

高频寻踪：再觅知情交易者的踪迹

VCV 因子的构建与解析

在非完全有效的市场中，知情交易者可以利用自己的私有信息来获取超越市场的回报。我们无法通过直接观测来确定知情交易行为，但是知情交易者的加入将改变原有个股成交量的分布，可以通过知情交易者的蛛丝马迹研究知情交易行为。从文献中寻找思路，我们构建了 VCV 指标作为知情交易的代理变量。

我们认为知情交易可能会影响个股后市的收益，利用 VCV 因子在不同股票池中测试，均获得了不错的收益表现。

- 不难发现，不论是在 A 股还是在相对成熟的美股市场均存在市场对突发事件反应不足和反应过度的情形，市场远没有达到有完全有效。因而市场上依然存在知情交易者，可以利用自己的私有信息来获取超越市场的回报。对于知情交易者的研究一直是学术界和业界关注的问题。
- 我们无法通过直接观测的方式来获得知情交易者在市场上的行踪，但是，知情交易者的出现很可能打破原有交易的既定规律，从而我们可以通过观测交易流的不平衡性来推导知情交易者存在的概率。参考 Lof, Matthijs & Jos van Bommel (2022) 的研究成果，我们设计了 VCV 指标，并用 VCV 指标构建因子在不同的 A 股股票池中进行选股测试。
- 知情交易者占比较多的个股可能在短期内被市场关注和炒作，影响未来个股回报，同时知情交易者占比高的股票也不被机构投资者所青睐，因而理论表现会更弱。我们猜想 VCV 指标是个有效的选股因子。
- VCV 因子比较灵活，可以在不同的时间维度上进行构建。首先我们用日频数据进行因子暴露度的计算，并尝试每月调仓和每周调仓。我们发现，VCV 因子在全 A 范围内确实有比较不错的选股能力，2015 年以来截止今年 3 月底，我们发现月频 VCV 因子的年化 IR 是 2.489，t 值为 7.113，月频胜率能达到 73.5%。因子多空组合的总收益为 252.60%，年化收益为 16.32%。此外，在中证 1000 股票池中的表现较大盘股票池更为优秀。
- 在更高频数据领域，我们又利用 5 分钟数据和 1 分钟数据对因子暴露度进行计算，并且在周频调仓下进行测试。高频数据构建的 VCV 因子在不同股票池中的表现均优于日频数据构建的因子。中证 1000 股票池中，表现最为突出，自 2015 年以来截止今年 3 月底，5 分钟数据构建的 VCV 因子 rank_IC 为 0.052，年化 IR 为 1.789，IC 的 t 值为 10.583，胜率为 68.8%。多空组合的总收益为 2884.26%，年化收益高达 42.00%。
- 最后，综合考量换手率和交易成本，我们模拟实际的交易环境，对 VCV 因子进行月频调仓的测试。在模拟实际情况之后，我们发现，即使在交易成本较高且资金没有办法充分利用的情况下，从 2021 年以来截止今年 3 月底，VCV 因子的表现仍然较为优秀，组合收益为 2.31%，同期的基准（中证全指）收益为 -9.74%，累积超额收益为 12.05%。

风险提示：本报告数据基于历史数据进行测试，如遇到较为极端的市场变化，可能会有模型失效的风险。

深度报告

金融工程

任瞳 S1090519080004
rentong@cmschina.com.cn
崔浩瀚 S1090519070004
cuihaohan@cmschina.com.cn

正文目录

一、寻觅知情交易者的蛛丝马迹.....	4
二、因子的构建.....	4
指标构建.....	4
模拟验证.....	5
三、因子测试与结果展示.....	7
月频因子表现.....	7
数据与因子计算.....	7
周频因子表现.....	12
数据与因子计算.....	12
四、日内数据构造因子测试.....	15
利用 5 分钟频率数据构建 VCV 因子.....	15
利用 1 分钟频率数据构建 VCV 因子.....	17
VCV 因子与大类风格因子相关性展示.....	18
五、模拟组合构建与回测.....	19
六、总结.....	20

图表目录

图 1 当 $\eta=0$ 时，成交量的分布情况.....	6
图 2 当 $\eta=0.1$ 时，成交量的分布情况.....	6
图 3 当 $\eta=0.5$ 时，成交量的分布情况.....	6
图 4 当 $\eta=1$ 时，成交量的分布情况.....	6
图 5 不同 M 取值下，VCV 值与知情交易者比例 η 的关系图.....	7
图 6 VCV 因子 10 组分层收益与多空组合收益（全 A，日频数据，月频调仓）.....	8
图 7 VCV 因子 rank_IC 与累积 rank_IC（全 A，日频数据，月频调仓）.....	8
图 8 VCV 因子多空组合收益与基准（全 A，日频数据，月频调仓）.....	8

图 9 VCV 因子 10 组分层收益与多空组合收益（沪深 300，日频数据，月频调仓）	9
图 10 VCV 因子 rank_IC 与累积 rank_IC（沪深 300，日频数据，月频调仓）	9
图 11 VCV 因子多空组合收益与基准（沪深 300，日频数据，月频调仓）	9
图 12 VCV 因子 10 组分层收益与多空组合收益（中证 500，日频数据，月频调仓）	10
图 13 VCV 因子 rank_IC 与累积 rank_IC（中证 500，日频数据，月频调仓）	10
图 14 VCV 因子多空组合收益与基准（中证 500，日频数据，月频调仓）	10
图 15 VCV 因子 10 组分层收益与多空组合收益（中证 1000，日频数据，月频调仓）	11
图 16 VCV 因子 rank_IC 与累积 rank_IC（中证 1000，日频数据，月频调仓）	11
图 17 VCV 因子多空组合收益与基准（中证 1000，日频数据，月频调仓）	11
图 18 VCV 因子 10 组分层收益与多空组合收益（全 A，日频数据，周频调仓）	12
图 19 VCV 因子 rank_IC 与累积 rank_IC（全 A，日频数据，周频调仓）	12
图 20 VCV 因子多空组合收益与基准（全 A，日频数据，周频调仓）	12
图 21 VCV 因子 10 组分层收益与多空组合收益（沪深 300，日频数据，周频调仓）	13
图 22 VCV 因子 rank_IC 与累积 rank_IC（沪深 300，日频数据，周频调仓）	13
图 23 VCV 因子多空组合收益与基准（沪深 300，日频数据，周频调仓）	13
图 24 VCV 因子 10 组分层收益与多空组合收益（中证 500，日频数据，周频调仓）	14
图 25 VCV 因子 rank_IC 与累积 rank_IC（中证 500，日频数据，周频调仓）	14
图 26 VCV 因子多空组合收益与基准（中证 500，日频数据，周频调仓）	14
图 27 VCV 因子 10 组分层收益与多空组合收益（中证 1000，日频数据，周频调仓）	15
图 28 VCV 因子 rank_IC 与累积 rank_IC（中证 1000，日频数据，周频调仓）	15
图 29 VCV 因子多空组合收益与基准（中证 1000，日频数据，周频调仓）	15
图 30 VCV 因子 10 组分层收益与多空组合收益（中证 1000 五分钟数据 周频调仓）	16
图 31 VCV 因子 rank_IC 与累积 rank_IC（（中证 1000 五分钟数据 周频调仓）	16
图 32 VCV 因子多空组合收益与基准（中证 1000 五分钟数据 周频调仓）	16
图 33 VCV 因子 10 组分层收益与多空组合收益（中证 1000 一分钟数据 周频调仓）	17
图 34 VCV 因子 rank_IC 与累积 rank_IC（（中证 1000 一分钟数据 周频调仓）	17
图 35 VCV 因子多空组合收益与基准（中证 1000 一分钟数据 周频调仓）	17
图 36 各调仓和数据频率的 VCV 因子与大类风格因子之间的相关性（全 A）	18
图 37 各调仓和数据频率的 VCV 因子与大类风格因子之间的相关性（中证 1000）	19
图 38 模拟组合 2021 年以来的回测表现	19

一、寻觅知情交易者的蛛丝马迹

过往三年，海内外市场发生了一系列黑天鹅事件，不管从 A 股的反应还是成熟市场（比如美股）的表现，市场均存在反应不足和反应过度的现象，也充分说明了哪怕是美股，也远远不是一个完全有效的市场。既然不是完美有效的市场，那么总有一小部分的知情交易者，可以根据私有信息获得超越市场的回报。

市场微观结构理论认为，信息是影响市场上投资者交易行为的重要因素之一，投资者可以根据自己所拥有的信息，判断市场的走势并采取相应的交易策略而获利；另一方面，金融市场的信息在不同的投资者之间的分布又是不均衡的，市场中客观存在的信息不对称会导致逆向选择行为的发生，知情交易者可以利用自己的私有信息从交易对手方处获得利润，通过对市场中信息结构的刻画可以更好地认识和理解金融市场运行的过程。

