

因子拥挤度专题

行业拥挤度

科大财经 2023年10月29日

大纲:

1. 华泰行业拥挤度指标
2. 国君行业拥挤度指标
3. 拥挤度有效性检验统计
4. 复合行业拥挤度收益模型表现

行业拥挤度

构建拥挤度指标的初衷是捕捉市场可能出现的交易过热状态。直观上股市最显而易见的拥挤特征就是成交量和价格的大幅提升。除此之外，市场波动率增大、量价变化趋势不匹配等异常现象，也表明当前市场风险正在累积，可能存在交易拥挤。本研究主要从以下几个方面进行拥挤度指标的构建，对市场的交易风险给出预警：

- 1. 判断行业指数价格或是流动性是否巨幅增长，相应指标：动量、流动性和乖离率指标
- 2. 判断行业指数的成交量和价格走势是否相吻合，相应指标：量价相关性指标
- 3. 判断当前市场是否存在较大波动，相应指标：波动率指标
- 4. 判断近期行业指数的收益率分布是否偏离统计规律，相应指标：分布特征指标

图表5： 本研究用于刻画行业拥挤度的量价指标



资料来源:华泰证券研究所

华泰行业拥挤度

图表9： 本研究构建的六类总计 20 项行业拥挤度指标

类别	指标名称	指标代码	判断是否拥挤的标准	窗口期长度（交易日）	历史分位数阈值
动量指标	普通动量拥挤度	normal_momentum_crowd	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
	夏普动量拥挤度	sharpe_momentum_crowd	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
	信息比率拥挤度	information_momentum_crowd	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
	路径调整动量拥挤度	distance_momentum_crowd	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
	尾部动量拥挤度	tail_momentum_crowd	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
流动性指标	成交量拥挤度	volume_crowd	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
	成交额拥挤度	amt_crowd	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
	换手率拥挤度	tum_crowd	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
乖离率指标	收盘价乖离率	close_bias	分位数是否处于历史高位	20/40/60/120/250	80%/90%/95%/99%
	成交量乖离率	volume_bias	分位数是否处于历史高位	20/40/60/120/250	80%/90%/95%/99%
	成交额乖离率	amt_bias	分位数是否处于历史高位	20/40/60/120/250	80%/90%/95%/99%
	换手率乖离率	tum_bias	分位数是否处于历史高位	20/40/60/120/250	80%/90%/95%/99%
量价相关性	成交量与收盘价的相关系数	corr_volume_close	分位数是否处于历史低位	20/40/60	1%/5%/10%/20%
	换手率与收盘价相关系数	corr_tum_close	分位数是否处于历史低位	20/40/60	1%/5%/10%/20%
	成交额与收盘价相关系数	corr_amt_close	分位数是否处于历史低位	20/40/60	1%/5%/10%/20%
波动率指标	波动率拥挤度	vol	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
	下行波动率拥挤度	downvol	分位数是否处于历史高位	20/40/60	80%/90%/95%/99%
分布特征指标	峰度	kurtosis	分位数是否处于历史低位	20/40/60	1%/5%/10%/20%
	偏度	skewness	分位数是否处于历史低位	20/40/60	1%/5%/10%/20%
	协偏度	coskewness	分位数是否处于历史低位	20/40/60	1%/5%/10%/20%

