

“持续异常交易量”选股因子 PATV

专题报告

——量化研究 2023 年中期投资策略

异常交易量因子（ATV）是在金融领域中被广泛研究的一个因子，用于衡量股票短期交易量的异常增加。但在实际测试中，我们发现 ATV 因子对于多头端股票的区分能力并不明显，因子收益的较大占比来自空头端负 Alpha。因此，在过去的研究基础上，本文提出了异常交易量因子的一种新的度量方式——“持续异常交易量”因子 PATV，对 ATV 因子的缺陷进行了改进。本文的主要结论如下：

- 测试结果表明，异常交易量因子 ATV 因子的实际选股效果却并不如学术研究中的那样理想，ATV 因子对于多头端股票的区分能力并不明显，其收益的较大占比来自空头端负 Alpha。
- 将“持续性”因素纳入到异常交易量因子中，本文构建了“持续异常交易量”因子 PATV。本报告的核心逻辑是零星的异常交易量尚不能形成确定性的反转趋势，异常交易的持续性越强，未来股票价格下跌的概率越大。
- 因子测试的结果表明，PATV 因子对股票未来收益具有显著的预测能力。PATV 因子与股票未来一个月的收益率呈显著的负相关关系，在全 A 股样本中的 IC 均值及 T 统计量分别为 -0.09 和 -11.273；根据 PATV 因子划分的分位数组具有明显的区分度，且各组之间具有严格的线性变化趋势。
- 通过稳健性检验，本文发现 PATV 因子的增量信息具备稳定性。本文通过敏感性测试，论证了 PATV 因子在剔除不同比例的非流动性尾部股票后仍具有稳定的超额收益，与同类因子正交后的残差因子（PATV residual）仍然具有明显的区分度和变化梯度，并且在过去十几年中稳定跑赢经典的异常交易量 ATV 因子。
- 在各类股票池中，基于 PATV 因子的组合优化增强策略回测表现突出，自 2007 年 2 月以来在沪深 300、中证 500 的年化收益分别为 8.41% 和 14.81%，自 2014 年 11 月以来在中证 1000 的年化收益为 11.56%。

风险提示：本报告结果通过历史数据统计、建模和测算完成，在政策、市场环境发生变化时模型存在失效的风险；本报告所提及个股或基金仅表示与相关主题有一定关联性，不构成任何投资建议。

任瞳 S1090519080004
rentong@cmschina.com.cn
麦元勋 S1090519090003
mai yuanxun@cmschina.com.cn
王武蕾 S1090519080001
wangwulei@cmschina.com.cn
周靖明 S1090519080007
zhoujingming@cmschina.com.cn
崔浩瀚 S1090519070004
cuihaohan@cmschina.com.cn
杨航 S1090523010004
yanghang4@cmschina.com.cn
罗星辰 S1090522070001
luoxingchen@cmschina.com.cn
梁雨辰 研究助理
liangyuchen2@cmschina.com.cn
李世杰 研究助理
lishijie1@cmschina.com.cn
许继宏 研究助理
xujihong@cmschina.com.cn
周游 研究助理
zhouyou4@cmschina.com.cn

正文目录

一、异常交易量因子 ATV: 空头端负 Alpha 的占比高?	5
二、“持续异常交易量”因子 PATV	6
2.1 对 ATV 因子空头逻辑的精细刻画	6
2.2 改进思路的初步验证	7
2.3 PATV 因子构建方法	8
三、单因子测试	9
3.1 IC 测试	9
3.2 分位数测试	10
3.3 在宽基指数成分股中的选股效果	11
四、稳健性测试	13
4.1 参数敏感性	13
4.2 与同类因子的相关性	14
4.3 相对经典 ATV 因子的改进效果	15
五、基于 PATV 因子的组合优化增强策略	16
5.1 构建方法	16
5.2 回测表现	16
六、主要结论	17

图表目录

图 1: 分组净值曲线 (ATV 因子)	5
图 2: 分组年化超额收益 (ATV 因子)	5
图 3: 分位数组合的多空净值曲线.....	6
图 4: 不同峰度下的数据分布	7
图 5: PATV 因子的 IC 变化 (全 A 股)	9
图 6: PATV 因子的 IC 变化 (沪深 300)	9
图 7: PATV 因子的 IC 变化 (中证 500)	9
图 8: PATV 因子的 IC 变化 (中证 1000)	9
图 9: 分位数测试——净值曲线 (PATV)	10
图 10: 分位数测试——年化超额收益 (PATV)	10
图 11: 分位数测试——净值曲线 (PATV)	11
图 12: 分位数测试——净值曲线 (ATV)	11
图 13: 分位数测试——净值曲线 (沪深 300)	11
图 14: 分位数测试——年化超额 (沪深 300)	11
图 15: 分位数测试——净值曲线 (中证 500)	12
图 16: 分位数测试——年化超额 (中证 500)	12
图 17: 分位数测试——净值曲线 (中证 1000)	12
图 18: 分位数测试——年化超额 (中证 1000)	12
图 19: 零成交数据占比分布 (全 A 股)	13
图 20: 不同剔除比例下的因子多空净值曲线.....	14
图 21: 不同剔除比例下的分位数组合年化超额	14
图 22: PATV 因子与同类因子的相关性	14
图 23: “残差因子”分组净值曲线.....	15
图 24: “残差因子”分组年化超额.....	15
图 25: PATV 因子与 ATV 的因子历史表现对比.....	15
图 26: PATV 策略_净值曲线 (沪深 300)	17
图 27: PATV 策略_净值曲线 (中证 500)	17
图 28: PATV 策略_净值曲线 (中证 1000)	17
表 1: IC 测试统计量 (ATV 因子)	5

表 2: 双分组检验 (单位: %) 7

表 3: 双分组检验 (单位: %) 8

表 4: PATV 因子的 IC 测试结果 9

表 5: 分位数测试——风险收益统计量 10

表 6: “残差因子”IC 测试结果 15

表 7: PATV 策略_回测指标 16

本文是招商量化“青出于蓝”系列的第五篇报告，“青出于蓝”这个词起源于中国传统文化，体现了文化传承中的一种演变和升华的过程，与因子投资中对传统量化因子的改进有着一定的共通点。传统量化因子能够大致体现策略的核心逻辑，但可能难以捕捉更加细微的市场变化和投资者情绪波动，因此需要不断改进和升级这些因子，以更好地适应市场的变化和投资者的需求，这也是我们招商量化团队不断追求“青出于蓝”的过程。

在本篇报告中，我们将介绍异常交易量因子（ATV）的一种新的度量方式——“持续异常交易量”因子 PATV。在过去的研究基础上，我们发现股票异常交易量持续的时间越长，随后产生的错误定价修正程度就越高，改进后的因子能够使得传统 ATV 因子的多头端收益得到有效增厚，因子的多头组合在未来一个月产生了显著的超额收益。

一、异常交易量因子 ATV：空头端负 Alpha 的占比高？

异常交易量因子（ATV, Abnormal Trading Volume）是在金融领域广泛研究的一个概念，通常被用于捕捉股票的异常交易活动，如投资者的情绪化买卖、信息泄露或者内幕交易等等。过去与异常交易量相关的研究非常丰富，如 Lee 和 Swaminat(2000)、Barber 和 Odean(2008)等人均表明，当异常交易量较高时，非理性投资者正主导着交易，此时股票价格往往会被高涨的交易情绪不断推高，超出其合理的范围。另一方面，由于市场长期存在非对称的风险偏好（Tversky& Kahneman, 1979），投资者对负面信息的反应通常比对正面信息的反应更加强烈。在这种非对称的风险偏好下，极端资金流所驱动的交易会对单只股票形成巨大价格压力，导致未来股价大幅下跌。

作为本文的重要基准变量，我们首先参照经典的构建方法，用股票过去 20 日的日均交易量（短期交易量）和过去 240 日的日均交易量（长期交易量）的商构建了 ATV 因子，并在 A 股中对其进行了回测（需要注意，为了避免缺乏流动性的股票影响回测结果的稳定性，我们仅对上市大于一年的样本进行研究，并且对流动性较差的 10%尾部样本进行了剔除）。自 2007 年 2 月以来，因子月度 IC 均值为 **-0.075**，IC_IR 为 **-0.644**，IC 的 T 统计量为 **-8.977**，若仅从 IC 测试的结果看，异常交易量因子 ATV 拥有相当强的选股能力。

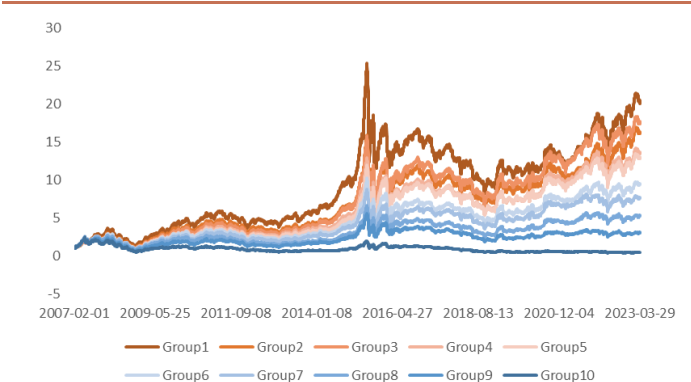
表 1: IC 测试统计量（ATV 因子）

因子	平均值	标准差	最小值	最大值	IC_IR	t 统计量	胜率
ATV 因子	-0.075	0.116	-0.436	0.346	-0.644	-8.977	76.29%