知情交易者往往会隐匿交易行踪，我们无法通过直接观测的方式了解知情交易者的存在，或追踪其行踪。对于知情交易者的研究其实早在上个世纪末就被广泛关注，学者一直都在尝试用不同的代理变量来刻画知情交易者的行为，并且也取得了一些成果。

一个比较容易想到的线索思路是：当交易者不知情且流动性需求不相关时，大多数订单将相互抵消，因此与观察到的交易量相比，订单不平衡相对较低，在这种情况下，交易量服从正态分布。一旦有知情交易者加入到某只股票的交易中，那么这只股票“固有”的交易流的分布就会发生变化。就像是在原本平静的湖面上，扔下了一记石子。

很典型地，不少人工做短线的交易员，往往会通过 K 线的走势来判断一只股票是不是“庄股”，这其实就是判断某只股票的“固有”交易流是否发生变化的一种方式。

在学术界，学者们会通过构建理论模型，用数学的语言来描述知情交易者和非知情交易者之间行为的差异，然后再通过实际的交易数据来验证猜想，构建合适的指标。本报告中，我们利用交易量变异系数（Volume Coefficient of Variation, VCV）来刻画知情交易者存在的可能性大小。

我们认为知情交易的存在会影响个股后市的表现：

猜想之一：知情交易者的加入会增加市场对该只个股的关注度，过分炒作会偏离股票的真实价值，未来将利空。

猜想之二：知情交易者过多的股票被视为“妖股”或者“庄股”，不被机构等大型投资者(大资金)看好，对股票价值本身是一种伤害。

为了验证 VCV 指标在 A 股选股能力的有效性，我们设计指标在 A 股进行测试，探究知情交易是否会影响个股未来的收益，即探究 VCV 指标是不是一个有效的选股因子。

二、因子的构建

指标构建

参考 Lof, Matthijs & Jos van Bommel (2022) 的研究成果，设计了以下理论模型。我们用 M 代表所有的投资者（流动性寻求者），普通的投资者所提供的订单服从均值为 0，标准差为 σ 的正态分布，而买卖单之间的不平衡性由做市商（在美股市场）吸收。我们将 σ 称之为交易强度。假设有 η 比例的投资者为知情交易者，那么，知情交易者的人数则为 ηM 。这里我们先假设 η 是个外生变量。

我们将每个交易者的个人交易需求标记成 y_i ，若 y_i 为正，这表示买入， y_i 为负代表卖出。这时，买卖单的不平衡性（净订单流）可以表示成 $\sum_M y_i$ ，这一不平衡性由做市商来弥补。这时，全部的成交量可以写成：

$$V = \frac{1}{2} \left(\sum_M |y_i| + \left| \sum_M y_i \right| \right)$$

值得注意的是，小括号里面其实把买单和卖单都进行了统计，由于对成交量做了双倍的统计，因而前面要乘以 $\frac{1}{2}$ 。

知情交易者的订单是完全相关的，而 $(1-\eta)M$ 的非知情交易者的订单则是不相关的 (i.i.d.)，这个假设之下，订单流的不平衡服从：

$$\sum_M y_i \sim N(0, \sigma^2(\eta^2 M^2 + (1-\eta)M))$$

由于知情和不知情需求的不同相关性，订单不平衡的方差是 η 的非线性函数。如果大多数流动性寻求者是不知情的，他们的订单将大多相互匹配，订单失衡预计会相对较低。当许多交易者被告知时，他们的相关需求可能会导致严重的失衡。因此，订单失衡的标准差随着知情交易比例 η 的增加而增加。

利用半正态分布的属性，我们可以推导出：

$$\begin{aligned} E[V] &= \frac{1}{2} \left(E \left[\sum_M |y_i| \right] + E \left[\left| \sum_M y_i \right| \right] \right) \\ &= \frac{\sigma M}{\sqrt{2\pi}} \left(1 + \sqrt{\eta^2 + (1-\eta)M^{-1}} \right) \end{aligned}$$

当 M 很大的时候（交易者人数很多），人均交易量可以写成：

$$\lim_{M \rightarrow \infty} E \left[\frac{V}{M} \right] = \frac{\sigma}{\sqrt{2\pi}} (1 + \eta)$$

综合考虑知情的流动性需求者，不知情的流动性寻求者和流动性提供者，对于较大的 M 来说，人均成交量的方差可以写成与 η^2 相关的公式。

$$\lim_{M \rightarrow \infty} \text{Var} \left(\frac{V}{M} \right) = \sigma^2 \left(1 - \frac{2}{\pi} \right) \eta^2$$

上式可以看出，当 M 很大的时候，人均交易量的标准差与人均交易量的波动之比，是严格与 η 正相关，而与 σ 以及 M 并不相关。

$$\lim_{M \rightarrow \infty} \frac{\sigma_V}{\mu_V} = \sqrt{2\pi - 4} \frac{\eta}{\eta + 1} = \sqrt{2\pi - 4} \left(1 - \frac{1}{\eta + 1} \right)$$

交易量变异系数 (VCV) 作为知情交易比例的自然衡量标准出现。以下，我们用 VCV 作为因子，看其在不同股票池中的表现。

VCV 的月度公司级观测值是用日交易量的月标准差除以日交易量的月平均值来计算的。股票 i 在 t 月中的交易量变异系数定义为：

$$VCV_{i,t} \equiv \frac{\sigma_{V(i,t)}}{\mu_{V(i,t)}}$$

$VCV_{i,t}$ 指标是比较灵活的，可以根据需要来算不同时间间隔的 VCV 指标，比如我们常见的可以使用日频数据进行计算，当然也可以将其扩展到更高频的领域。

同时，对于上文提到的“成交量”的这个变量，其实也可以用其他变量，比如成交额、个股日成交量在全市场成交量中的占比等替换。为了避免除权等因素的变化对指标构建的影响，本报告中 y_i 用成交额进行计算。

模拟验证

关于上述猜想是否成立，实际情况中，我们无法通过直接观测市场来验证上述的推导，但是我们可以通过计算机模拟来验证知情交易者的比例 η 是否和 VCV 指标存在单调关系。

我们分析了不同的 η 值（知情交易的比例）和 M （流动性寻求者的数量）的组合下，以上模型产生的交易量的分布。从标准正态分布中抽取 $1 + (1-\eta)M$ 个随机观测值来模拟个人交易需求（即我们假设 $\sigma=1$ ）。第一个观测值乘以 ηM ，

代表总知情需求，其余观测值代表非知情交易者的个人需求。通过 $V = \frac{1}{2}(\sum_M |y_i| + |\sum_M y_i|)$ 计算交易量 V 。对于每对 (M, η) ，我们生成 T 个交易日观测样本，从中计算交易量变异系数 VCV 。

下图显示了流动性寻求者数量 $M=1000$ 时，不同 η 值下模拟的交易量的直方图。样本量 $T=1000$ 个交易日时段。在没有知情交易者的情况下 ($\eta = 0$)，成交量分布遵循一个略微倾斜的钟形曲线，而在存在知情交易者的情况下，成交量水平更高且更分散。可见，知情交易者的出现显著改变了“固有”的交易量分布的形态。

图 1 当 $\eta=0$ 时，成交量的分布情况

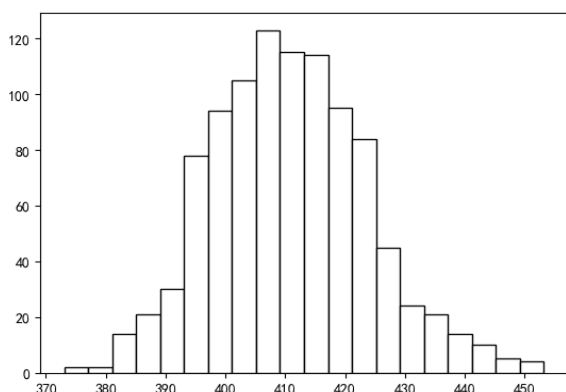
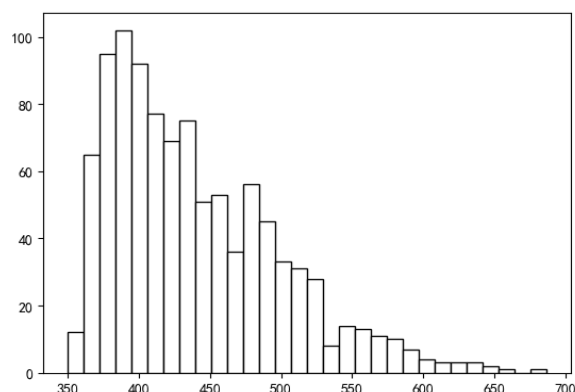


