spread

	DFY(%)			Default yield spread			
	INFL(%)			Inflation			
表 3-3：12 个经济预测因子描述性统计							
	Mean	Std	Skew	Kurt	Min	Max	$\rho(1)$
DP	-3.65	0.42	0.06	-0.99	-4.52	-2.75	0.99
EP	-2.87	0.45	-0.51	2.12	-4.84	-1.90	0.99
DE	-0.78	0.31	3.00	16.35	-1.24	1.38	0.99
SAVR(%)	0.24	0.52	9.90	117.96	0.02	7.32	0.39
BM	0.48	0.27	0.88	-0.32	0.12	1.21	1.00
NTIS	0.01	0.02	-0.43	-0.19	-0.06	0.05	0.98
TBL(%)	4.50	3.36	0.61	0.40	0.01	16.30	0.99
LTY(%)	6.37	2.94	0.40	-0.12	0.62	14.82	0.99
LTR(%)	0.61	3.08	0.35	2.03	-11.24	15.23	0.06
TMS(%)	1.86	1.46	-0.39	-0.10	-3.65	4.55	0.96
DFY(%)	1.06	0.43	1.91	4.84	0.55	3.38	0.96
INFL(%)	0.33	0.37	-0.07	2.98	-1.92	1.81	0.60

首先，我们考虑对单一经济变量的预测回归。分别用 12 个因子对市场超额收益率进行预测，方程如下：

$$R_{t+1}^m = \alpha + \psi Z_t^k + \epsilon_{t+1}, \quad k = 1, \dots, 12. \quad (14)$$

其中， R_{t+1}^m 是第 $t+1$ 期的市场超额收益率， Z_t^k 是第 t 期的 12 个经济预测因子之一。表 3 中前十二列报告了的回归结果。根据回归结果可知，在 12 个因子中，仅有 SAVR、TBL、LTY、LTR、INFL 五个因子在 5% 以上的显著性水平上对市场有显著的预测能力。而在上述五个显著的因子中，仅有一个因子（LTR）的 R^2 大于 S^{PLS} 。因此可以说明在对市场的预测能力方面， S^{PLS} 优于除 LTR 外的 11 个经济预测指标，H1 成立。

表 3-4：单变量回归结果			
Panel A: Univariate predictive regressions			
$R_{t+1}^m = \alpha + \psi Z_t^k + \epsilon_{t+1}$			
	ψ	t-stat	$R^2(\%)$
DP	-0.42	-1.22	0.23
EP	-0.52	-1.62	0.40
DE	0.32	0.69	0.07
SAVR	-0.95***	-3.47	1.78
BM	-0.92	-1.73	0.45
NTIS	-9.18	-1.31	0.26
TBL	-0.13**	-3.06	1.39
LTY	-0.12*	-2.44	0.89

LTR	0.19***	4.09	2.47
TMS	0.21	2.12	0.67
DFY	0.43	1.29	0.25
INFL	-1.35***	-3.48	1.79
ECON ^{PC}	-0.16*	-2.37	0.84
ECON ^{PLS}	0.54***	5.74	4.74
(S + ECON) ^{PC}	-0.16*	-2.53	0.96
(S + ECON) ^{PLS}	0.74***	7.85	8.51

*, **和***分别表示在 10%、5%和 1%水平上的显著性

接着，我们研究在控制经济预测因子后， S^{PLS} 的预测能力是否仍然显著。为了分析 S^{PLS} 的增量预测能力，我们将 S^{PLS} 分别与 12 个因子进行双变量预测回归，方程如下：

$$R_{t+1}^m = \alpha + \beta S_t^{PLS} + \psi Z_t^k + \epsilon_{t+1}, \quad k = 1, \dots, 12 \quad (15)$$

表 4 报告了方程（15）的回归结果。根据回归结果可知，方程（15）中 β 的估计值为负且较大，与方程（12）的预测回归结果一致。更重要的是，当加入经济预测因子之后，方程（15）中的所有 R^2 都远大于方程（14）中的 R^2 ， β 仍然具有统计学意义。以上结果表明， S^{PLS} 包含了大量的互补预测信息，具有增量预测能力，H2 成立。