样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券、Wind

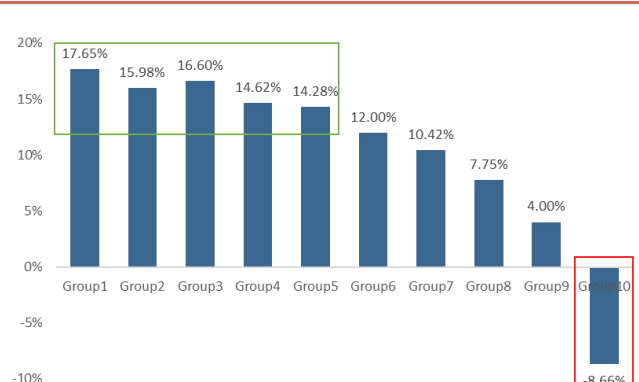
但从分位数测试的结果看，ATV 因子的实际效果却并不如学术研究中的那样理想。图 1 和图 2 分别为 ATV 因子在全 A 股中的分组净值曲线以及各组的年化超额收益率。从图 1 看，尽管 ATV 因子在历史上确实获得了最高的收益，但多头端各组的区分度始终不明显，并且因子多头组合的表现自 2017 年以来已经出现了明显逆转。另一方面，ATV 因子前五个分位数组合年化收益的差距不超过 4%（见图 2），第 1 组至第 3 组的年化超额分别为 17.65%、15.98% 和 16.60%，换言之，做多 ATV 因子前 10%股票和做多 ATV 因子前 30%股票所得到的年化收益率差距不超过 2%。可见，ATV 因子对于多头端股票的区分能力并不明显，因子收益的较大占比来自空头端负 Alpha。

图 1: 分组净值曲线（ATV 因子）



样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券

图 2: 分组年化超额收益（ATV 因子）



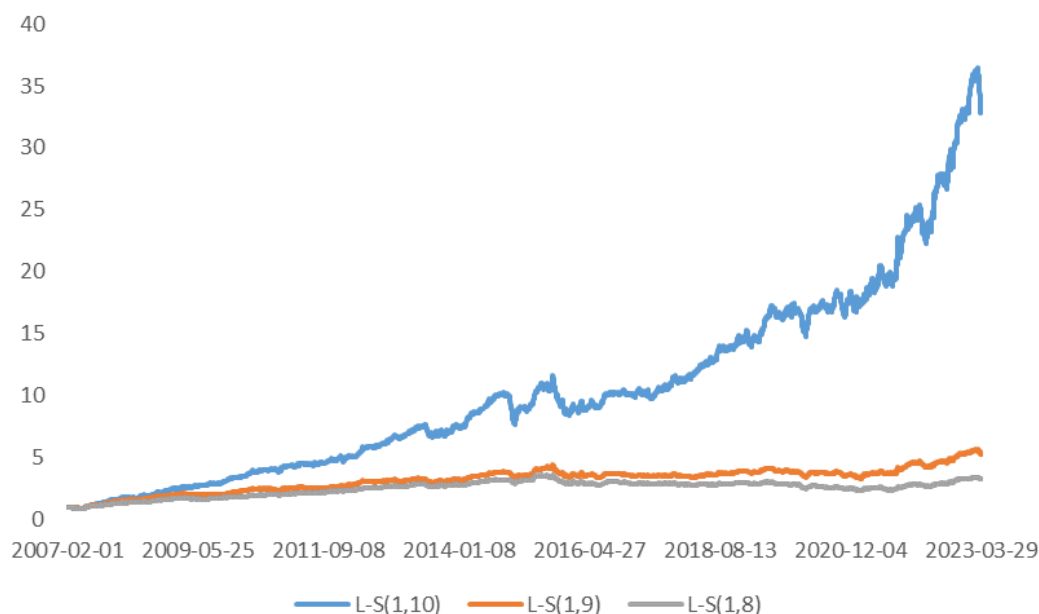
样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券

更直观地，我们分别计算了分位数组合中第 1 组（多头组）和第 10（空头组）、9、8 组的多空净值曲线，记为 L-S(1,10)、L-S(1,9)、L-S(1,8)，如图 3 所示。由图可知，L-S(1,10)自 2007 年 2 月以来累计涨幅超过 30 倍，而 L-

S(1,9)和 L-S(1,8)的累计涨幅仅仅不到 5 倍。三者显著的差异向我们清楚地展示了 ATV 因子收益的较大占比来自空头端负 Alpha。

鉴于原始的异常交易量因子（ATV）构建方法相对简单，我们认为因子依然具有较大的改进空间。因此，在接下来的部分中，我们将对 ATV 因子的逻辑进行更多的思考以及对因子进行更加精细的刻画，尝试进一步改善因子对股票的筛选效果。

图 3：分位数组组合的多空净值曲线



样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券

二、“持续异常交易量”因子 PATV

通常来说，大多数因子的改进方向应当是去寻找因子多头端的增强逻辑，并与原始因子进行结合以达到改进因子效果的目的。但对于异常交易量因子这类只有空头端逻辑而没有多头端逻辑的因子（只有“异常高”交易量，没有“异常低”交易量），我们几乎无法从逻辑出发挖掘因子多头端的增强逻辑。因此在本章中，我们尝试转换思路，从因子的空头端入手，通过更加细致地刻画因子的空头逻辑，从而提升整个因子的选股效果。

2.1 对 ATV 因子空头逻辑的精细刻画

现有研究普遍认为，由异常交易量造成的错误定价终将得到纠正，所以股票未来收益与异常交易量 ATV 应当呈显著负相关。然而，我们认为这种负相关还会在一定程度上受到异常交易量的持续性所影响。由于市场上存在大量可能导致股票出现异常交易量的因素，如公司定期报告的发布、国家政策影响、宏观经济和行业因素、技术分析指标达到阈值等等，均可能导致股票出现短期的异常交易量。但这些因素所导致的异常交易量属于正常的市场波动，是短期且不连续的，这种类型的异常交易量或许还无法形成强烈的价格反转。真正能够代表情绪化交易、具有非理性特征的应当是那些“持续异常”的高交易量。

这一想法已经有部分研究进行过论述，如 Black（1986）和 Barberis 等人（1998）发现，投资者的高情绪在短期内更有可能继续持续下去，并不断推动股价偏离其合理基本面。Bali 等人（2014）认为，如果投资者对流动性冲击反应不足，那么在异常交易量的负向效应出现之前，可能还会存在一个十分短暂的正收益漂移期。换言之，短期内零星的异常交易量尚不能形成确定性的反转趋势。相反，投资者情绪处于极端状态的时间越长，累计的股价偏差就越大，随后的错误定价纠正幅度也会越大，这也是本报告的核心逻辑。

2.2 改进思路的初步验证

对于如何评价“持续异常交易”，一个简单的方法就是计算过去一段时间某只股票的异常交易量位于全市场前 $n\%$ 的天数。如果某只股票的异常交易量多次位于全市场的前 $n\%$ ，则可以认为这只股票的交易量具有“持续异常”特征。

我们通过一个简单的统计来进行初步验证。在根据异常交易量因子（ATV）将全市场股票分为 5 组的基础上，我们统计了所有股票在过去 20 天里 ATV 的排名位于全市场前 20% 的天数 DAYS。在同一个 ATV 分位数组中，**DAYS 数值较低的股票对应“偶然异常”的交易量，DAYS 数值较高的股票对应“持续异常”的交易量。**

进一步，我们在各个 ATV 分位数组内根据 DAYS 再将股票分为 5 组，并统计了这 25 个组合未来一个月的平均收益率，具体见表 2。我们发现，对于异常交易量较低四个分位数组（ATV_1 至 ATV_4），交易量的不同分布形态对未来收益没有产生明显的影响。但对于本文重点关注的高异常交易组（ATV_5），具有“高持续性”异常交易量（DAYS 5）的股票组合未来收益显著低于其它组。