资料来源：华泰证券研究所

国泰君安行业拥挤度

表 1: 指标参数设置

指标类别	指标名称	指标代码	指标方向	计算参数
成交额	成交额占比	volume_proportion	正向	0/5/10
市值	市值占比	market_value_proportion	正向	0/5/10
偏度	超额收益偏度	excess_ret_skew	正向	20/40/60
量价相关系数	成交额-股价相关系数	price_volume_corr	负向	20/40/60
	成交额占比-市值占比相关系数	marketvalue_volume_proportion_corr	负向	20/40/60
波动率	超额收益波动率	excess_ret_vol	正向	20/40/60
峰度	超额收益峰度	excess_ret_kurt	正向	20/40/60
换手率	换手率	turnover	正向	0/5/10
	换手率占比	turnover_proportion	正向	0/5/10
配对相关性	收益率配对相关性	pair_corr_ret	负向	20/40/60
	成交额配对相关性	pair_corr_volume	负向	20/40/60
	换手率配对相关性	pair_corr_turnover	负向	20/40/60
买卖非均衡指标	机构净流入	inst_diff	正向	20/40/60
	散户净流入	indi_diff	正向	20/40/60
	机构主动净买入	inst_diff_act	正向	20/40/60
	散户主动净买入	indi_diff_act	正向	20/40/60
	行业整体主动净买入	total_diff_act	正向	20/40/60

数据来源：国泰君安证券研究

图表6： 本研究采用的动量指标

类型	计算公式
普通动量	$Mom_{t,K} = \frac{Close_{t,K}}{Close_{t-N+1,K}} - 1$
夏普比率	$Sharpe_{t,K} = \frac{\sum_{i=0}^{N-1} r_{t-i,K}/N}{std_{t,K}}$ $std_{t,K}^2 = \sum_{i=0}^{N-1} \frac{(r_{t-i,K} - \bar{r}_{t,K})^2}{N-1}, \bar{r}_{t,K} = \frac{\sum_{i=0}^{N-1} r_{t-i,K}}{N}$
信息比率	$IR_{t,K} = \frac{\sum_{i=0}^{N-1} (r_{t-i,K} - r_{t-i,m})/N}{TE_{t,K}}$ $TE_{t,K}^2 = \sum_{i=0}^{N-1} \frac{(r_{t-i,K} - r_{t-i,m} - \bar{r}'_{t,K})^2}{N-1}, \bar{r}'_{t,K} = \frac{\sum_{i=0}^{N-1} r_{t-i,m}}{N}$
路程调整动量	$Distance_{t,K} = \frac{Mom_{t,K}}{\sum_{i=0}^{N-1} r_{t-i,K} }$
尾部动量	$Tail_{t,K} = VaR_{t,K}$

说明：K表示行业，t为交易日，N为窗口期，r表示收益率，M为全行业数，VaR取下1%分位数

资料来源：华泰证券研究所

VaR计算：<https://blog.csdn.net/mfsdmlove/article/details/126081926>

乖离率指标

乖离率是较为常见的一种技术指标，通常是指最新市场价格与近期平均值之间的偏离度，以t时刻K行业的收盘价乖离率（窗口期N）为例：

$$close_bias_{t,K} = \frac{close_{t,K} - \sum_{i=0}^{N-1} close_{t-i,K}/N}{\sum_{i=0}^{N-1} close_{t-i,K}/N}$$

乖离率越大，说明当前市场价格超出过去N天的平均价越高，当前市场出现了剧烈的价格抬升，呈超买状态。在乖离率较大时，价格出现回弹的可能性更大。本研究对乖离率的做法做出延伸，计算了成交量、成交额和换手率的乖离率，分析市场最新的流动性指标是否与平均值出现了明显的偏离。收盘价或流动性指标的乖离率较大时，说明市场成交活跃度非常高，量和价都超出了均衡水平，此时可以考虑降低仓位，规避风险。

$$volume_bias_{t,K} = \frac{volume_{t,K} - \sum_{i=0}^{N-1} volume_{t-i,K}/N}{\sum_{i=0}^{N-1} volume_{t-i,K}/N}$$

$$amt_bias_{t,K} = \frac{amt_{t,K} - \sum_{i=0}^{N-1} amt_{t-i,K}/N}{\sum_{i=0}^{N-1} amt_{t-i,K}/N}$$

$$turn_bias_{t,K} = \frac{turn_{t,K} - \sum_{i=0}^{N-1} turn_{t-i,K}/N}{\sum_{i=0}^{N-1} turn_{t-i,K}/N}$$