表 3-5: 双变量回归结果

Panel B: Bivariate predictive regressions					
$R_{t+1}^m = \alpha + \beta S_t^{PLS} + \psi Z_t^k + \epsilon_{t+1}$					
	β	t-stat	ψ	t-stat	$R^2(\%)$
DP	-0.43***	-3.33	-0.32	-0.94	1.87
EP	-0.42***	-3.29	-0.42	-1.33	2.00
DE	-0.44***	-3.41	0.29	0.65	1.80
SAVR	-0.45***	-3.56	-0.97***	-3.60	3.62
BM	-0.41**	-3.18	-0.64	-1.20	1.95
NTIS	-0.43**	-3.16	-1.01	-0.14	1.74
TBL	-0.34*	-2.51	-0.09*	-1.99	2.32
LTY	-0.38**	-2.81	-0.07	-1.46	2.05
LTR	-0.42***	-3.31	0.18***	4.00	4.05
TMS	-0.40**	-3.08	0.15	1.52	2.08
DFY	-0.45***	-3.53	0.51	1.55	2.09
INFL	-0.39**	-3.09	-1.22**	-3.15	3.19
ECON ^{PC}	-0.38**	-2.93	-0.11	-1.58	2.10
ECON ^{PLS}	-0.25	-1.95	0.48***	4.98	5.28

*, **和***分别表示在 10%、5%和 1%水平上的显著性

在本小节中我们的复刻结果和原文结果在定量上有所出入，原因可能有以下两点：第一，

我们拓展了数据样本，原文样本周期为 1965 年 7 月至 2010 年 12 月，而本小节我们选取的样本周期为 1967 年 1 月至 2022 年 5 月；第二，本小节所用的经济预测因子的数据来源于 Goyal 学者网页，在计算方法上可能和原文有所不同。但尽管存在差别，我们的复刻结果依然可以从定性上证明原文的结论，即 S^{PLS} 的单变量预测能力优于大多数经济指标，且在控制了其他经济指标之后， S^{PLS} 具有增量预测能力。

最后我们研究的问题是，当 PLS 和 PC 应用于所有经济变量或将经济变量与 Baker 和 Wurgler (2006) 代理变量相结合时，它们的表现如何。 $ECON^{PC}$ 是从 12 个经济变量中提取的第一个主成分 (PC) 因子。 $ECON^{PLS}$ 是基于 12 个经济变量提取的 PLS 预测因子。 $(S + ECON)^{PC}$ 和 $(S + ECON)^{PLS}$ 是分别应用 PC 方法和 PLS 方法对 12 个经济变量和 5 个个人情绪代理变量的联合所提取的两个预测因子。表 3 的最后四行报告了单变量回归结果，表 4 的最后两行报告了双变量回归结果。根据回归结果可知， $ECON^{PC}$ 的 R^2 远小于 $ECON^{PLS}$ 的 4.74%，且显著水平较低。将所有经济变量与情绪代理变量相结合， $(S + ECON)^{PLS}$ 的样本内 R^2 为 8.51%，并且在 1% 水平下显著，而 $(S + ECON)^{PC}$ 的 R^2 再次较小，且显著水平较低。结合先前的结论我们可以得到，PLS 在大多数情况下都优于 PC，并且在预测因子组合起来时也具有更大的 R^2 。

在这一步骤中我们的复刻结果和原文依然有定量上的差别，原因除了上述两条之外，还可能包括插值法填补空缺数据所导致的误差、用于提取预测因子的变量较原文减少以及在执行 PC 方法和 PLS 方法过程中的偏差等等，但依旧不影响我们对原文定性结论的证明，即应用 PLS 方法提取到的预测因子相比于 PC 方法具有更强的显著性和更大的 R^2 。

3.3. Out-of-sample forecasts

在本节，我们进行样本外检测。这么做的原因主要有二：1. 根据已有文献，样本外测试似乎更适合实时评估真正的回报可预测性，并避免样本内过拟合问题。2. 样本外测试收到的小样本大小扭曲的影响，比如 Stambaugh 偏差和 PLS 方法的前瞻性偏差关注，要小得多。