图 2 当 $\eta=0.1$ 时，成交量的分布情况



资料来源：招商证券量化团队

资料来源：招商证券量化团队

图 3 当 $\eta=0.5$ 时，成交量的分布情况

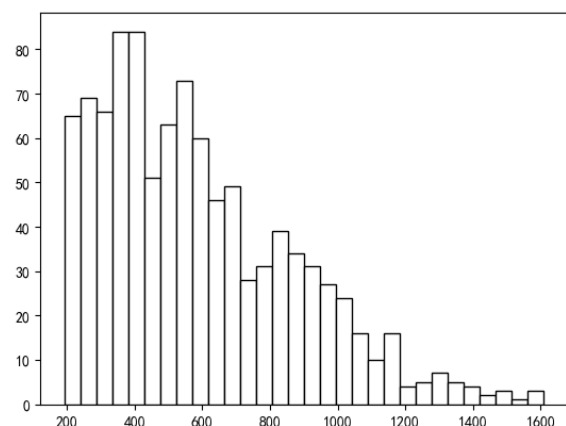
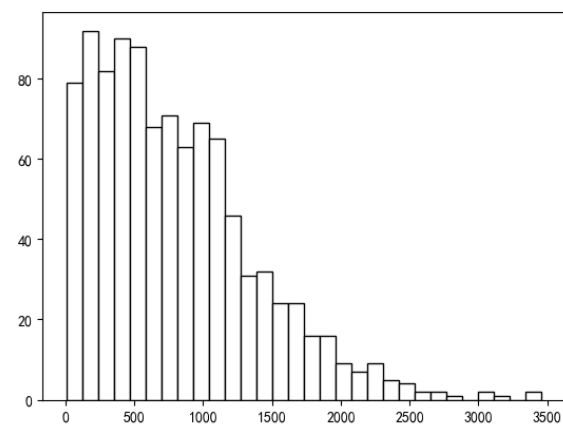


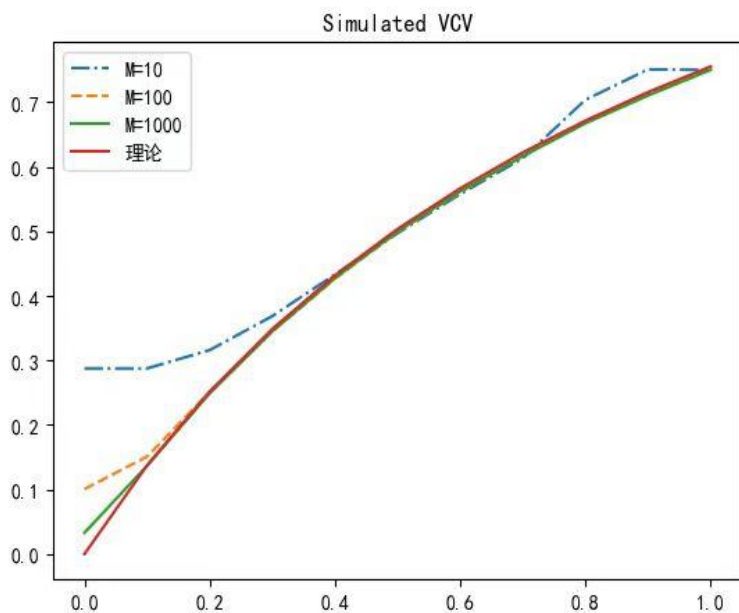
图 4 当 $\eta=1$ 时，成交量的分布情况



资料来源：招商证券量化团队

资料来源：招商证券量化团队

下图为不同的 η 和 M 值下，10000 次重复模拟的平均 VCV ，每次样本量 $T=100$ 个交易日时段。我们可以看到，只有当 M 和 η 都很低时，平均 VCV 才大幅偏离其渐近值。即使对于较小的 M ，平均 VCV 也随 η 严格递增。

图 5 不同 M 取值下，VCV 值与知情交易者比例 η 的关系图

资料来源：招商证券量化团队，横轴为 η ，纵轴为 VCV

三、因子测试与结果展示

月频因子表现

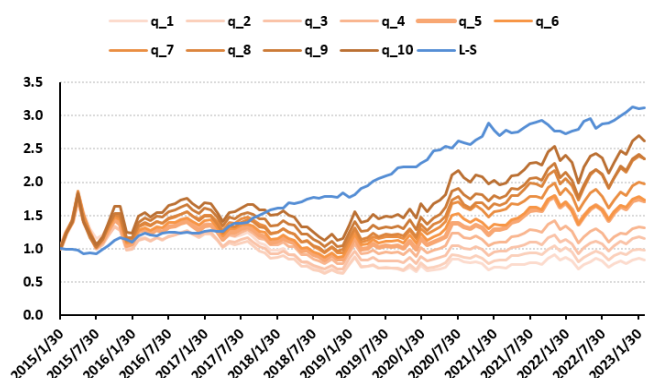
数据与因子计算

为了测试 VCV 因子在 A 股历史上的表现，回测方式如下：

1. 股票池：历史时点上的全 A 股（沪深 300、中证 500 和中证 1000），剔除 ST、PT 股票，剔除上市时间不足 3 个月的个股。
2. 回测数据区间：2015 年 1 月 1 日至 2023 年 3 月 31 日。
3. 截面期：每个自然月的最后一个交易日计算因子暴露度，分别计算出个股 i 的 $VCV_{i,m}$ 。
4. rank_IC 计算：个股截面因子暴露度序列与下个月个股收益率序列的次序相关系数（斯皮尔曼相关系数）。
5. 分层回测模型：则假定在次月第一个交易日调仓，根据个股的因子暴露度差异，将股票池个股分成 n 组，每月重新计算，求出每组的累积收益率，来考察因子在时间序列维度上对个股的区分能力。

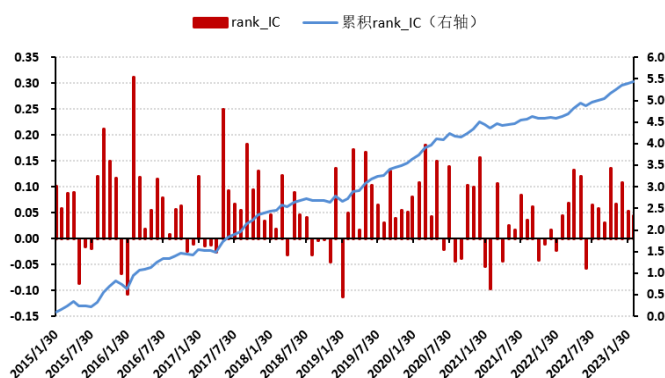
以下，所有因子的单因子测试均以上述计算方式为准。注意：VCV 因子是负向因子，也就是说，VCV 值越大，知情交易者占比越高，个股未来的走势越弱，为了因子展示结果更易于观察，这里因子取了 VCV 的倒数，故方向变为正向。

图 6 VCV 因子 10 组分层收益与多空组合收益（全 A，日频数据，月频调仓）



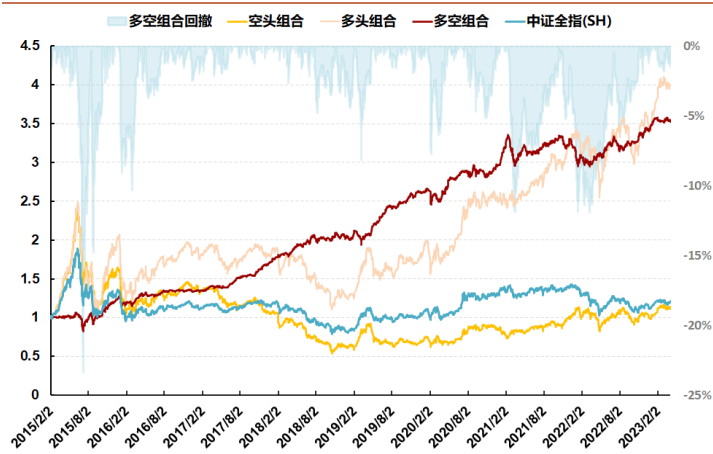
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 7 VCV 因子 rank_IC 与累积 rank_IC（全 A，日频数据，月频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 8 VCV 因子多空组合收益与基准（全 A，日频数据，月频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