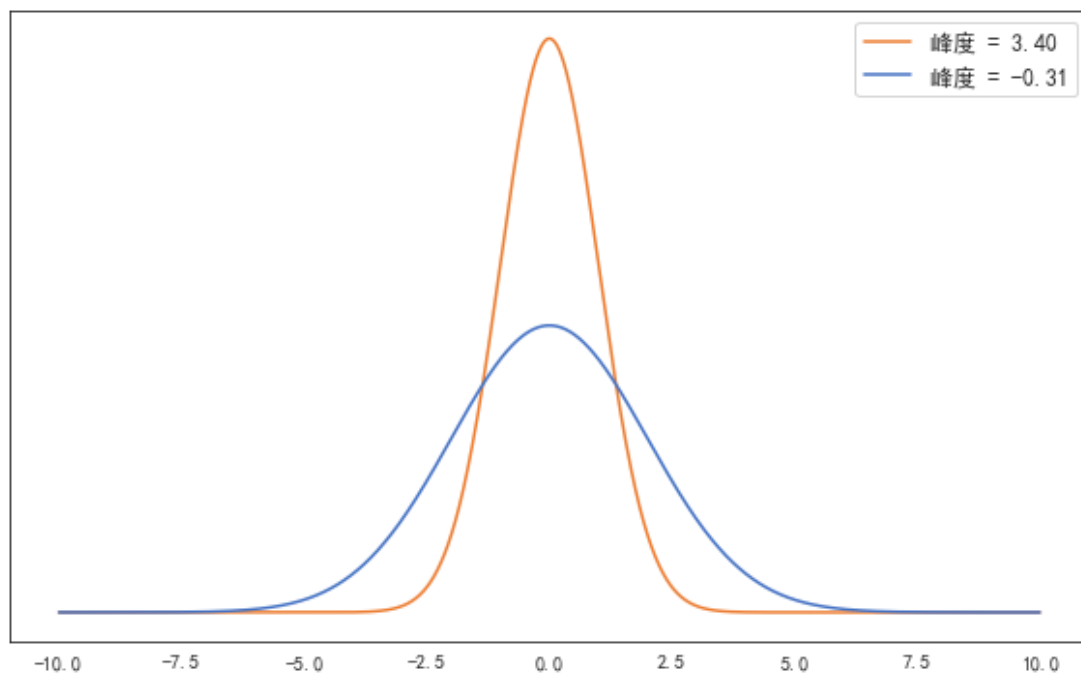
表 2：双分组检验（单位：%）

	DAYS_1	DAYS_2	DAYS_3	DAYS_4	DAYS_5
ATV_1	1.603	1.722	1.797	1.793	1.760
ATV_2	1.539	1.676	1.724	1.755	1.615
ATV_3	1.399	1.443	1.661	1.615	1.561
ATV_4	1.250	1.351	1.042	1.304	1.230
ATV_5	0.483	0.701	0.474	0.418	0.150

资料来源：招商证券、Wind；样本区间：20070201-20230331

说到持续性，我们自然而然还可以想到使用峰度这一更高阶的矩来描绘数据分布。如图 4 所示，在高峰态下，数据分布更集中（橙线），而低峰态下的数据分布则相对更分散（蓝线），峰度数值越高，代表股票异常交易量的持续性越强。此外，由于高峰态具有厚尾特征，能够帮助我们捕捉真正的“异常”，可谓一举两得。

图 4：不同峰度下的数据分布



资料来源：招商证券

同样地，我们也尝试对 ATV 及其峰度进行了双分组检验，结果如表 3 所示。与我们的猜想一致，峰度指标对股票也具有比较明显的区分度，**在异常交易量较高的股票组合中，峰度最高的组合未来一个月的收益率显著低于峰度最低的股票组合。**因此，我们同样可以借助峰度来对 ATV 因子的空头逻辑进行优化。

表 3：双分组检验（单位：%）

	KURT_1	KURT_2	KURT_3	KURT_4	KURT_5
ATV_1	1.694	1.703	1.892	1.751	1.773
ATV_2	1.564	1.729	1.809	1.777	1.497
ATV_3	1.603	1.653	1.450	1.555	1.525
ATV_4	1.397	1.215	1.131	1.336	1.166
ATV_5	0.466	0.518	0.492	0.607	0.297

资料来源：招商证券、Wind；样本区间：20070201-20230331

2.3 PATV 因子构建方法

基于上节所述的逻辑，我们将“持续性”因素纳入到异常交易量因子（ATV）中，尝试构建了改进的“持续异常交易量”因子（PATV）。

在研究数据的选取上，考虑到：①股票的日频交易数据粒度较粗，改进的空间有限，不足以让我们更加精细地描绘异常交易量的持续性；②一分钟频率的行情数据中存在大量的零成交数据。已有大量的文献证明了“零成交异常”的存在，可能会对我们的研究形成较多的噪声。因此在本篇报告中，我们将使用五分钟频率的高频成交数据开展研究（便于表达，后文中的“分钟”均指五分钟。并且我们与上文保持相同的操作，剔除流动性最低的尾部 10% 股票后，将剩余股票作为最终的研究样本）。具体来说，“持续异常交易量”因子 PATV 的构建可以分为以下几步：

■ Step1: 计算“日内异常交易量” ATV_d ;

在经典定义中，异常交易量因子等于短期交易量和长期交易量之比。因此类似地，我们以每只股票的分钟交易量衡量“短期交易量”，用股票过去一段时间所有的分钟交易量的均值衡量“分钟交易量中枢”，并将两者做商计算得到股票的日内异常交易量 ATV_d ；

■ Step2: 计算 ATV_d 在截面上的排名分位数 $rank_ATV_d$;

通过计算每只股票的日内异常交易量 ATV_d 在全市场的排名百分位 $rank_ATV_d$ ，我们可以对股票某分钟的异常交易量在全市场的相对水平进行衡量；

■ Step3: 计算“日内持续异常交易量” $PATV_d$;

对于“持续异常交易量”的计算，最直观且简单的方法就是像上文所介绍的，计算过去一段时间股票的异常交易量位于全市场前 20% 分位数的次数，但如此计算得到的“持续次数”是一个整数型变量，除了数据较为稀疏外，其右偏分布的特征也十分明显。使用这样的因子，我们将难以对股票池进行分层。

为此，我们提出了另一种用于衡量“持续异常”的方式。每一天，我们对每只股票进行以下计算：将第二步中 $rank_ATV_d$ 的均值除以标准差，再加上其峰度，作为股票“日内持续异常交易量” $PATV_d$ 的衡量，具体如下式所示：

$$PATV_d = \frac{Mean(rank_ATV_d)}{Std(rank_ATV_d)} + Kurt(rank_ATV_d)$$

这个构建方法的思路在于， $rank_ATV_d$ 的均值越高，股票日内的异常交易量相对更高； $rank_ATV_d$ 的标准差越小，则该股票日内异常交易量相对更稳定；最后，如上文所述，高峰态的数据分布往往更集中。因此，三方面因素综合得到的股票“日内持续异常交易量” $PATV_d$ ，其数值越高，代表股票的日内异常交易量的持续性越强。

■ Step4: EMA 低频化为月度“持续异常交易量”因子PATV;

最后，由于本文构建的是月频选股因子，因此我们需要将第三步中得到的“日内持续异常交易量”低频化为月频的“持续异常交易量”因子 PATV。在时间序列数据平滑的方法上，相比于 SMA（Simple Moving Average，简单移动平均），EMA（Exponential Moving Average，指数移动平均）方法具有反应更快、滞后信息更少以及对趋势敏感性更强等众多优势，因此我们最终通过 EMA 法（half life），于每月月末构建得到月度的“持续异常交易量”因子 PATV。

三、单因子测试

3.1 IC 测试

为了检验 PATV 因子与股票下一期收益率之间是否存在显著的相关关系，我们在不同股票池里对各个因子进行了 IC 测试。表 4 是 PATV 因子分别在全 A 股、沪深 300 成分股、中证 500 成分股、中证 1000 成分股中的 IC 测试结果。从中我们可以发现：

- 1) 在全 A 股样本中，PATV 因子的 IC 均值为 **-0.09**，月度 ICIR 为 **-0.809**，IC 测试的 T 统计量为 **-11.273**。**结果表明因子与股票未来收益之间存在明显的负相关关系，并且该关系在统计上显著。**这意味着在全 A 股市场中，PATV 因子可以作为一个有效的选股因子，用于预测股票未来的表现；
- 2) 在**沪深 300**、**中证 500**和**中证 1000**三个不同的股票池中，PATV 因子的 IC 均值分别为**-0.044**、**-0.063**和**-0.07**。结果同样显示出 PATV 因子在这三个股票池中具有较强的选股能力；

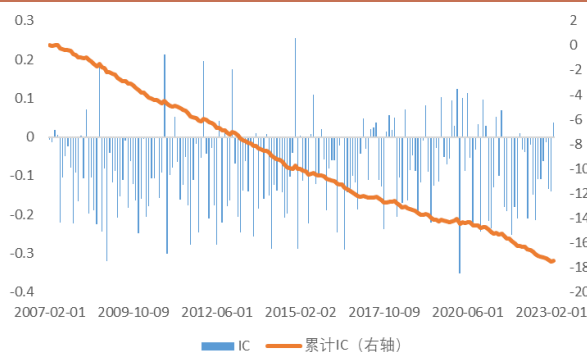
表 4: PATV 因子的 IC 测试结果

因子	平均值	标准差	最小值	最大值	IC IR	t 统计量	胜率
全 A 股	-0.09	0.111	-0.351	0.255	-0.809	-11.273	78.35%
沪深 300	-0.044	0.131	-0.337	0.355	-0.336	-4.675	63.92%
中证 500	-0.063	0.131	-0.326	0.289	-0.484	-6.738	68.04%
中证 1000	-0.07	0.107	-0.32	0.181	-0.654	-6.574	76.24%