我们使用 1966:12 至 1984:12 的数据作为初始估计期，因此预测评估期为 1985:01 至 2010:12 以及加入最新数据后的 1985:01 至 2022:10。

第一步：预测样本外的股票收益 R 。关键要求是在这个时间序列的数据里，只能使用到 t 的可用信息来预测 $t+1$ 时的股票收益。先选择一个固定的数字 p ，作为初始样本训练的终点，从而有 $T-p$ 个样本外预测。具体来说，我们用初始估计期的实际市场超额收益与各个情

绪指数，得到不同情绪指数对应的 $\hat{\alpha}$ 与 $\hat{\beta}$ 。接着在两个预测评估期，估计超额市场收益。

第二步：进行三个显著性检验。它们分别是 R_{OS}^2 统计量、DM 检验、CW 检验。特别地，我们在第一个显著性检验中也分别计算了经济扩张期和经济衰退期的 R_{OS}^2 。

两个预测评估期的结果如表所示。

表 3-6：第一步预测评估期的结果

	$R_{OS}^2(\%)$	DM-test	CW-test	$R_{OS,up}^2(\%)$	$R_{OS,down}^2(\%)$
Panel A: Investor sentiment indexes					
S^{BW}	-1.57	-1.16	-0.12	-7.11	4.84
S^{EW}	-1.34	-0.76	0.64	-1.61	-4.29
S^{Com}	0.93	0.36	2.01**	-5.64	9.26
S^{PLS}	1.32	1.34	1.74**	-1.49	2.98
Panel B: Individual investor sentiment proxies					
RIPO	1.56	0.62	1.86**	-5.59	10.94
NIPO	-0.61	-0.62	0.03	-5.34	4.39
CEFD	-1.93	-1.58	-0.46	-7.36	4.26
Kitchen Sink	4.56	1.74*	4.42***	1.22	7.86

表 3-7：第二步预测评估期的结果

	$R_{OS}^2(\%)$	DM-test	CW-test	$R_{OS,up}^2(\%)$	$R_{OS,down}^2(\%)$
Panel A: Investor sentiment indexes					
S^{BW}	-3.00	-2.43**	-0.99	-10.41	6.18
S^{EW}	-1.00	-0.70	0.82	-2.54	-2.58
S^{Com}	0.00	0.00	2.06**	-8.92	12.10
S^{PLS}	0.45	0.56	1.19	-3.78	3.86

Panel B: Individual investor sentiment proxies					
RIPO	0.67	0.29	1.91**	-8.30	12.89
NIPO	-1.94	-2.08**	-1.01	-8.48	5.65
CEFD	-3.32	-3.01**	-1.48	-10.56	5.51
Kitchen Sink	5.99	2.83**	5.69***	2.55	8.20

预测评估期为 1985:01 至 2010:12 的结果显示，三个用于对比的情绪指数中，只有组合预测 S^{Com} 产生正 R_{OS}^2 统计量（0.93%），因此，其 MSFE 值低于历史平均值。然而，根据 DM 测试，这种优越性在统计上并不显著。因此， S^{Com} 对总体股市的样本外预测能力较弱。而 S^{BW} 、 S^{EW} 在三个显著性检验中均表现不显著，说明这两个情绪指标对总体股市的样本外预测能力很弱。

相比之下， S^{PLS} 对总体市场的样本外预测能力要强得多。其 R_{OS}^2 为 1.32%，大大超过其他预测方法的所有 R_{OS}^2 。 S^{PLS} 的 DM 检验统计量和 CW 检验统计量分别为 1.34 和 1.74，DM 检测的结果明显优于三个对比的情绪指数，而 CW 检验的结果表明 S^{PLS} 的 MSFE 在 5% 的显著水平上明显小于历史平均值。

为了方便与 3.1 节的结果进行比较，我们在该表格的 B 面板显示了三个情绪代用指标 RIPO、NIPO 与 CEFD 的样本外表现。三个变量中只有 RIPO 产生了正的 R_{OS}^2 统计量，为 1.56。而根据 DM 检验，RIPO 的预测能力在样本外也不显著。