波动率指标

波动率就是股票价格的风险，一般用股票收益率的标准差进行表征。股票市场普遍存在低波动异常，长期来看低波动率的股票相对高波动率的股票具有更高收益。从拥挤度角度来看，如果行业指数短期的波动率急剧上升，说明股票风险加大，在投资过程中应该规避。行业K在t时刻的波动率可以表示为：

$$vol_{t,K} = \sqrt{\sum_{i=0}^{N-1} \frac{(r_{t-i,K} - \bar{r}_{t,K})^2}{N-1}}$$

$$\bar{r}_{t,K} = \frac{\sum_{i=0}^{N-1} r_{t-i,K}}{N}$$

波动率还可以进一步分为上行波动率和下行波动率，两者的区别在于统计正收益率还是负收益率的标准差。从定义来看，下行波动率指标关注过去一段时间市场下行风险，更贴近我们对于拥挤度指标的定义。因此本研究还基于下行波动率进行拥挤度指标构建：

$$down_vol_{t,K} = \sqrt{\sum_{i=0}^{n_d-1} \frac{(r_{t-i,K} - \bar{r}_{t,K})^2}{n_d-1}}$$

$$\bar{r}_{t,K} = \frac{\sum_{i=0}^{N-1} r_{t-i,K}}{N}$$

其中d代表日收益率小于0的日期集合，n_d为日收益率小于0总天数。

量价相关系数

本研究采用行业指数收盘价和流动性指标的相关系数来定量分析量价背离现象。量价相关系数越大，说明量价指标的走势越接近，反之则表示两者出现背离。举例来说，行业 K 在时刻 t 的成交量与收盘价相关系数可以表示为：

$$\text{corr_volume_close}_{t,K} = \frac{\sum_{i=0}^{N-1} (\text{close}_{t-i,K} - \overline{\text{close}}_{t,K})(\text{volume}_{t-i,K} - \overline{\text{volume}}_{t,K})}{\sum_{i=0}^{N-1} (\text{close}_{t-i,K} - \overline{\text{close}}_{t,K})^2 \sum_{i=0}^{N-1} (\text{volume}_{t-i,K} - \overline{\text{volume}}_{t,K})^2}$$
$$\overline{\text{close}}_{t,K} = \sum_{i=0}^{N-1} \text{close}_{t-i,K} / N$$
$$\overline{\text{volume}}_{t,K} = \sum_{i=0}^{N-1} \text{volume}_{t-i,K} / N$$

当股价上涨时，如果某个行业指数的量价相关系数下降，意味着出现了量价背离的现象。当量价相关系数处于历史低位时，需要警惕可能出现的下跌风险。我们还计算了收盘价和成交额以及换手率的相关系数对量价背离现象进行分析：

$$\text{corr_amt_close}_{t,K} = \frac{\sum_{i=0}^{N-1} (\text{close}_{t-i,K} - \overline{\text{close}}_{t,K})(\text{amt}_{t-i,K} - \overline{\text{amt}}_{t,K})}{\sum_{i=0}^{N-1} (\text{close}_{t-i,K} - \overline{\text{close}}_{t,K})^2 \sum_{i=0}^{N-1} (\text{amt}_{t-i,K} - \overline{\text{amt}}_{t,K})^2}$$
$$\text{corr_turn_close}_{t,K} = \frac{\sum_{i=0}^{N-1} (\text{close}_{t-i,K} - \overline{\text{close}}_{t,K})(\text{turn}_{t-i,K} - \overline{\text{turn}}_{t,K})}{\sum_{i=0}^{N-1} (\text{close}_{t-i,K} - \overline{\text{close}}_{t,K})^2 \sum_{i=0}^{N-1} (\text{turn}_{t-i,K} - \overline{\text{turn}}_{t,K})^2}$$