样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券、Wind

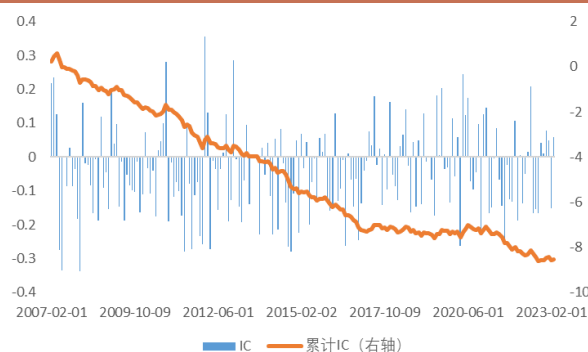
从 PATV 因子在各股票池中的 IC 变化看（见图 5 到图 8），无论在全 A 股、沪深 300 成分股、中证 500 成分股还是中证 1000 成分股中，PATV 因子的选股能力都具有较强的稳定性。一方面，IC 值大多在负值区域内波动，具有较高的胜率；另一方面，累计 IC 曲线的单调趋势平稳，表明**在各股票池中，PATV 因子与股票下期收益率之间的负相关性能长期维持稳定。**

图 5: PATV 因子的 IC 变化（全 A 股）



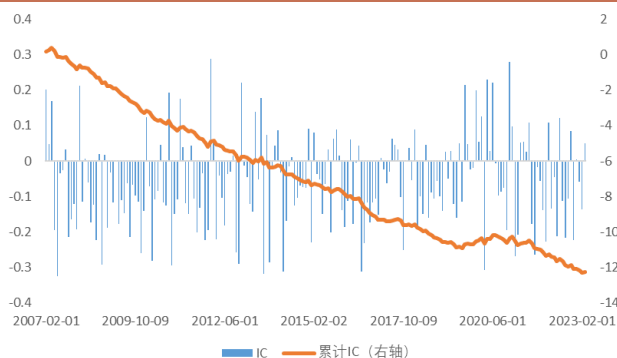
样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券、Wind

图 6: PATV 因子的 IC 变化（沪深 300）



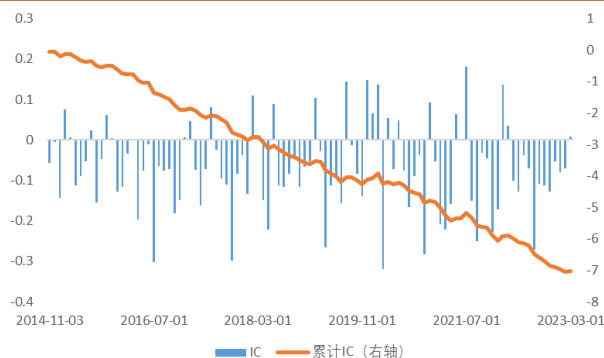
样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券、Wind

图 7: PATV 因子的 IC 变化（中证 500）



样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券、Wind

图 8: PATV 因子的 IC 变化（中证 1000）



样本区间：2014/11/03 至 2023/03/31；资料来源：招商证券、Wind

3.2 分位数测试

为了检验 PATV 因子的分组结果是否符合逻辑，我们进一步按 PATV 因子从低到高将全 A 股样本均匀分为 10 组，以等权的方式设定股票权重并进行了历史的回测（与 ATV 因子的构建过程保持一致，我们对全市场中流动性最差的 10% 尾部样本进行了剔除）。我们于每月结束后的首个交易日进行组合调整，回溯期为 2007 年 2 月 1 日至 2023 年 3 月 31 日。

表 5、图 9 和图 10 分别为 PATV 因子在全 A 股票池里进行分位数测试的风险收益统计量、净值曲线以及年化超额收益。其中，多头组合（Top）是 PATV 因子数值最小的股票组合，空头组合（Bottom）为 PATV 数值最大的股票组合。可以发现在全 A 股样本中：

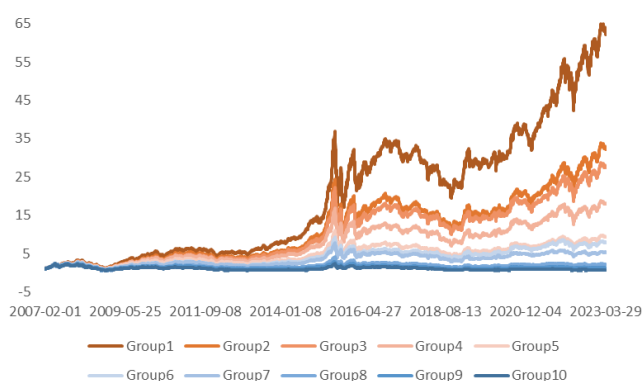
- 1) **PATV 因子的 Top 组合整体表现最好，年化收益率和夏普比率分别达到了 30.10% 和 0.951**，Bottom 组合的表现最差，年化收益率和夏普比率分别为 -1.58% 和 -0.141；
- 2) 从 Top 组合到 Bottom 组合，年化收益率、年化超额收益率和夏普比率整体的区分度较强。**随着 PATV 因子的因子值从大到小变化，股票分位数组的投资收益也越来越高，因子的分位数组具有较好的单调性。**

表 5：分位数测试——风险收益统计量

	总收益率	年化收益	年化超额	年化波动	夏普比率	最大回撤	平均换手率
Group1(Top)	6159.15%	30.10%	26.70%	28.50%	0.951	67.78%	44.79%
Group2	3141.62%	24.77%	21.37%	30.24%	0.72	67.99%	72.41%
Group3	2672.71%	23.53%	20.14%	31.02%	0.662	68.87%	80.09%
Group4	1708.26%	20.22%	16.82%	31.54%	0.546	69.53%	83.21%
Group5	842.44%	15.34%	11.94%	32.19%	0.383	69.66%	84.05%
Group6	701.28%	14.15%	10.76%	32.66%	0.342	72.89%	84.48%
Group7	439.24%	11.31%	7.92%	32.59%	0.255	73.53%	84.10%
Group8	130.58%	5.46%	2.06%	32.92%	0.075	73.22%	82.21%
Group9	35.16%	1.94%	-1.46%	33.02%	-0.032	74.74%	78.37%
Group10(Bottom)	-22.09%	-1.58%	-4.97%	32.52%	-0.141	75.82%	59.75%

样本区间：2007/02/01 至 2023/03/31；基准：沪深 300；资料来源：招商证券、Wind

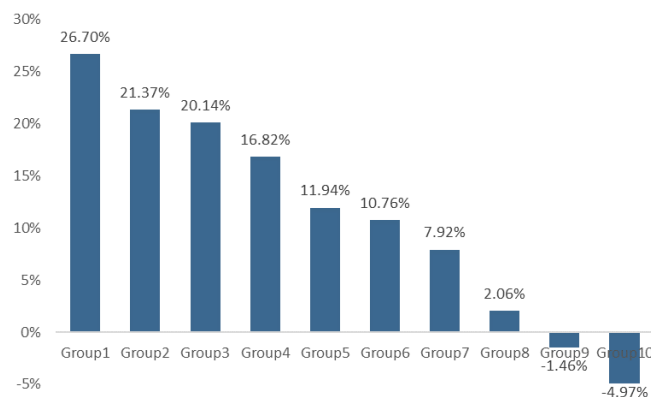
图 9：分位数测试——净值曲线（PATV）



样本区间：2007/02/01 至 2023/03/31；

资料来源：招商证券、Wind

图 10：分位数测试——年化超额收益（PATV）



样本区间：2007/02/01 至 2023/03/31；

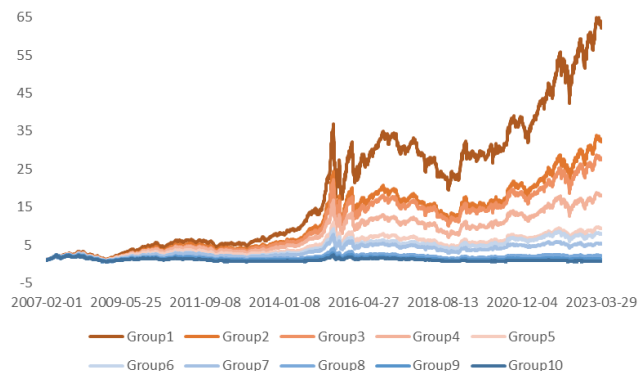
资料来源：招商证券、Wind

更加清晰地，我们把“持续异常交易量”因子 PATV 和“异常交易量”因子 ATV 进行对比，可以发现：

- 1) **相比于 ATV 因子，PATV 因子的分位数组单调性明显改善。**如上文所述，尽管 ATV 因子在历史上确实具有较好的表现，但多头端各组的区分度始终不明显，且自 2017 年以来多头端的收益几乎出现了逆转，因子的选股能力受空头端的影响较大。而对比之下，对异常交易现象进行更精细刻画的“持续异常交易量”因子 PATV 不仅在空头端中适用，其在全市场样本中均具有比较好的区分度（见图 11）。从 Top 组到 Bottom 组，各相邻分位数组组合间变化的梯度明显增加；