我们还对 Kitchen-sink 模型进行了递归估计，并对其样本外性能进行了评估。面板 B 底行报告了结果。厨房水槽模型产生了 4.56% 的正 R_{OS}^2 ，并且在 DM 检验和 CW 检验中分别在 10% 和 1% 的水平上表现显著。通常情况下，厨房水槽在预测因子较多时表现较差。而在本文中，只有 5 个预测因子，因此其表现较好。

与原论文结果相比，复刻结果仍有所差异，尤其是在 $R_{OS,up}^2$ 、 $R_{OS,down}^2$ 的显著性上。我们分析主要原因有三：1. 缺失 TURN 后，情绪指数的预测能力受到影响。2. 进行回归、预测的初始样本为 1966:12-1984:12，相较原文的 1965:07-1984:12，缺失了 1965:07-1966:11 一年多的数据，影响了样本外的股票收益 \hat{R} 的构建。3. 在构建 S^{PLS} 指标、搜集 RIPO 等情绪代理变量的数据处理过程中，存在数据缺失等问题，进行数据填充时，应当严格按照“在时间序列的数据里，只能使用到 t 的可用信息来预测 $t+1$ 时的股票收益”这一要求，且其中存在许多处

理细节问题。

加入了更新数据的预测评估期 1985:01 至 2022:10 的结果，与原文产生较大差异，尤其是 S^{PLS} 的显著性与预测能力。我们分析原因有以下几点：1. 2010 年至 2022 年期间，互联网的应用、环境产生了巨大变化，平台的多样化、社群的丰富化等现象，使得投资者的情绪代理变量需要更多改进与修正，才能更准确地显示投资者情绪。2. 最近几年，国际政治经济环境动荡，大国经济政策给股票市场带来巨大变化，所以需要与时俱进的新理论成果，来支撑投资者情绪指数的构建。

3.4. Predictability with longer horizons

前文主要检测投资者情绪在一个月内的可预测性，但在本小节中，我们研究其在更长期限内的预测能力。具体来说，即 3、6、9、12 与 24 个月内的预测能力。下表报告了 S^{PLS} 对长期超额市场回报率的样本内和样本外预测结果。为便于比较，我们还显示了 BW 指数的结果。通过比较可以发现，首先， S^{PLS} 可以预测 12 个月以内的长期超额市场回报率。预测能力随着时间跨度的增加而先增加，然后下降，在样本内和都如此。更具体地说，样本内预测能力在 12 个月时达到峰值，样本外预测能力在 9 个月时达到峰值。其次， S^{PLS} 对长期超额市场回报率的预测能力远远超过 S^{BW} 。

表 3-8：长期预测结果

Horizon	S^{PLS}				S^{BW}			
	β	t-stat	R^2 (%)	$R^2_{OS}(\%)$	β	t-stat	$R^2(\%)$	$R^2_{OS}(\%)$
1-month	-0.44***	-3.47	1.80	1.32	-0.24	-1.65	0.40	-1.57
3-month	-1.33***	-5.31	4.08	2.64	-0.59**	-2.03	0.62	-3.10
6-month	-2.64***	-7.09	7.09	3.68	-1.12**	-2.54	0.97	-5.04
9-month	-3.60***	-7.71	8.32	3.40	-1.50***	-2.69	1.09	-6.76
12-month	-4.37***	-7.98	8.88	2.56	-1.57**	-2.39	0.87	-8.38
24-month	-4.87***	-6.11	5.51	-19.32	0.91	0.96	0.14	-12.62

3.5. Asset allocation implications

我们使用 1982 年 1 月—2022 年 5 月共 485 个月的数据进行实证检验。使用每月超额市场收益率作为资产收益,根据历史 5 年数据计算每月收益率的方差。利用递归估计方法获得基于投资者情绪指数 SPLS 的每月超额收益率预测值。即每月仅使用过去可获得的信息进行回归,获得当月的预测系数。按照均值-方差模型,结合风险厌恶系数、收益率预测和方差,可以动态计算得到每月持股权重。限制权重在 0 到 1.5 之间。计算资产组合每月实际收益。基于预测收益率进行资产配置的投资者收益率高于基于历史平均收益率的投资者。分别考虑无交易成本和 50 个基点交易成本两种情况。计算在风险厌恶系数为 1、3、5 时的投资组合的夏普比率并进行显著性检验。