表 1: VCV 因子回测数据展示（全 A，日频数据，月频调仓）

rank_IC 数据	rank_IC 均值	波动（年化）	IR(年化)	IC t 值	IC 胜率	最大值	最小值
多空组合	0.056	0.268	2.489	7.113	73.5%	0.312	-0.112
	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额	
	252.60%	16.32%	23.36%	299.99%	21.31%	16.20%	

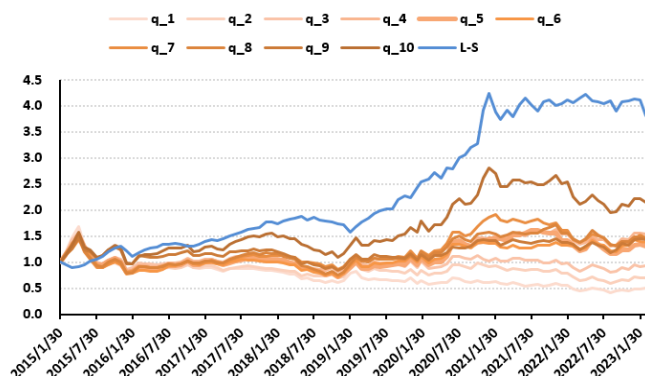
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

我们用 VCV 因子在 A 股进行了测试，整体来看，VCV 因子在全 A 范围内确实有比较不错的选股能力，根据过去近 8 年的回测，我们发现 VCV 的年化 IR 是 2.489，t 值为 7.113，在统计学上是显著的，胜率能达到 73.5%。

多空组合收益方面，整体而言，个股分层的单调性还属稳健，多空组合的总收益为 252.60%，年化收益为 16.32%，历史最大回撤为 23.36%。

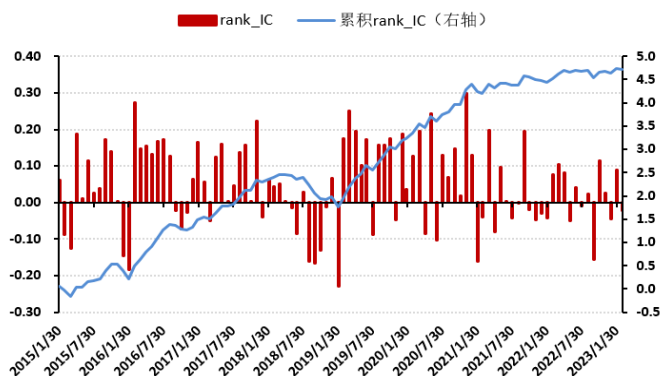
另外，我们也讨论了 VCV 因子在沪深 300、中证 500、中证 1000 股票池中的表现，结果如下：

图 9 VCV 因子 10 组分层收益与多空组合收益（沪深 300，日频数据，月频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 10 VCV 因子 rank_IC 与累积 rank_IC（沪深 300，日频数据，月频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 11 VCV 因子多空组合收益与基准（沪深 300，日频数据，月频调仓）



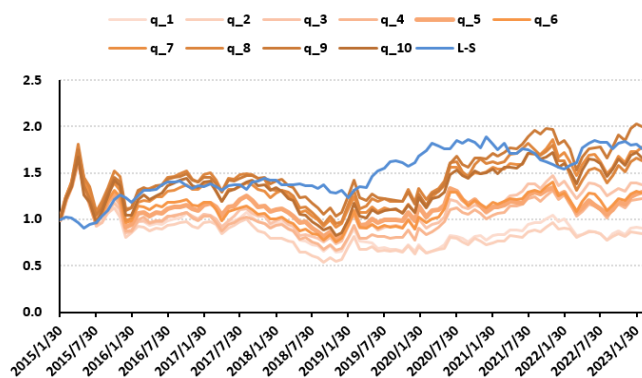
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

表 2: VCV 因子回测数据展示（沪深 300，日频数据，月频调仓）

rank_IC 数据	rank_IC 均值	波动 (年化)	IR(年化)	IC t 值	IC 胜率	最大值	最小值
多空组合	0.048	0.397	1.453	4.152	65.3%	0.298	-0.229
	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额	
	332.45%	19.20%	23.80%	205.31%	21.96%	12.24%	

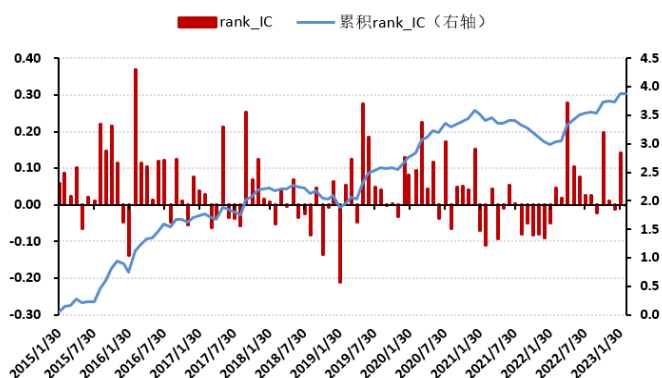
资料来源：Wind 资讯、招商证券，截止 2022 年 3 月 31 日

图 12 VCV 因子 10 组分层收益与多空组合收益（中证 500，日频数据，月频调仓）



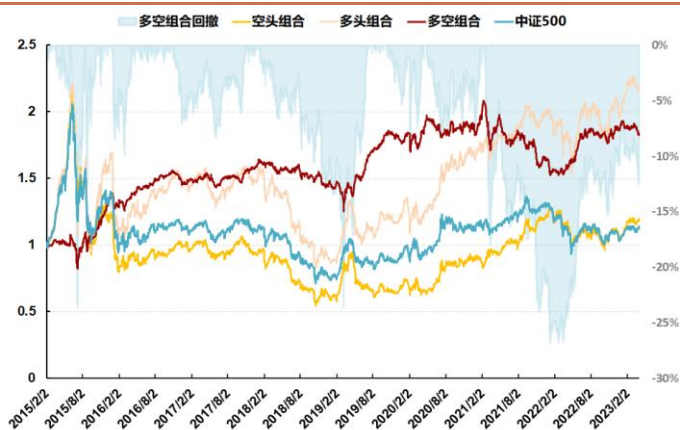
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 13 VCV 因子 rank_IC 与累积 rank_IC（中证 500，日频数据，月频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 14 VCV 因子多空组合收益与基准（中证 500，日频数据，月频调仓）



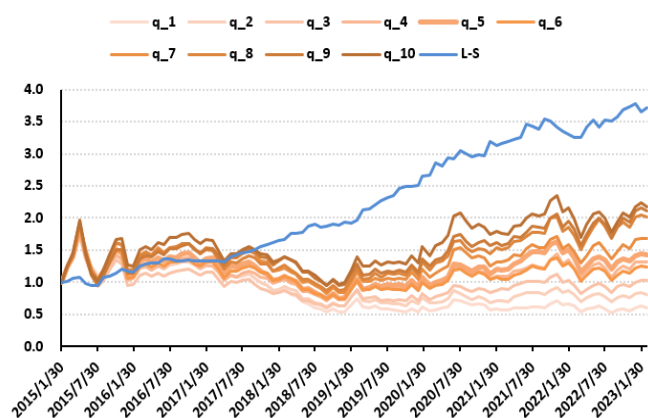
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

表 3: VCV 因子回测数据展示（中证 500，日频数据，月频调仓）

rank_IC 数据	rank_IC 均值	波动 (年化)	IR(年化)	ICt 值	IC 胜率	最大值	最小值
	0.040	0.350	1.355	3.873	63.3%	0.369	-0.210
多空组合	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额	
	82.15%	7.46%	26.86%	118.29%	14.34%	8.48%	

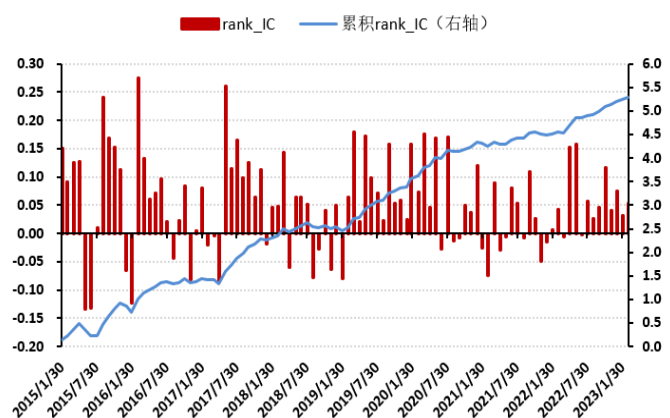
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 15 VCV 因子 10 组分层收益与多空组合收益（中证 1000，日频数据，月频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 16 VCV 因子 rank_IC 与累积 rank_IC（中证 1000，日频数据，月频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 17 VCV 因子多空组合收益与基准（中证 1000，日频数据，月频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