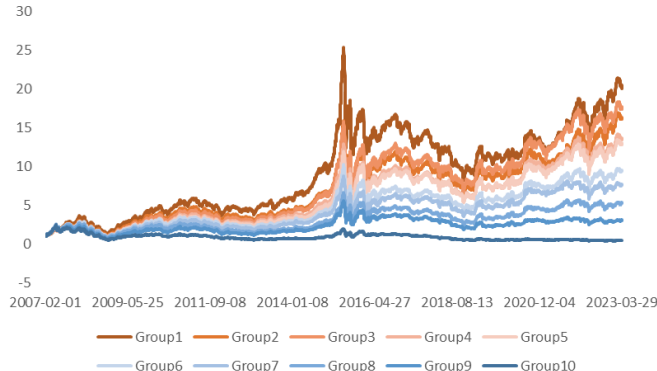
- 2) **PATV 因子的多头收益显著**。自 2007 年 2 月至今，PATV 因子多头、空头组合的年化收益率分别为 **30.1%**和 **-1.58%**，同期 ATV 因子多头、空头组合的年化收益率分别为 **20.7%**和 **-4.67%**。可见，PATV 因子不仅对空头端维持了较好的筛选效果，同时多头端相对 ATV 因子具有更突出的相对优势。

图 11: 分位数测试——净值曲线 (PATV)



样本区间: 2007/02/01 至 2023/03/31;
资料来源: 招商证券、Wind

图 12: 分位数测试——净值曲线 (ATV)



样本区间: 2007/02/01 至 2023/03/31;
资料来源: 招商证券、Wind

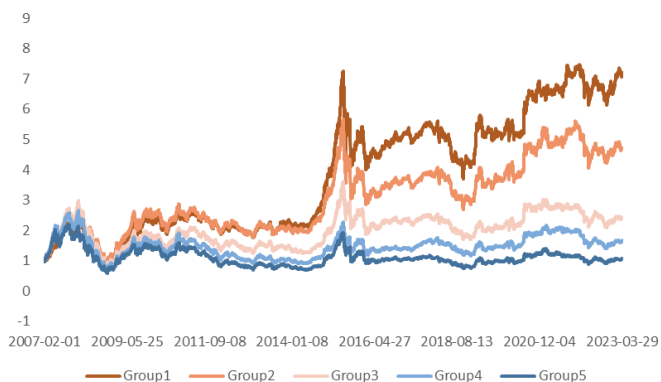
3.3 在宽基指数成分股中的选股效果

进一步，我们也对 PATV 因子在各宽基指数中的分层效果进行展示。由于指数成分股数量较少，因此在宽基指数样本中我们将股票分为五组进行测试。

3.3.1 沪深 300 成分股

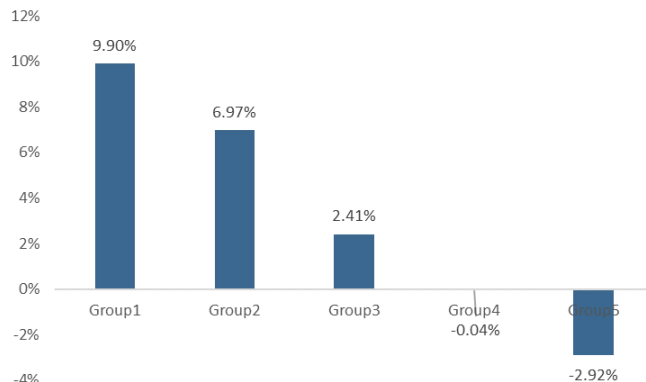
图 13 和图 14 分别为 PATV 因子在沪深 300 指数成分股中进行分位数测试的分组净值曲线和分组年化收益。从结果中可知: (1) 在沪深 300 指数中，根据 PATV 因子筛选的 Top 组合整体表现最好，Bottom 组合的表现最差，**相对沪深 300 指数的年化超额分别为 9.90%和-2.92%**; (2) **从分位数组的单调性看，PATV 因子在沪深 300 指数中具有显著的区分度**。从 Top 组到 Bottom 组，年化超额收益下降的梯度明显，并且从净值曲线看，表现最好的 Top 组合一直在最上方波动，在整个样本期内和 Bottom 组合的表现基本维持较大的差距。

图 13: 分位数测试——净值曲线 (沪深 300)



样本区间: 2007/02/01 至 2023/03/31;
资料来源: 招商证券、Wind

图 14: 分位数测试——年化超额 (沪深 300)

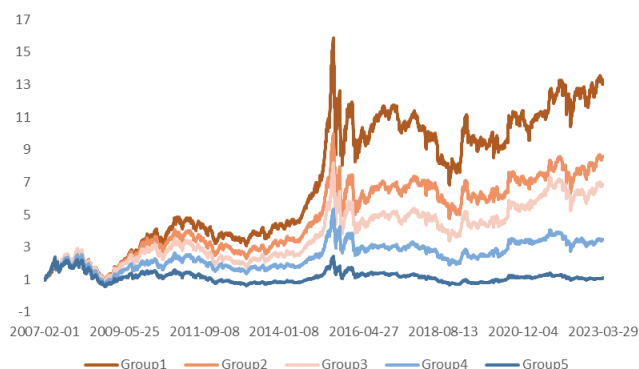


样本区间: 2007/02/01 至 2023/03/31;
资料来源: 招商证券、Wind

3.3.2 中证 500 成分股

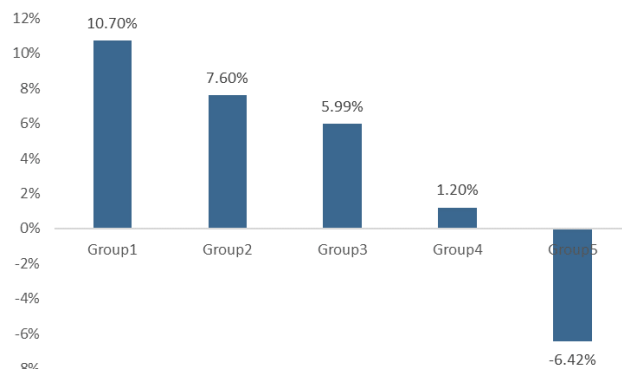
图 15 和图 16 分别为 PATV 因子在中证 500 指数成分股中进行分位数测试的分组净值曲线和分组年化收益。从结果中可知：（1）在中证 500 指数中，根据 PATV 因子筛选的 Top 组合整体表现最好，Bottom 组合的表现最差，相对中证 500 指数的年化超额分别为 10.70%和-6.42%；（2）从分位数组合的单调性看，PATV 因子在中证 500 指数中具有显著的区分度。从 Top 组到 Bottom 组，年化超额收益下降的梯度明显，并且从净值曲线看，表现最好的 Top 组合一直在最上方波动，在整个样本期内和 Bottom 组合的表现基本维持较大的差距。

图 15: 分位数测试——净值曲线（中证 500）



样本区间：2007/02/01 至 2023/03/31；
资料来源：招商证券、Wind

图 16: 分位数测试——年化超额（中证 500）

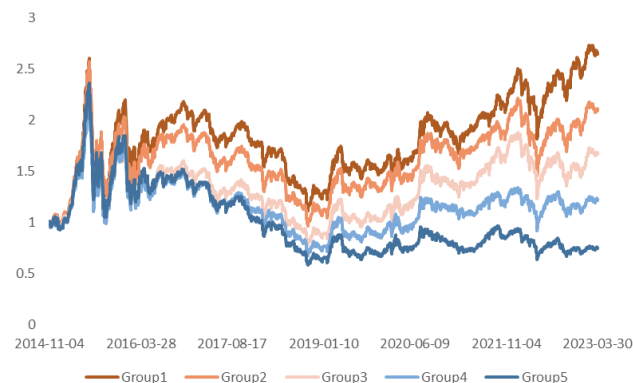


样本区间：2007/02/01 至 2023/03/31；
资料来源：招商证券、Wind

3.3.3 中证 1000 成分股

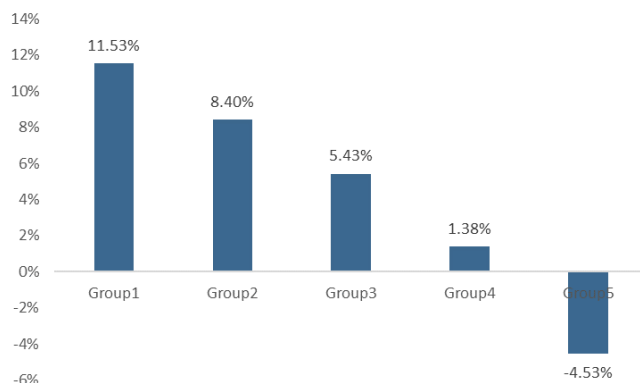
图 17 和图 18 分别为 PATV 因子在中证 1000 指数成分股中进行分位数测试的分组净值曲线和分组年化收益。从结果中可知：（1）在中证 1000 指数中，根据 PATV 因子筛选的 Top 组合整体表现最好，Bottom 组合的表现最差，相对中证 1000 指数的年化超额分别为 11.53%和-4.53%；（2）从分位数组合的单调性看，PATV 因子在中证 1000 指数中具有显著的区分度。从 Top 组到 Bottom 组，年化超额收益下降的梯度明显，并且从净值曲线看，表现最好的 Top 组合一直在最上方波动，在整个样本期内和 Bottom 组合的表现基本维持较大的差距。