使用的方程为:

$$\text{在 } t \text{ 期末, 投资者进行最优配置: } \omega_t = \frac{1}{\gamma} \frac{\hat{R}_{t+1}^{ms}}{\hat{\sigma}_{t+1}^2}$$

$$t+1 \text{ 实现的投资组合收益: } R_{t+1}^p = \omega_t R_{t+1}^{ms} + R_{t+1}^f$$

$$\text{投资组合收益的夏普比率为: } \text{Sharpe ratio} = \frac{E(R_p - R_f)}{\sigma_p}$$

$\gamma=1$ 时,四个指数 SPLS、SBW、SEW 和 SCOM 预测的夏普比率都高于历史平均水平。这说明在低风险厌恶条件下,基于投资者情绪的预测效果较好,与原文结论一致。其中 SPLS 的预测夏普比率最高预测效果最好。

表 3-9: $\gamma=1$ 时,四个指数的预测结果

Predictor	No transaction cost		50pbs transaction cost	
	Sharpe ratio	P-value	Sharpe ratio	P-value
S ^{PLS}	0.29	0.14	0.28	0.11
S ^{BW}	0.23	0.89	0.23	0.94
S ^{EW}	0.17	0.35	0.17	0.58
S ^{COM}	0.28	0.27	0.28	0.76

$\gamma=3$ 时,中等风险厌恶条件下,部分指数如 SPLS 预测效果仍较强,与原文结论一致。SPLS 和 SBW 预测夏普比率仍高于历史平均值,而 SEW 和 SCOM 略低。这说明中等风险厌恶条件下,部分指数如 SPLS 预测效果仍较强,与原文结论一致。但其余投资者情绪指数预测效果弱于平均值,这一点与原文结论不同,可能是受样本期的影响。

表 3-10: $\gamma=3$ 时,四个指数的预测结果

Predictor	No transaction cost		50pbs transaction cost	
	Sharpe ratio	P-value	Sharpe ratio	P-value
S^{PLS}	0.25	0.01	0.25	0.06
S^{BW}	0.23	0.77	0.22	0.84
S^{EW}	0.15	0.86	0.15	0.72
S^{COM}	0.16	0.21	0.13	0.13

$\gamma=5$ 时,四个指数预测的夏普比率都接近历史平均值。说明高风险厌恶条件下,情绪指数预测效果较弱,与原文结论一致。这可能因为高风险厌恶投资者追求稳定,对情绪指数预测反应不明显。

表 3-11: $\gamma=5$ 时,四个指数的预测结果

Predictor	No transaction cost		50pbs transaction cost	
	Sharpe ratio	P-value	Sharpe ratio	P-value
S^{PLS}	0.18	0.03	0.18	0.05
S^{BW}	0.14	0.10	0.14	0.09
S^{EW}	0.11	0.81	0.10	0.64
S^{COM}	0.25	0.06	0.24	0.08

总体来说,随着投资者风险厌恶系数的增加,夏普比率会下降。这是因为更风险厌恶的投资者会要求更高的风险准备金,从而降低了调整后的期望收益。相对于 S^{BW} 和 S^{EW} 等其他情绪指数, S^{PLS} 具有更好的预测能力。这些结论和原文基本一致。部分与原论文结果不一致的可能原因:基于数据的可得性,我们分析的样本期可能与原文使用的样本期有所不同。

3.6. Forecasting characteristics portfolios

在这一小节中,我们考虑投资者情绪在不同的行业、规模、账面市值比、动量分组下对于回报的预测是否存在差异。因此,我们将股票依次按照 industry、size、B/M、momentum 分别分成十组,以探究 S^{PLS} 对于不同的投资组合的预测能力。回归方程如下:

$$R_{t+1}^j = \alpha_j + \beta_j S_t^{PLS} + \epsilon_{t+1}^j$$

回归结果如下表所示。总体上看, S^{PLS} 较 S^{BW} 具有更高的回报预测能力, S^{PLS} 所有的 R^2

几乎都比 S^{BW} 对应的 R^2 更高，且两者的回归斜率几乎均为负。具体而言，在行业层面，不同行业之间的回归斜率存在较明显差异，我们的回归结果与论文基本一致，除了能源行业的显著性较原文下降了以外，制造业、技术、电信行业的可预测性仍较高。在规模层面，回归结果基本显著，且斜率的绝对值随着公司规模变小而单调增加的现象仍然成立。在市值层面，结果与原文稍有不同，投资者情绪对于个别投资组合的预测能力较差，可能的原因是小组采取了与原论文不同的数据选取方法。而在动量层面，回归结果也是基本显著的。