分布特征指标

分布特征指标一般用于描述行业指数历史收益率的概率分布情况。通过峰度和偏度两种指标可以分析近期行业指数日收益率是否出现异常分布，进而判断当前是否处于拥挤状态。

峰度用来描述数据概率分布顶端尖峭或扁平的程度。峰度越大，代表分布越集中；峰度越小，分布越平滑。如果过去一段时间日收益率峰度越低，说明收益率取值较为分散，短期市场波动率更高，风险更大，拥挤度更高。行业 K 在 t 时刻的峰度（kurtosis）为：

$$\text{kurt}_{t,K} = \frac{\frac{1}{N} \sum_{i=0}^{N-1} (r_{t-i,K} - \bar{r}_{t,K})^4}{(\frac{1}{N} \sum_{i=0}^{N-1} (r_{t-i,K} - \bar{r}_{t,K})^2)^2}$$

偏度用来度量数据概率分布的偏斜方向和程度。当偏度大于零时，数据处于正偏态，概率密度呈右侧长尾分布；当偏度小于零时，数据处于负偏态，概率密度呈左侧长尾分布。如果过去一段时间行业指数收益率序列处于负偏态时，意味着近期行业指数出现了非常态的低收益率，有可能存在较大风险。行业 K 在 t 时刻的偏度（skewness）为：

$$\text{skew}_{t,K} = \frac{\frac{1}{N} \sum_{i=0}^{N-1} (r_{t-i,K} - \bar{r}_{t,K})^3}{(\frac{1}{N} \sum_{i=0}^{N-1} (r_{t-i,K} - \bar{r}_{t,K})^2)^{3/2}}$$

进一步地，我们还引入协偏度（coskewness）考察行业收益率相比于全市场整体收益率概率分布的偏离程度，行业 K 在 t 时刻相对于市场基准的协偏度为：

$$\text{CS}_{t,K} = \frac{\sum_{i=0}^{N-1} [(r_{t-i,K} - \bar{r}_{t,K})(r_{t-i,m} - \bar{r}_{t,m})^2]}{\sum_{i=0}^{N-1} [(r_{t-i,m} - \bar{r}_{t,m})^3]}$$

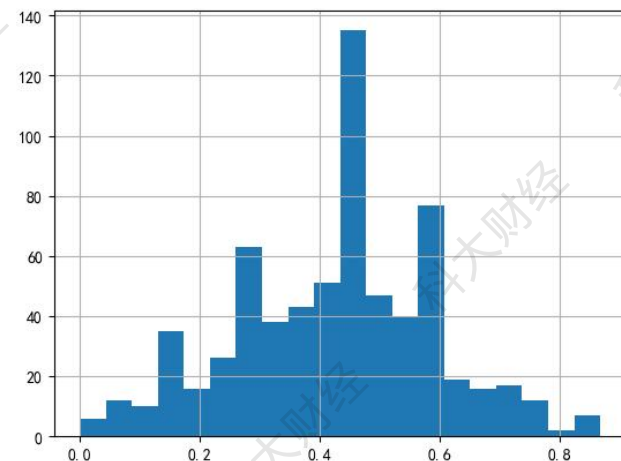
公式中 $r_{t,K}$ 表示 t 时刻 K 行业的日收益率， $r_{t,0}$ 为基准在 t 时刻的收益率，其中行业 K 平均收益率 $\bar{r}_{t,K}$ 以及市场收益率 $r_{t,m}$ 可以表示为：

$$r_{t,m} = \frac{\sum_{K=1}^M r_{t,K}}{M}$$
$$\bar{r}_{t,m} = \frac{\sum_{i=0}^{N-1} r_{t-i,m}}{N}$$