表 4: VCV 因子回测数据展示（中证 1000，日频数据，月频调仓）

rank_IC 数据	rank_IC 均值	波动 (年化)	IR (年化)	IC t 值	IC 胜率	最大值	最小值
	0.054	0.291	2.229	6.370	72.4%	0.275	-0.134
多空组合	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额	
	268.99%	16.96%	31.09%	213.99%	6.52%	14.57%	

资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

在沪深 300 股票池中，因子 rank_IC 为 0.048，IR 年化为 1.453，t 值为 4.152，胜率是 65.3%。多空组合方面过去 8 年时间，多空的收益为 332.45%，年化收益率为 19.20%。

在中证 500 股票池中，因子 rank_IC 为 0.04，IR 年化为 1.355，t 值为 3.873，胜率是 63.3%。多空组合方面过去 8 年时间，多空的收益为 82.15%，年化收益率为 7.46%。

在中证 1000 股票池中，因子 rank_IC 为 0.054，IR 年化为 2.229，t 值为 6.370，胜率是 72.4%。多空组合方面过去 8 年时间，多空的收益为 268.99%，年化收益率为 16.96%。

VCV 因子在小盘中的表现优于在中大盘上的表现，在中证 500 中的表现最弱，我们认为很可能跟近年来量化私募在中证 500 指增和中性产品的大量布局有关系。在中证 500 的成分上，量价类因子被交易得较为充分，因而超额最弱。

周频因子表现

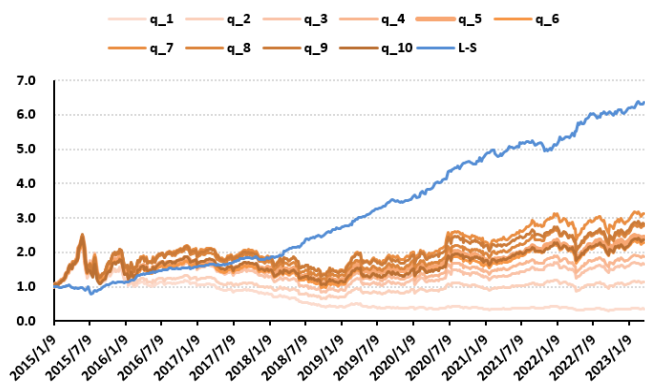
我们将该因子的换仓频率推向更高的维度，看看 VCV 因子是否在周频上会有更好的表现。

数据与因子计算

为了测试各因子在 A 股历史上的周频调仓表现，回测方式如下：

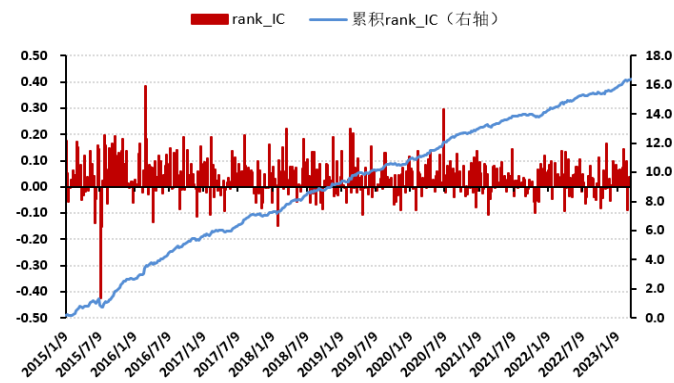
1. 股票池：历史时点上的全 A 股（沪深 300、中证 500 和中证 1000），剔除 ST、PT 股票，剔除上市时间不足 3 个月的个股。
2. 回测数据区间：2015 年 1 月 1 日至 2023 年 3 月 31 日。
3. 截面期：每个自然周的最后一个交易日计算因子暴露度，我们采用日频数据数据来计算 VCV 因子。
4. rank_IC 计算：个股截面因子暴露度序列与下一周个股收益率序列的次序相关系数（斯皮尔曼相关系数）。
5. 分层回测模型：则假定在次周第一个交易日调仓，根据个股的因子暴露度差异，将股票池个股分成 n 组，每周重新计算，求出每组的累积收益率，来考察因子在时间序列维度上对个股的区分能力。

图 18 VCV 因子 10 组分层收益与多空组合收益（全 A，日频数据，周频调仓）



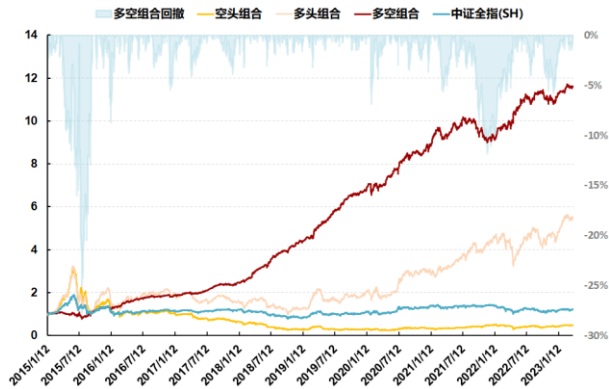
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 19 VCV 因子 rank_IC 与累积 rank_IC（全 A，日频数据，周频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 20 VCV 因子多空组合收益与基准（全 A，日频数据，周频调仓）



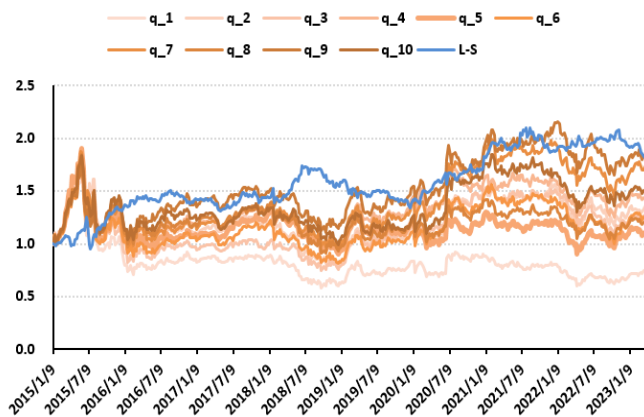
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

表 5: VCV 因子回测数据展示 (全 A, 日频数据, 周频调仓)

rank_IC 数据	rank_IC 均值	波动 (年化)	IR(年化)	IC t 值	IC 胜率	最大值	最小值
	0.039	0.257	1.839	10.825	70.7%	0.386	-0.423
多空组合	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额	
	1060.90%	28.87%	28.07%	451.92%	21.85%	20.76%	

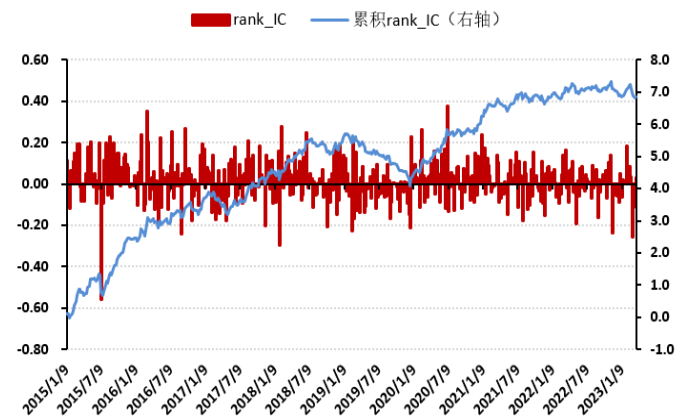
资料来源: Wind 资讯、招商证券, 截止 2023 年 3 月 31 日

图 21 VCV 因子 10 组分层收益与多空组合收益 (沪深 300, 日频数据, 周频调仓)



资料来源: Wind 资讯、招商证券, 截止 2023 年 3 月 31 日

图 22 VCV 因子 rank_IC 与累积 rank_IC (沪深 300, 日频数据, 周频调仓)



资料来源: Wind 资讯、招商证券, 截止 2023 年 3 月 31 日

图 23 VCV 因子多空组合收益与基准 (沪深 300, 日频数据, 周频调仓)