表 3-12：在不同分组下对回报的预测结果

	$S^{PLS}(\%)$	t -stat	$R^2(\%)$	$S^{BW}(\%)$	t -stat	$R^2(\%)$
Panel A: Industry portfolios						
Nondurable	-0.18	-1.18	0.21	-0.01	-0.08	0.00
Durable	-0.22	-0.89	0.12	-0.16	-0.57	0.05
Manufacture	-0.35**	-1.99	0.60	-0.22	-1.10	0.18
Energy	-0.11	-0.50	0.04	-0.13	-0.54	0.04
Technology	-0.68***	-3.04	1.37	-0.63**	-2.47	0.91
Telecom	-0.35**	-2.16	0.70	-0.38**	-2.01	0.61
Shop	-0.33*	-1.82	0.50	-0.11	-0.54	0.04
Health	-0.15	-0.91	0.12	-0.01	-0.06	0.00
Utility	-0.07	-0.51	0.04	-0.06	-0.37	0.02
Other	-0.29	-1.54	0.36	-0.24	-1.12	0.19
Panel B: Size portfolios						
Lo 10	-0.62***	-2.79	1.16	-0.72***	-2.80	1.17
2-Dec	-0.52**	-2.32	0.80	-0.63**	-2.43	0.88
3-Dec	-0.52**	-2.41	0.87	-0.54**	-2.21	0.73
4-Dec	-0.50**	-2.44	0.89	-0.57**	-2.42	0.88
5-Dec	-0.47**	-2.38	0.84	-0.49**	-2.14	0.68
6-Dec	-0.51***	-2.72	1.10	-0.50**	-2.34	0.82
7-Dec	-0.45**	-2.44	0.89	-0.42**	-1.98	0.59
8-Dec	-0.35*	-1.95	0.57	-0.34*	-1.69	0.43
9-Dec	-0.33**	-2.05	0.63	-0.25	-1.32	0.26
Hi 10	-0.36**	-1.79	0.80	-0.22	-1.31	0.26
Panel C: Book-to-market portfolios						
BM1	-0.85***	-3.67	1.99	-0.37*	-1.82	0.49
BM2	-0.60***	-2.92	1.27	-0.25	-1.34	0.27
BM3	-0.48**	-2.48	0.92	-0.24	-1.28	0.25
BM4	-0.39**	-2.10	0.66	-0.28	-1.46	0.32
BM5	-0.32*	-1.81	0.49	-0.22	-1.23	0.23
BM6	-0.26	-1.49	0.33	-0.19	-1.01	0.15
BM7	-0.24	-1.38	0.29	-0.23	-1.24	0.23
BM8	-0.32*	-1.72	0.44	-0.22	-1.15	0.20
BM9	-0.50***	-2.57	0.99	-0.18	-0.88	0.12

BM10	-0.41	-1.60	0.39	-0.26	-1.06	0.17
Panel D: Momentum portfolios						
Lo PRIOR	-0.89***	-2.99	1.32	-0.79**	-2.33	0.81
PRIOR 2	-0.38*	-1.65	0.41	-0.26	-1.01	0.15
PRIOR 3	-0.39**	-2.06	0.63	-0.25	-1.13	0.19
PRIOR 4	-0.28	-1.59	0.38	-0.21	-1.05	0.17
PRIOR 5	-0.29*	-1.85	0.51	-0.18	-1.01	0.15
PRIOR 6	-0.42***	-2.63	1.03	-0.29	-1.60	0.39
PRIOR 7	-0.31**	-2.04	0.62	-0.21	-1.19	0.21
PRIOR 8	-0.36**	-2.31	0.80	-0.26	-1.48	0.33
PRIOR 9	-0.37**	-2.24	0.75	-0.41**	-2.14	0.69
Hi PRIOR	-0.57***	-2.69	1.08	-0.62**	-2.53	0.96