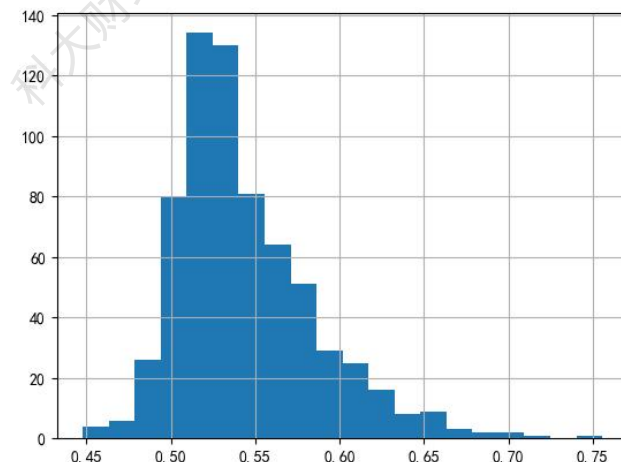
拥挤度指标有效性检验

1. 显著性检验：取拥挤度信号对于未来一个月行业指数收益率进行回归分析，对回归系数进行显著性检验。显著性检验通过意味着拥挤度指标对行业指数收益率有一定的解释力度。
2. 信号胜率检验：汇总出现拥挤度信号之后一个月行业指数收益率，统计其中收益率为负的次数占比。胜率越高则意味着拥挤度指标捕捉下跌风险的准确度越高。
3. 回测收益检验：在所有行业组合中剔除处于拥挤状态的行业，构建处于非拥挤状态的行业组合。如果非拥挤行业组合在回测过程中可以取得超越行业等权基准的收益，说明通过拥挤度指标剔除了收益会出现下降的行业。

显著性



胜率



复合行业拥挤度模型收益表现

细分拥挤指标回测表现

	年化收益率	超额年化收益率	波动率	夏普比率	最大回撤
ind_close_bais_20_p240_95	18.28	9.32	20.52	0.74	19.77
ind_close_bais_40_p360_90	15.88	9.75	20.27	0.64	21.23
ind_close_bais_60_p480_80	14.57	7.35	18.51	0.63	21.14
ind_close_bais_60_p360_90	15.66	10.69	21.08	0.60	19.88
ind_close_vol_20_p360_90	16.45	11.36	22.77	0.59	22.31
ind_close_bais_20_p360_90	13.93	7.18	18.99	0.58	23.24
ind_mtm_20_p360_90	13.49	6.77	19.39	0.54	22.59
ind_close_vol_20_p240_80	14.85	7.25	22.28	0.53	24.32
ind_close_vol_60_p480_95	15.17	7.35	23.51	0.52	31.43
ind_mtm_20_p360_95	13.63	8.16	21.29	0.50	23.16
ind_sharpe_40_p360_90	12.67	7.24	20.81	0.46	26.85
ind_close_bais_40_p360_80	11.66	5.75	18.75	0.46	22.25
ind_volume_40_p360_90	12.93	7.13	21.83	0.45	28.82
ind_close_bais_40_p360_95	12.89	8.13	22.93	0.43	25.61
ind_volume_bais_120_p360_95	11.05	6.16	19.26	0.42	24.81
ind_close_bais_60_p360_95	13.05	9.17	24.27	0.41	30.75
ind_close_bais_120_p480_95	11.90	7.64	22.64	0.39	31.86
ind_distance_40_p360_90	10.99	5.65	20.57	0.39	24.36
ind_volume_40_p360_95	11.53	5.66	22.48	0.38	32.56
ind_volume_20_p240_90	10.72	2.33	20.50	0.38	27.84
ind_mtm_20_p240_95	10.50	2.13	20.97	0.36	23.24
ind_close_bais_40_p240_95	10.46	5.39	21.29	0.35	21.67
ind_close_vol_60_p480_90	11.31	3.75	23.42	0.35	29.38
ind_close_bais_120_p360_90	10.28	5.43	21.67	0.34	28.46
ind_close_bais_60_p240_90	10.12	6.16	21.18	0.34	20.74
ind_volume_bais_120_p360_90	9.18	4.37	18.91	0.33	24.60
ind_close_bais_60_p360_80	9.38	4.68	19.19	0.33	21.62
ind_total_turnover_40_p240_99	9.88	4.84	21.62	0.32	31.80
ind_VaR_60_p480_95	10.24	2.75	22.61	0.32	24.94
ind_turn_ratio_20_p240_80	9.15	0.88	20.06	0.31	23.83

年化收益率	超额年化收益率	波动率	夏普比率	最大回撤
17.87	10.07	20.99	0.71	22.06