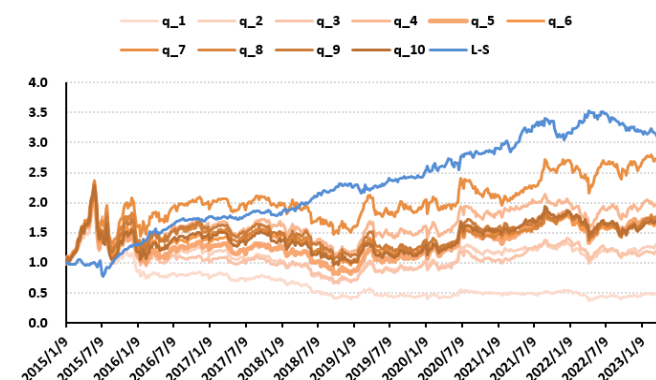
资料来源: Wind 资讯、招商证券, 截止 2023 年 3 月 31 日

表 6: VCV 因子回测数据展示 (沪深 300, 日频数据, 周频调仓)

rank_IC 数据	rank_IC 均值	波动 (年化)	IR(年化)	IC t 值	IC 胜率	最大值	最小值
	0.016	0.366	0.538	3.169	56.7%	0.376	-0.558
多空组合	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额	
	97.25%	7.28%	27.37%	99.41%	16.42%	6.95%	

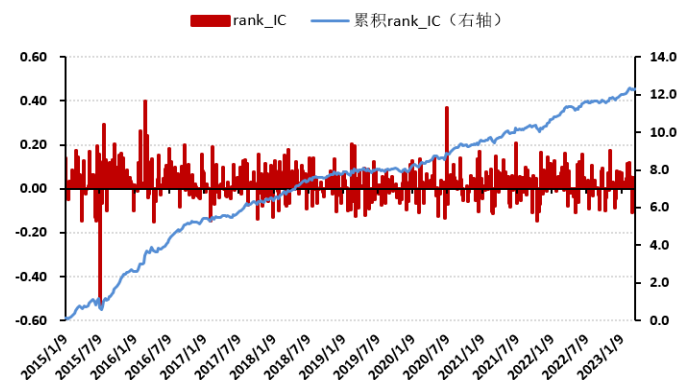
资料来源: Wind 资讯、招商证券, 截止 2023 年 3 月 31 日

图 24 VCV 因子 10 组分层收益与多空组合收益（中证 500，日频数据，周频调仓）



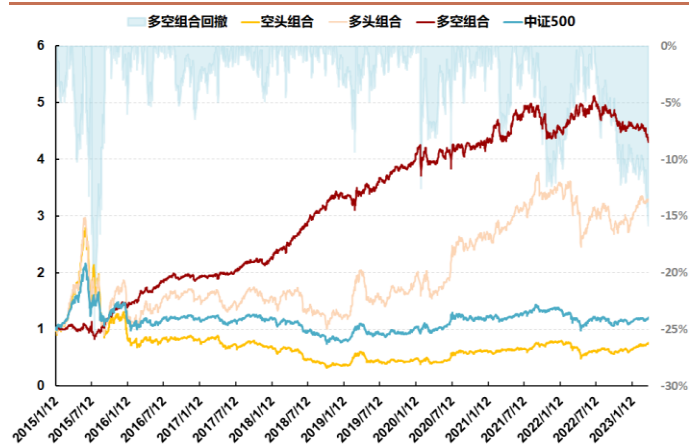
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 25 VCV 因子 rank_IC 与累积 rank_IC（中证 500，日频数据，周频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 26 VCV 因子多空组合收益与基准（中证 500，日频数据，周频调仓）



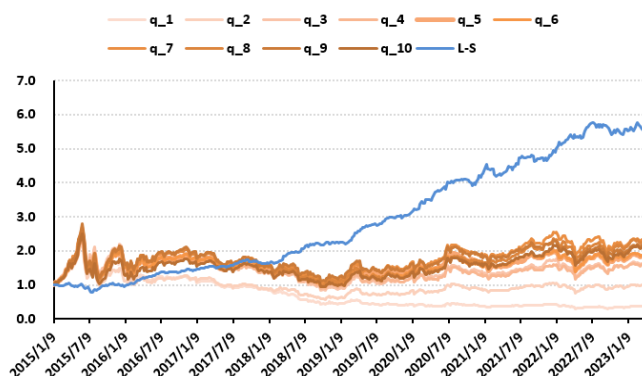
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

表 7：VCV 因子回测数据展示（中证 500，日频数据，周频调仓）

rank_IC 数据	rank_IC 均值	波动（年化）	IR(年化)	IC t 值	IC 胜率	最大值	最小值
	0.030	0.302	1.173	6.905	64.9%	0.398	-0.528
多空组合	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额	
	330.2%	16.3%	25.8%	229.92%	20.28%	13.43%	

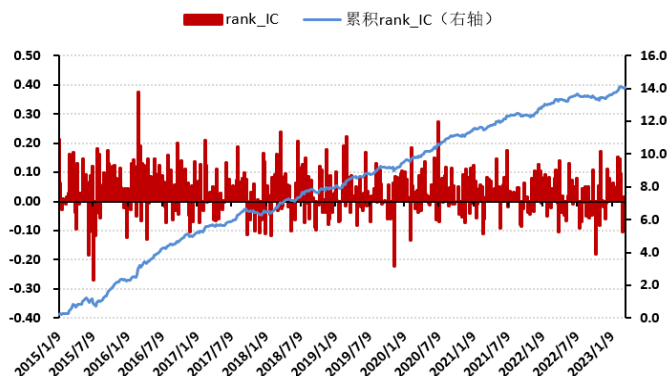
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 27 VCV 因子 10 组分层收益与多空组合收益（中证 1000，日频数据，周频调仓）



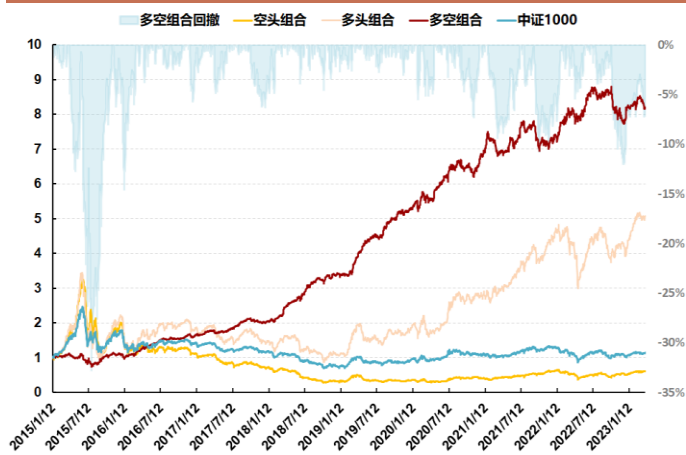
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 28 VCV 因子 rank_IC 与累积 rank_IC（中证 1000，日频数据，周频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 29 VCV 因子多空组合收益与基准（中证 1000，日频数据，周频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

表 8：VCV 因子回测数据展示（中证 1000，日频数据，周频调仓）

rank_IC 数据	rank_IC 均值	波动（年化）	IR(年化)	IC t 值	IC 胜率	最大值	最小值
	0.034	0.276	1.467	8.636	66.8%	0.376	-0.271
多空组合	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额	
	719.4%	24.3%	32.7%	407.97%	14.84%	20.40%	

资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

当我们将日频数据构造的因子调仓频率调整成周频之后，多空组合的收益有明显的提升。自 2015 年以来，全 A 股票池中，多空组合累积收益为 1060.90%；沪深 300 股票池中，多空累积收益为 97.25%；中证 500 股票池中，多空累积收益为 330.2%，中证 1000 股票池中则为 719.4%。整体看，依然是在小盘中的表现更优，也符合量价因子的典型特征。

四、日内数据构造因子测试

利用 5 分钟频率数据构建 VCV 因子

正如前文所述，VCV 指标比较灵活，可以采用不同的数据频率进行计算。前面我们已经测试了利用日频数据进行计算，以下，我们使用日内数据，比如 5 分钟 bar 数据和 1 分钟的 bar 数据对因子进行构造。

具体计算中，我们利用 5 分钟或者 1 分钟的数据来计算每周的成交额均值和标准差，并用此来计算 VCV 因子。

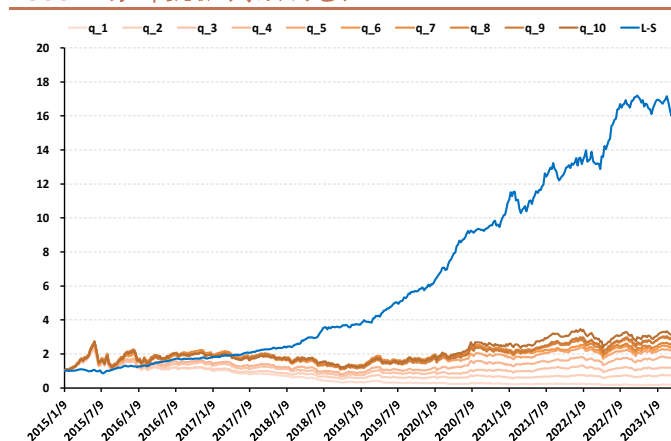
用 5 分钟数据计算 VCV 进行周频调仓，回测结果如下：

表 9：5 分钟数据计算的 VCV 因子周频调仓关键数据展示

股票池	rank_IC					多空组合		
	rank_IC 均值	波动（年化）	IR(年化)	IC t 值	IC 胜率	总收益	年化收益	历史最大回撤
全 A	0.053	0.331	1.911	11.305	71.0%	2566.6%	40.4%	31.8%
沪深 300	0.041	0.437	1.118	6.612	63.1%	1187.42%	30.19%	31.39%
中证 500	0.050	0.379	1.573	9.305	67.6%	1711.94%	34.87%	29.09%
中证 1000	0.052	0.351	1.789	10.583	68.8%	2884.26%	42.00%	23.80%