图 17: 分位数测试——净值曲线（中证 1000）



样本区间：2014/11/01 至 2023/03/31；
资料来源：招商证券、Wind

图 18: 分位数测试——年化超额（中证 1000）



样本区间：2014/11/01 至 2023/03/31；
资料来源：招商证券、Wind

四、稳健性测试

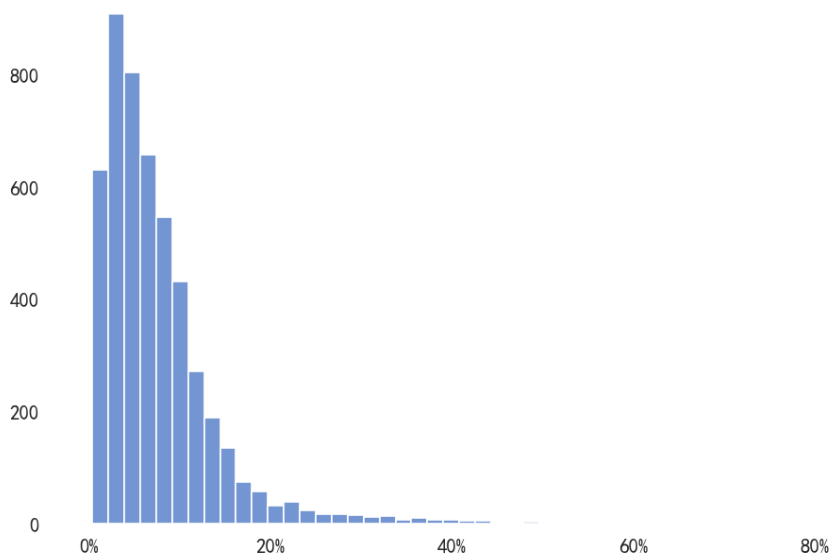
4.1 参数敏感性

我们在前文提到 ATV 和 PATV 因子的构建流程中采取了流动性尾部剔除的处理。在本节，我们将详细阐述尾部样本剔除对构建 PATV 因子选股组合的必要性，以及投资组合对剔除尾部样本的参数敏感性。

进行流动性剔除的主要原因有两个。**首先，大量现有研究结果表明，价量类因子普遍与低波因子、非流动性因子具有较高的相关性。**换言之，价量类因子的多头组合通常容易选出流动性较差的股票。流动性差的股票往往存在交易成本高、交易量小以及风险较高等问题，并且会对股票的价格产生显著的影响，使得我们对投资组合收益和因子有效性的测度不够准确可靠。

进行尾部剔除的**第二个原因在于高频行情数据中大量的零成交数据对因子多头组合的合理性有较大影响。**图 19 为所有股票样本的零交易量样本占比分布图。根据统计，在所有的 1 分钟频交易量样本中，**平均有 8% 的分钟交易量数据均为零，并且零交易量样本占比大于 20% 的股票也不在少数。**若直接使用未经筛选的样本构建投资组合，因子的多头组合中会存在大量流动性极差的股票，其投资收益或许有很大一部分来自于我们上文所提到的“零成交异常”，这样构建的投资组合是理想化的，其投资收益在实践中往往难以实现。

图 19：零成交数据占比分布（全 A 股）

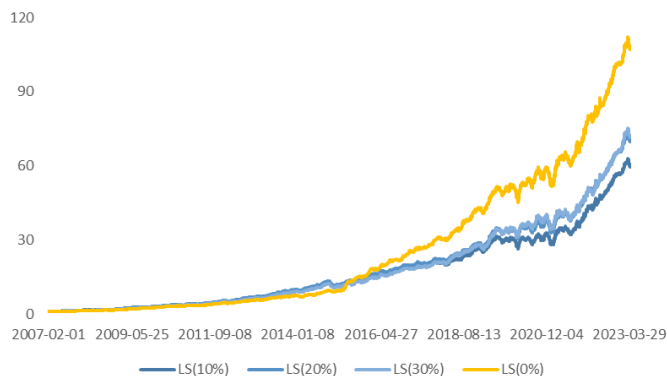


资料来源：招商证券、Wind

我们对比了不剔除、剔除流动性尾部 10%、20% 和 30% 这几种情况下的 PATV 因子多空净值曲线，如图 20 所示。可以清楚地发现，**当不剔除流动性较差的尾部样本时，PATV 因子多空组合自 2007 年至今的累计收益率高达 110 倍以上，显著高于其它三类情况下的收益**，这样的异常实际上就是由非流动性所导致的。而另一边，剔除流动性尾部 10%、20% 和 30% 样本后的因子多空净值曲线和分组年化超额收益十分接近，因子多头组合的年化超额收益稳定在 25% 以上，且各分位数组合呈现出比较良好的梯度变化（见图 21），**PATV 因子具有更强的稳定性，受尾部样本影响的敏感性也显著降低。**

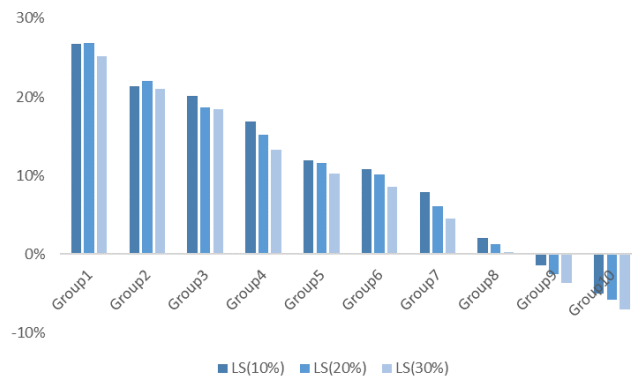
综上，虽然剔除流动性较差的尾部样本后，PATV 因子的收益显著下降，但我们依然坚持这种更合理和符合实际情况的做法。对流动性最差的尾部样本进行剔除既可以提高因子的准确性和可靠性，同时也能够让我们构建的因子投资策略具有更高的可执行性。

图 20: 不同剔除比例下的因子多空净值曲线



样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券

图 21: 不同剔除比例下的分位数组年化超额



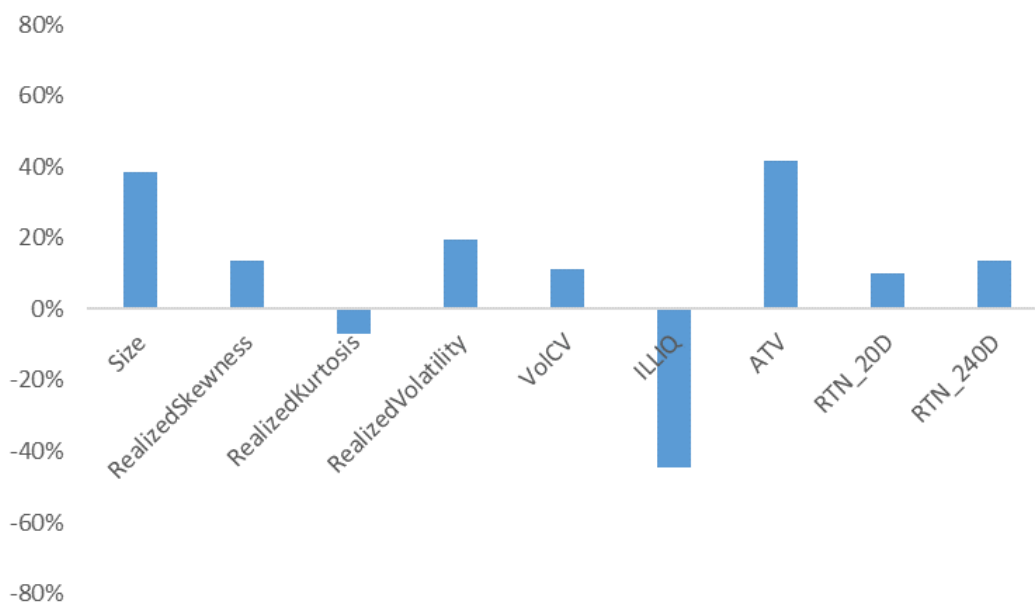
样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券

4.2 与同类因子的相关性

我们检验了 PATV 因子与部分相似因子的相关性，包括 Size（总市值）、RealizedSkewness（已实现偏度）、RealizedKurtosis（已实现峰度）、RealizedVolatility（已实现波动）、VolCV（交易波动）、ILLIQ（非流动性）、ATV（异常交易量）、RTN_20D（20 日动量）和 RTN_240D（240 日动量）等，这些因子均在逻辑上或者数据源上与 PATV 因子存在或多或少的相似性。

从统计结果看，与大多数的价量类因子类似，**PATV 因子在市值、波动和流动性因子上具有一定的暴露**，而与偏度、峰度等收益率分布特征因子以及动量类因子相关性较低。

图 22: PATV 因子与同类因子的相关性



样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券、Wind

因此，我们进一步通过多元线性回归的方式将**市值、波动、流动性**三大高相关特征，以及**成长、盈利、价值、动量**等基本面风格特征进行剔除，以回归残差作为**剔除相似因子暴露后的“残差因子”**（PATV_residual），并对“残差因子”进行了 IC 测试与分位数测试。