4. Economic Explanations

4.1. Cash flow and discount rate predictability

我们使用 1966-2021 年的年度数据进行本节的复刻。探讨了投资者情绪对股票市场收益的预测是否来自于现金流量还是贴现率这两个渠道。我们使用 DG、EG、GDPG,作为现金流量的代理变量,使用 D/P 作为贴现率的代理变量。

股息增长率和股息收益率根据文献构建:股息增长率为当年对数股息减上年对数股息。股息收益率通过总市场收益和除股息收益计算。收益增长率和 GDP 增长率直接取年度数据。所有变量进行标准化处理。进行回归分析,解释变量为投资者情绪指数,控制变量为股息收益率。检验投资者情绪指数对各现金流量指标的预测能力。按照资产定价理论,如果投资者情绪在控制股息收益率后仍能预测股票超额收益,则其效应来自于现金流量预测或贴现率预测。

根据图表分析得出: 投资者情绪对未来股息收益率的预测作用很弱,回归系数很小且不显著,投资者情绪显著负向预测未来的股息增长率,这与原文的结论一致。综上所述,可以得出这表明投资者情绪主要通过现金流量通道预测股票收益,而非贴现率通道,与原文结论一致。

表 4-1：渠道检验结果

<i>Panel A: Aligned investor sentiment, S^{PLS}</i>					
	β	t-stat	ψ	t-stat	R^2 (%)
D/P	0.02	0.89	0.93	16.54	84.16
DG(%)	-1.69	-2.25	-0.05	-0.03	8.93
EG(%)	-3.94	-0.90	4.70	0.42	1.82
GDPG(%)	-0.12	-0.45	-0.12	-0.18	0.47

<i>Panel B: BW investor sentiment, S^{BW}</i>					
	β	t-stat	ψ	t-stat	R^2 (%)
D/P	0.00	-0.01	0.94	16.2	83.92
DG(%)	-1.57	-1.94	-0.94	-0.48	6.76
EG(%)	-8.04	-1.76	0.72	0.07	5.88
GDPG(%)	-0.24	-0.87	-0.24	-0.35	1.51

4.2. The cross-section of cash-flow channel

在本章节中，我们使用预测回归检验了 S^{PLS} 对现金流量横截面的预测能力，方程如下：

$$DG_{i+1}^j = a_j + \Phi_j S_i^{PLS} + \vartheta_{i+1}^j$$

接着，我们检验投资者情绪预测股票收益的能力是否与其预测交叉市场现金流的能力呈正相关。我们使用横截面回归来统计测试现金流渠道，如果这个假设成立，那些更容易被投资者情绪预测的公司也应该有更高的现金流可预测性。截面回归的方程如下：

$$\beta_j = a + g\Phi_j + e_j$$

表 4-2： S^{PLS} 对现金流量横截面的预测能力

Aligned investor sentiment index, S^{PLS}			
Panel A: Cross-sectional regression, $\beta_j = a + g\Phi_j + e_j$			
	g	t-stat	R^2 (%)
	0.38	4.27	69.6
Panel B: Forecasting dividend growth of size portfolios			
	Φ_j	t-stat	R^2 (%)
Small	-1.05	-4.00	2.41

2	-1.06	-5.35	4.23
3	-1.15	-6.08	5.40
4	-0.97	-4.95	3.65
5	-0.84	-3.72	2.09
6	-1.04	-5.06	3.81
7	-0.65	-2.40	0.88
8	-0.58	-2.05	0.65
9	-0.69	-2.35	0.85
Large	-0.74	-2.87	1.26

表 4-4 的 Panel A 显示，投资者情绪的回报敞口较高的公司，其现金流对投资者情绪的敞口也显著较高。对于 10 个规模的投资组合，第二个方程中 g 的 OLS 估计为 0.38， t 统计量为 4.27，表明 β_j 和 Φ_j 存在显著的正相关关系。因此，具有绝对值较大负 β_j 的 S^{PLS} 可预测性的小企业具有较高的 S^{PLS} 现金流可预测性，也具有较大的 Φ_j 。