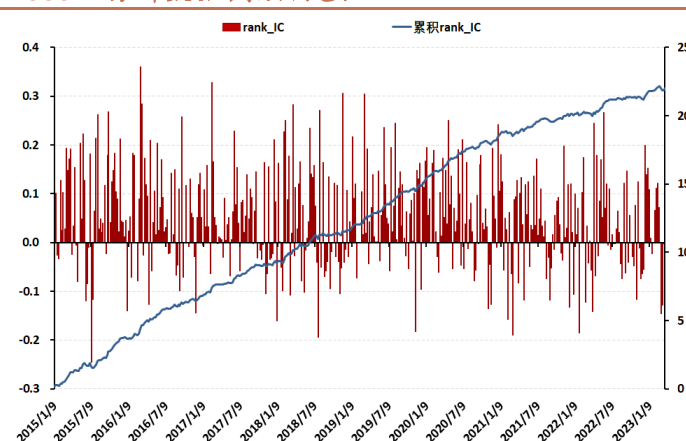
资料来源：米筐、招商证券，回测期：2015 年 1 月 1 日至 2023 年 3 月 31 日

图 30 VCV 因子 10 组分层收益与多空组合收益（中证 1000 五分钟数据 周频调仓）



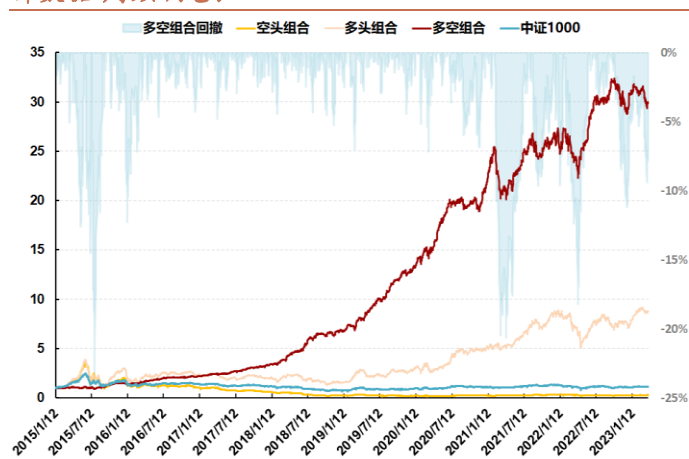
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 31 VCV 因子 rank_IC 与累积 rank_IC（（中证 1000 五分钟数据 周频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 32 VCV 因子多空组合收益与基准（中证 1000 五分钟数据 周频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

表 10：VCV 因子回测数据展示（中证 1000 五分钟数据 周频调仓）

rank_IC 数据	rank_IC 均值	波动（年化）	IR(年化)	IC t 值	IC 胜率	最大值	最小值
	0.052	0.351	1.789	10.583	68.8%	0.360	-0.246

多空组合	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额
	2884.26%	42.00%	23.80%	783.34%	14.84%	29.02%

资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

高频数据构建的 VCV 因子在不同股票池中的表现均优于日频数据构建的因子。中证 1000 中的表现较为突出，自 2015 年以来，VCV 因子的 rank_IC 为 0.052，年化 IR 为 1.789，IC 的 t 值为 10.583，胜率为 68.8%。多空组合的总收益为 2884.26%，年化收益高达 42%。

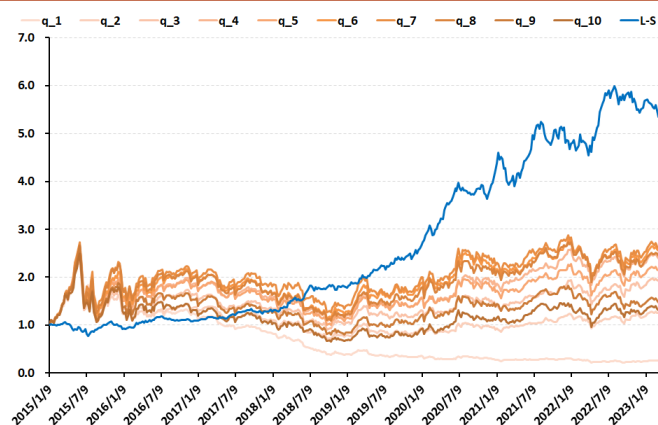
利用 1 分钟频率数据构建 VCV 因子

表 11：1 分钟数据计算的 VCV 因子周频调仓关键数据展示

股票池	rank_IC					多空组合		
	rank_IC 均值	波动（年化）	IR(年化)	IC t 值	IC 胜率	总收益	年化收益	历史最大回撤
全 A	0.017	0.397	0.501	2.960	54.4%	557.99%	21.54%	39.89%
沪深 300	0.024	0.502	0.584	3.453	57.3%	527.26%	20.93%	39.44%
中证 500	0.026	0.478	0.648	3.831	56.3%	697.02%	23.97%	32.21%
中证 1000	0.022	0.433	0.605	3.577	56.3%	860.92%	26.39%	27.08%

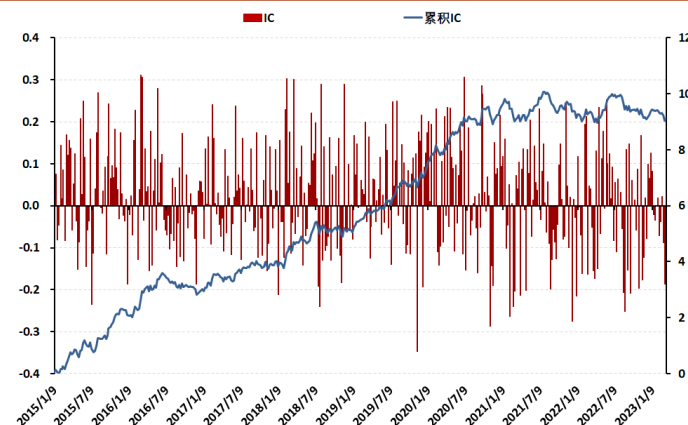
资料来源：米筐、招商证券，回测期：2015 年 1 月 1 日至 2023 年 3 月 31 日

图 33 VCV 因子 10 组分层收益与多空组合收益（中证 1000 一分钟数据 周频调仓）



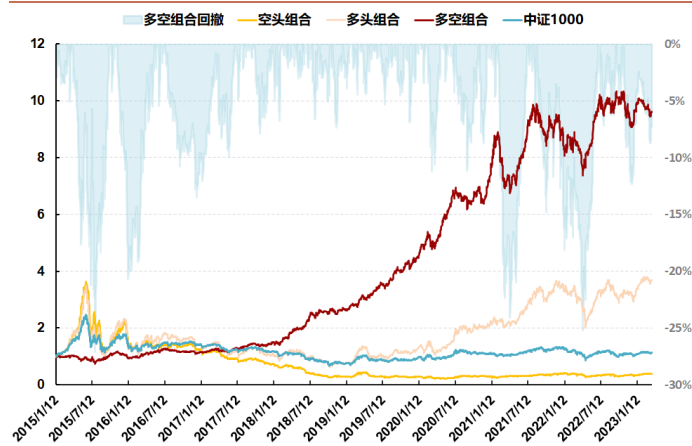
资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 34 VCV 因子 rank_IC 与累积 rank_IC（（中证 1000 一分钟数据 周频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 35 VCV 因子多空组合收益与基准（中证 1000 一分钟数据 周频调仓）



资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 31 日

表 12: VCV 因子回测数据展示（中证 1000 一分钟数据 周频调仓）

rank_IC	rank_IC 均值	波动（年化）	IR(年化)	IC t 值	IC 胜率	最大值	最小值
数据	0.022	0.433	0.605	3.577	56.3%	0.311	-0.347
多空组合	总收益	年化收益	历史最大回撤	多头总收益	基准总收益	多头年化超额	
	860.92%	26.39%	27.08%	269.94%	13.60%	14.06%	