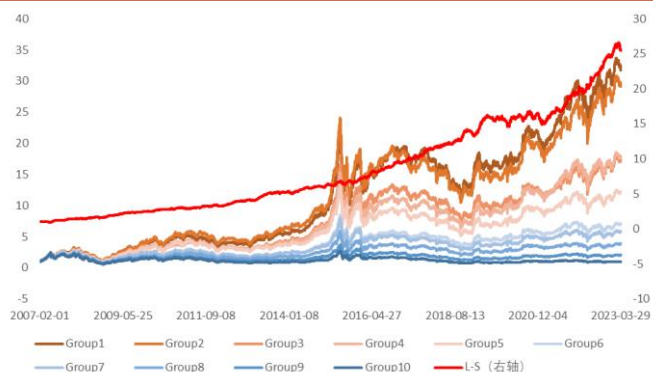
从表 6、图 23、图 24 中可以得知，和原始 PATV 因子相比，剔除相似因子暴露后的 PATV_residual 因子有效性有所下降，IC 均值从-9.0%下降至-6.8%，因子多头组合的年化收益率从 30.1%下降到了 24.68%。但尽管如此，**因子的 IC 依然维持在较高的水平**，且具有相当的显著性。各**分位数组的净值曲线仍然具有一定的区分度和变化梯度**，多空净值曲线上趋势平稳。

表 6: “残差因子” IC 测试结果

因子	平均值	标准差	最小值	最大值	IC_IR	t 统计量	胜率
PATV	-0.09	0.111	-0.351	0.255	-0.809	-11.273	78.35%
PATV_residual	-0.068	0.085	-0.268	0.211	-0.799	-11.126	66.8%

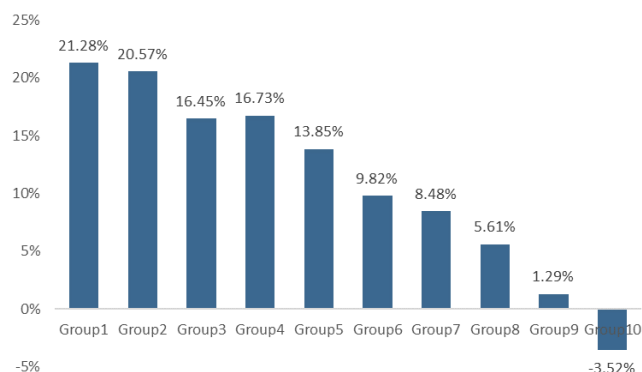
样本区间: 2007/02/01 至 2023/03/31; 资料来源: 招商证券、Wind;

图 23: “残差因子” 分组净值曲线



样本区间: 2007/02/01 至 2023/03/31; 资料来源: 招商证券

图 24: “残差因子” 分组年化超额



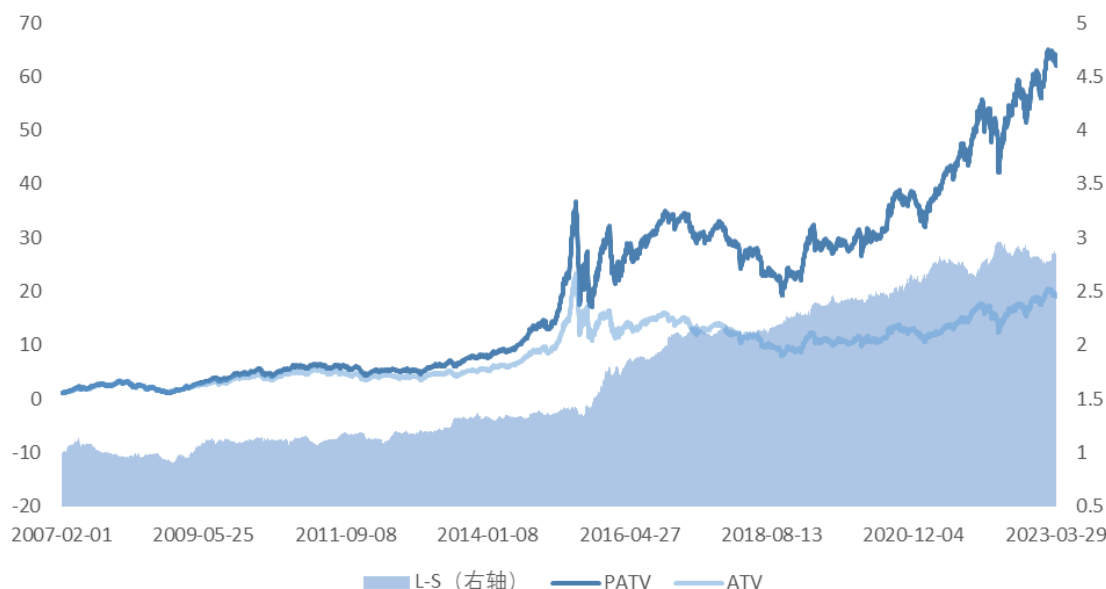
样本区间: 2007/02/01 至 2023/03/31; 资料来源: 招商证券

4.3 相对经典 ATV 因子的改进效果

回到本篇报告的出发点，我们的核心目标是“改进经典异常交易量因子 ATV”。因此我们把 PATV 策略组合与 ATV 策略组合历史上的回测表现进行直接对比，探究纳入异常交易量的“持续性”信息后能否更加细致地对全市场的股票进行区分，将 ATV 因子的选股效果从空头端扩展至全样本，从而获得更可观的收益。

将 PATV 因子多头组合与 ATV 因子多头组合的净值曲线进行对比后（见图 25）我们发现，PATV 因子在过去的十余年里几乎稳定跑赢原始 ATV 因子，并且在 ATV 因子逐渐失效的近些年，PATV 因子的相对收益变得更加明显，可见 PATV 因子确实在 ATV 因子之上取得了进一步的增强。

图 25: PATV 因子与 ATV 的因子历史表现对比



样本区间: 2007/02/01 至 2023/03/31; 资料来源: 招商证券、Wind

五、基于 PATV 因子的组合优化增强策略

5.1 构建方法

基于上文的一系列测试，我们已经对“持续异常交易量”因子 PATV 的选股能力及其相对于经典 ATV 因子的信息增量有了较具体的了解。因此在本章中，我们将进一步利用 PATV 因子在不同宽基指数股票池中构建组合优化增强策略（简称 PATV 策略）。由于篇幅的限制，我们接下来仅对沪深 300 和中证 500 指数以及中证 1000 指数中的 PATV 策略回测结果进行展示。PATV 策略的主要构建步骤如下：

- **初始股票池：**沪深 300 指数成分股、中证 500 指数成分股、中证 1000 指数成分股；
- **股票池筛选：**剔除上市时间小于半年以及交易当天不可交易、停牌、ST/ST*和涨跌停的股票样本，以排除新股和不可交易情况对策略回测的影响；
- **策略的构建：**每月月末，以投资组合的 PATV 因子暴露最小化作为目标函数，并加以卖空约束，使得投资组合的个股权重和行业权重限制在基准指数成分股权重和行业权重的一定范围内，以此构建得到当期的 PATV 股票组合；
- **交易设置：**我们于每月结束后的首个交易日进行持仓的调整，以股票的后复权收盘价进行交易；
- **比较基准：**沪深 300 指数（000300.SH）、中证 500 指数（000905.SH）和中证 1000 指数（000852.SH）；
- **回测时间区间：**2007 年 2 月 1 日至 2023 年 03 月 31 日；
- **交易费用：**扣除单边 2‰ 的交易费用。

5.2 回测表现

表 7 为 PATV 策略在各类宽基指数成分股中的回测统计量。从中可知，PATV 策略在沪深 300、中证 500 和中证 1000 指数成分股中表现突出：（1）沪深 300 和中证 500 策略自 2007 年 2 月至今相对沪深 300 指数和中证 500 指数的年化超额收益率分别为 5.01%、7.74%；（2）中证 1000 策略自 2014 年 11 月以来同样获得了较可观的收益，策略组合相对中证 1000 指数的年化超额收益率为 10.44%；