为了更深入地研究预测通道，表 4-4 的 Panel B 显示了第一个方程的回归结果。对于所有 10 个规模的投资组合， S^{PLS} 是现金流 DG_{i+1}^j 的显著负预测因子，与我们此前在总体市场证据一致。最重要的是，我们发现了一个有趣的横截面模式：更具投机性和难以套利的投资组合的现金流更容易被投资者情绪预测。较大规模投资组合的 R^2 普遍小于较小规模投资组合的 R^2 ，小规模投资组合通常被认为更具投机性，更难以套利。

4.3. Market-volatility risk

这一小节研究市场风险是否可以解释投资者的股票回报可预测性。按照波动性风险的观点，高 S^{PLS} 应该预测股市总波动性较低，从而降低市场风险，进而降低股票风险溢价。为了研究这个问题，我们构建以下方程进行回归，

$$\beta_j = \alpha + g\varphi_j + e_j$$

$$DG_{t+1}^j = \alpha_j + \varphi_j S_t^k + v_{t+1}^j, \quad k = PLS, BW, \quad j = 1, \dots, 10$$

其中， β 是 3.6 小节面板 B 中对于 10 个规模投资组合的回归结果中的斜率。回归结果如下表所示。

表 4-3：对于 10 个规模投资组合的回归结果

Aligned investor sentiment index, S^{PLS}	BW investor sentiment index, S^{BW}
Panel A: Cross-sectional regression, $\beta_j = a + g\varphi_j$	

	g	t-stat	R ² (%)		g	t-stat	R ² (%)
	0.05	5.57	79.51		0.06	8.40	89.81
Panel B: Forecasting dividend growth of size portfolios							
	φ _j	t-stat	R ² (%)		φ _j	t-stat	R ² (%)
Small	-5.44	-3.99	2.41		-7.50	-4.83	3.48
2	-5.66	-5.35	4.23		-6.11	-5.02	3.74
3	-5.54	-6.08	5.40		-5.07	-4.79	3.43
4	-4.59	-4.95	3.65		-4.77	-4.47	2.99
5	-3.60	-3.72	2.09		-3.06	-2.75	1.15
6	-4.05	-5.06	3.81		-2.86	-3.08	1.44
7	-2.44	-2.39	0.88		-1.30	-1.11	0.19
8	-1.99	-2.05	0.65		-1.59	-1.42	0.31
9	-2.15	-2.35	0.85		-0.72	-0.68	0.07
Large	-2.26	-2.87	1.26		-0.75	-0.83	0.11

接着，我们构建了以下方程进行回归。我们将标普 500 指数的每日超额收益按月求平方和，得到每月股市总方差 $SVAR_{t+1}$ ，再对其取对数得到 $LVOL_{t+1}$ 进行回归。

$$LVOL_{t+1} = \alpha + \beta S_t^{PLS} + \psi LVOL_t + v_{t+1}$$

$$LVOL_{t+1} = \log(\sqrt{SVAR_{t+1}})$$

$$SVAR_{t+1} = \sum_{i=1}^{N_{t+1}} R_{i,t+1}^2$$

结果如下表所示， S^{PLS} 对市场波动表现出正向的预测能力， β 值为 0.0094，这与波动率假设不一致。尽管我们不能完全排除基于风险的解释，但市场风险似乎不太可能推动 S^{PLS} 对于股票市场回报的预测能力。

表 4-4: S^{PLS} 对市场波动的预测能力

$LVOL_{t+1}$	β	t-stat	R ² (%)
S_t^{PLS}	0.0094	0.914	44.51

参考文献:

- [1] Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor Sentiment Aligned: A Powerful Predictor of Stock Returns. *Review of Financial Studies*, 28(3), 791–837.
<https://doi.org/10.1093/rfs/hhu080>
- [2] <http://www.econ.yale.edu/~shiller/>
- [3] <https://pages.stern.nyu.edu/~jwurgler/data/>