资料来源：Wind 资讯、招商证券，截止 2023 年 3 月 27 日

VCV 因子与大类风格因子相关性展示

我们展示了在全 A 股票池和中证 1000 股票池中，不同方式构建的 VCV 因子与大类风格因子之间的相关性。总体而言，VCV 因子与大类风格因子之间的相关性并不是很高。其中，日频数据构建的因子，不论是在月频调仓还是周频调仓的维度下，其相关性均较低；而高频数据构建的因子与大类风格因子之间的相关性略高。

图 36 各调仓和数据频率的 VCV 因子与大类风格因子之间的相关性（全 A）

	月频调仓	周频调仓因子		
	日频数据	日频数据	5分钟数据	1分钟数据
贝塔	0.123	0.050	0.130	0.215
账面市值比	0.034	0.019	0.013	-0.087
盈利率	0.074	0.019	0.117	0.065
成长	0.027	-0.001	0.060	0.081
杠杆	0.005	-0.002	0.002	-0.009
流动性	-0.072	0.024	-0.012	0.167
动量	0.066	0.019	0.116	0.209
非线性市值	-0.004	-0.016	0.072	0.129
残差波动率	-0.257	-0.071	-0.178	0.001
规模	0.125	0.013	0.259	0.368

资料来源：米筐、Wind 资讯、招商证券，截止 2023 年 3 月 31 日

图 37 各调仓和数据频率的 VCV 因子与大类风格因子之间的相关性（中证 1000）

	月频调仓	周频调仓因子		
	日频数据	日频数据	5分钟数据	1分钟数据
贝塔	0.151	0.050	0.164	0.264
账面市值比	-0.005	0.014	-0.059	-0.188
盈利率	0.039	0.013	0.054	-0.023
成长	0.027	-0.006	0.060	0.081
杠杆	-0.030	-0.010	-0.069	-0.108
流动性	-0.007	0.039	0.074	0.305
动量	0.087	0.021	0.125	0.223
非线性市值	-0.050	-0.025	-0.010	0.008
残差波动率	-0.221	-0.063	-0.146	0.051
规模	0.065	-0.009	0.116	0.222

资料来源：米筐、Wind 资讯、招商证券，截止 2023 年 3 月 31 日

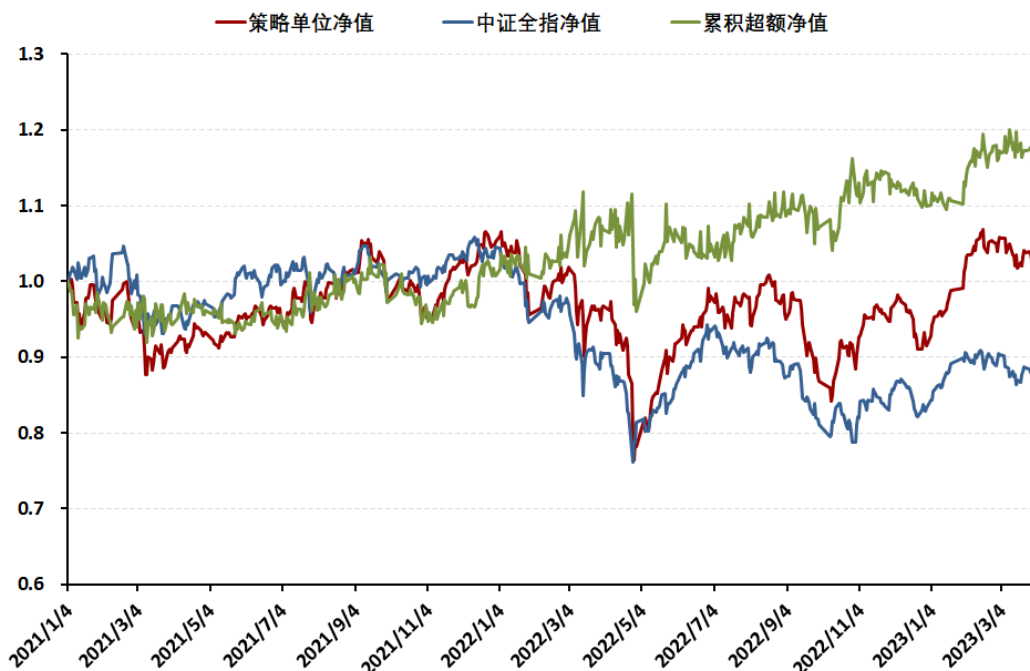
五、模拟组合构建与回测

我们已经对 VCV 因子做了各方面的测试，最后我们模拟实际交易环境，构建模拟组合，对该因子进行测试。以 VCV 因子为例，在中证全指的成分股中进行模拟组合的构建。具体方式如下：

模拟账户初始资金为 1 亿元，按照月底最后一个交易日计算个股因子暴露度，从中证全指(000985)的成分股中，挑选出 VCV 因子暴露度最高的 100 只个股，进行等权组成模拟组合。次月第一个交易日进行交易，设置单边交易成本为 3‰，根据最新持仓目标对组合进行再平衡。充分考虑当天个股停牌、涨跌停等无法交易的情况，若无法在当天交易，则将交易指令顺延到下一个交易日执行，直到撮合为止。

将中证全指作为基准，观察 2021 年以来的模拟组合实际表现：

图 38 模拟组合 2021 年以来的回测表现



资料来源：米筐、招商证券，回测期：2021 年 1 月 4 日至 2023 年 03 月 31 日

表 13：模拟组合 2021 年以来回测关键指标

组合收益	最大回撤	基准收益	超额累积收益	超额最大回撤
2.31%	28.24%	-9.74%	12.05%	8.78%

资料来源：米筐、招商证券，回测期：2021 年 1 月 4 日至 2023 年 03 月 31 日

在模拟实际情况之后，我们发现，即使在交易成本较高且资金没有办法充分利用的情况下，VCV 因子从 2021 年以来的表现仍然较为优秀，组合收益为 2.31%，同期的基准收益为-9.74%，累积超额收益为 12.05%。

六、总结

不难发现，不论是在 A 股还是在相对成熟的美股市场均存在市场对突发事件反应不足和反应过度的问题，市场远没有达到有完全有效。因而市场上依然存在知情交易者，可以利用自己的私有信息来获取超越市场的回报。对于知情交易者的研究一直是学术界和业界关注的问题。

我们无法通过直接观测的方式来获得知情交易者在市场上的行踪，但是，知情交易者的出现很可能打破原有交易的既定规律，从而我们可以通过观测交易流的不平衡性来推导知情交易者存在的概率。参考 Lof, Matthijs & Jos van Bommel (2022) 的研究成果，设计了 VCV 指标，并用 VCV 指标构建因子在不同的 A 股股票池进行选股测试。

VCV 因子比较灵活，可以在不同的时间维度上进行构建。首先我们用日频数据进行因子暴露度的计算，并尝试每月调仓和每周调仓。我们发现，VCV 因子在全 A 范围内确实有比较不错的选股能力，根据过去近 8 年的回测，我们发现 VCV 的年化 IR 是 2.489，t 值为 7.113，在统计学上是显著的，胜率能达到 73.5%。多空组合收益方面，多空组合的总收益为 252.60%，年化收益为 16.32%。在中证 1000 股票池中的表现较大盘股票池更为优秀。

我们又利用 5 分钟数据和 1 分钟数据对因子暴露度进行计算，并且在周频调仓下进行测试。高频数据构建的 VCV 因子在不同股票池中的表现均优于日频数据构建的因子。中证 1000 中的表现较为突出，自 2015 年以来，5 分钟数据构建的 VCV 因子 rank_IC 为 0.052，年化 IR 为 1.789，IC 的 t 值为 10.583，胜率为 68.8%。多空组合的总收益为 2884.26%，年化收益高达 42.00%。

最后，综合考量换手率和交易成本，我们模拟实际的交易环境，对 VCV 因子进行月频调仓的测试。在模拟实际情况之后，我们发现，即使在交易成本较高且资金没有办法充分利用的情况下，VCV 因子从 2021 年以来的表现仍然较为优秀，组合收益为 2.31%，同期的基准收益为-9.74%，累积超额收益为 12.05%。

风险提示：本报告数据基于历史数据进行测试，如遇到较为极端的市場变化，可能会有模型失效的风险。

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：首席分析师，量化团队负责人，管理学硕士，18 年证券研究经验，2010 年、2015 年、2016、2017、2020、2021 年、2022 年新财富最佳分析师。在量化选股择时、基金研究以及衍生品投资方面均有深入独到的见解。

崔浩瀚：量化分析师，浙江大学经济学硕士，6 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。

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。