表 7：PATV 策略_回测指标

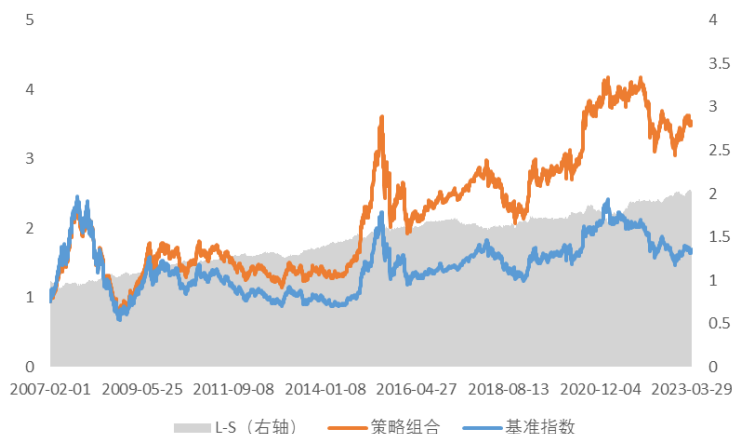
		总收益率	年化收益	年化超额	年化波动	夏普比率	最大回撤	平均换手率
沪深 300	策略组合	254.35%	8.41%	5.01%	25.87%	0.208	70.05%	34.35%
	基准指数	69.13%	3.40%	-	26.42%	0.015	72.30%	-
	L-S	101.47%	4.56%	-	4.58%	0.34	12.77%	-
中证 500	策略组合	776.27%	14.81%	7.74%	28.60%	0.413	68.57%	44.00%
	基准指数	192.78%	7.07%	-	29.90%	0.136	72.41%	-
	L-S	172.66%	6.59%	-	6.43%	0.558	22.07%	-
中证 1000	策略组合	145.11%	11.56%	10.44%	27.78%	0.308	61.38%	49.02%
	基准指数	9.56%	1.12%	-	28.11%	-0.067	72.35%	-
	L-S	117.06%	9.92%	-	7.53%	0.919	10.70%	-

样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券、Wind

图 26 至图 28 分别为 PATV 策略在各类宽基指数成分股中的策略回测净值曲线。可见：（1）无论是在沪深 300、中证 500 还是中证 1000 指数成分股中，在施加了一定的个股权重约束和行业权重约束后，三个指数增强策略均能够相对基准指数获得一定的超额收益，策略组合的净值曲线始终位于基准指数上方。该收益并非来自于个股权重与

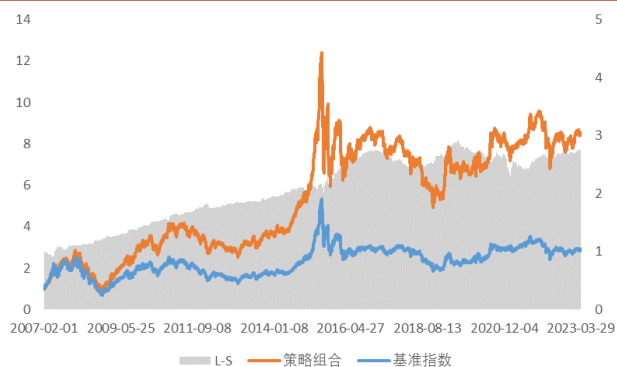
行业权重的极端偏离，而是主要来自于 PATV 因子的 Alpha 贡献。（2）总体而言，PATV 策略在中证 1000 指数成分股中超额收益最高，但在沪深 300 指数成分股中最稳。

图 26: PATV 策略_净值曲线（沪深 300）



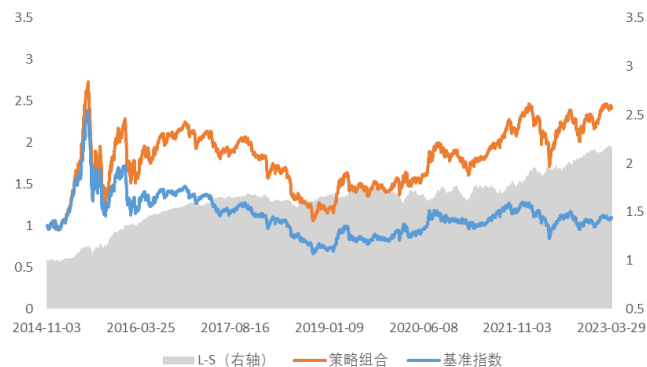
样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券

图 27: PATV 策略_净值曲线（中证 500）



样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券

图 28: PATV 策略_净值曲线（中证 1000）



样本区间：2007/02/01 至 2023/03/31；资料来源：招商证券

六、主要结论

本文提出了异常交易量因子的一种新的度量方式——“持续异常交易量”因子 PATV。在过去 ATV 因子的研究基础上，我们对因子的缺陷进行了改进，发现基于 PATV 构建的投资组合会在未来一个月产生显著的超额收益。本文的主要结论如下：

- **测试结果表明，异常交易量因子 ATV 因子的实际选股效果却并不如学术研究中的那样理想。**ATV 因子对于多头端股票的区分能力并不明显，因子收益的较大占比来自空头端负 Alpha；
- **将“持续性”因素纳入到异常交易量因子中，本文构建了“持续异常交易量”因子 PATV。**我们认为，零星的异常交易量尚不能形成确定性的反转趋势，异常交易的持续性越强，未来股票价格下跌的概率越大；
- **因子测试的结果表明，PATV 因子对股票未来收益具有显著的预测能力。**PATV 因子与股票未来一个月的收益率呈显著的负相关关系，在全 A 股样本中的 IC 均值及 T 统计量分别为 -0.09 和 -11.273；根据 PATV 因子划分的分位数组合具有明显的区分度，且各组之间具有严格的线性变化趋势；
- **通过稳健性检验，本文发现 PATV 因子的增量信息具备稳定性。**本文通过敏感性测试，论证了 PATV 因子在剔除不同比例的非流动性尾部股票后仍具有稳定的超额收益，与同类因子正交后的残差因子仍然具有明显的区分度和变化梯度，并且在过去十几年中稳定跑赢经典的异常交易量 ATV 因子；

- 在各类股票池中，PATV 选股策略均表现突出。单因子策略自 2007 年 2 月以来在沪深 300、中证 500 的年化收益分别为 8.41%和 14.81%，自 2014 年 11 月以来在中证 1000 的年化收益为 11.56%。

分析师承诺

负责本研究报告的每一位证券分析师，在此申明，本报告清晰、准确地反映了分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告中的具体推荐或观点直接或间接相关。

任瞳：研究发展中心执行董事，量化与基金评价团队负责人，管理学硕士，18 年证券研究经验，2010 年、2015 年至 2018 年、2020 年至 2022 年新财富最佳分析师（金融工程方向），在量化选股择时、基金研究以及衍生品投资方面有深入独到的见解。

麦元勋：研究发展中心董事，资深量化分析师，金融学博士，10 年量化策略研究开发经验，2015 年至 2018 年、2020 年至 2022 年新财富最佳分析师（金融工程方向），主要从事 ESG 投资策略、量化因子选股策略、事件驱动策略和基金评价等方向的研究开发。

王武蕾：资深量化分析师，武汉大学金融工程硕士，7 年量化策略研究开发经验。研究方向是量化资产配置与创新金融产品设计，在市场择时、风格判断、养老金产品设计等方面有长期深入的研究。

周靖明：资深量化分析师，武汉大学金融工程硕士，7 年量化策略研究开发经验。研究方向是多因子量化选股，另类 Alpha 研究，行业轮动，因子择时，此外，对 SmartBeta 基金和量化基金产品也有长期深入的研究。

崔浩瀚：量化分析师，浙江大学经济学硕士，5 年量化策略研究开发经验。研究方向是机器学习在金融领域的应用和多因子选股策略开发。

杨航：美国马里兰大学公园分校数量金融硕士，厦门大学金融工程学士，2021 年加入招商证券研究发展中心。

罗星辰：复旦大学应用统计硕士，中央财经大学经济学学士，2021 年加入招商证券研究发展中心，研究方向为择时，形态识别，技术面研究，文本挖掘等。

梁雨辰：研究助理，法国里昂高等商学院金融学硕士，应用数学与建模工程师，2021 年加入招商证券研究发展中心。

李世杰：研究助理，华南理工大学计算机科学和经济学双学士、金融硕士，2022 年加入招商证券研究发展中心。

许继宏：研究助理，中国科学技术大学金融硕士，2022 年加入招商证券研究发展中心。

周游：研究助理，北京大学人工智能学院，理学硕士，曾任职于华西证券金融工程组，2023 年加入招商证券研究发展中心。

评级说明

报告中所涉及的投资评级采用相对评级体系，基于报告发布日后 6-12 个月内公司股价（或行业指数）相对同期当地市场基准指数的市场表现预期。其中，A 股市场以沪深 300 指数为基准；香港市场以恒生指数为基准；美国市场以标普 500 指数为基准。具体标准如下：

股票评级

强烈推荐：预期公司股价涨幅超越基准指数 20%以上

增持：预期公司股价涨幅超越基准指数 5-20%之间

中性：预期公司股价变动幅度相对基准指数介于±5%之间

减持：预期公司股价表现弱于基准指数 5%以上

行业评级

推荐：行业基本面向好，预期行业指数超越基准指数

中性：行业基本面稳定，预期行业指数跟随基准指数

回避：行业基本面转弱，预期行业指数弱于基准指数

重要声明

本报告由招商证券股份有限公司（以下简称“本公司”）编制。本公司具有中国证监会许可的证券投资咨询业务资格。本报告基于合法取得的信息，但本公司对这些信息的准确性和完整性不作任何保证。本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价，在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。除法律或规则规定必须承担的责任外，本公司及其雇员不对使用本报告及其内容所引发的任何直接或间接损失负任何责任。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突。

本报告版权归本公司所有。本公司保留所有权利。未经本公司事先书面许可，任何机构和个人均不得以任何形式翻版、复制、引用或转载，否则，本公司将保留随时追究其法律责任的权利。