

基于日内高频数据的短周期选股因子研究

高频数据因子研究系列一

报告摘要:

● 传统多因子选股

在国内 A 股市场,传统的多因子量化选股模型得到了广泛的应用,在实际表现中,传统的多因子模型在过去几年中也表现出较为稳定的超额收益率。但随着传统多因子模型应用越来越广泛,历史长期有效的因子逐渐失效,对新因子的挖掘提出了迫切的需求。

● 新因子挖掘

传统的因子指标挖掘主要集中于财务报表、个股中低频率的量价等相关的数据维度,而这部分数据维度的增量价值的挖掘已逐渐饱和,需从其他新的数据维度中挖掘新的因子指标,本篇报告从个股日内高频数据出发尝试挖掘出新的因子指标。

● 基于高频数据因子的策略构建

基于个股日内高频数据,构建了已实现波动 (Realized Volatility) $RVol$, 已实现偏度 (Realized Skewness) $RSkew$ 、已实现峰度 (Realized Kurtosis) $RKurt$ 因子指标,考察这三个因子在回测区间内对个股收益率的区别度。

● 策略实证结果分析

在实证区间内,报告对 $RVol$, $RSkew$ 、 $RKurt$ 三个因子指标进行了详细测算。实证结果表明, $RVol$ 、 $RKurt$ 因子指标对个股收益率区分度不明显,而 $RSkew$ 在全市场以及中证 500 中对个股收益率区分度明显。

因子指标 $RSkew$ 在全市场中选股,从 2007 年至今,IC 均值为-0.028,负 IC 占比为 68.7%,多头组合在回测期内表现优异,年化收益率为 26.7%,多头组合对冲中证 800 指数后年化收益率为 17.7%,最大回撤为 23.6%,信息比率为 1.291。

因子指标 $RSkew$ 在中证 500 指数成分股中选股,从 2007 年至今,IC 均值为-0.04,负 IC 占比为 64.6%,多头组合取得了 23.10%的年化收益率,多头组合对冲中证 500 指数后年化收益率为 11.20%,最大回撤为 5.70%,信息比率为 2.076。

● 核心假设风险:

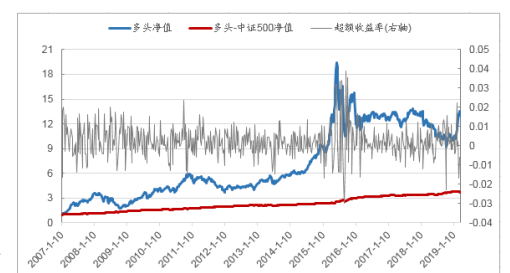
本文所做的数据测算完全基于过去数据的推演,市场未来环境可能发生变化。投资者制定投资策略时,必须结合市场环境和自身投资理念。

图: 因子 $RSkew$ 在全市场选股表现



数据来源: 广发证券发展研究中心

图: 因子 $RSkew$ 在中证 500 中选股表现



数据来源: 广发证券发展研究中心

分析师: 陈原文



SAC 执证号: S0260517080003



0755-82797057



chenyuanwen@gf.com.cn

分析师: 罗军



SAC 执证号: S0260511010004



020-66335128



luojun@gf.com.cn

分析师: 安宁宁



SAC 执证号: S0260512020003



SFC CE No. BNW179



0755-23948352

anningning@gf.com.cn

请注意,陈原文、罗军并非香港证券及期货事务监察委员会的注册持牌人,不可在香港从事受监管活动。

相关研究:

风险收益平衡: 可转债基金深 2019-04-20

度解析:基金产品专题研究系

列之十一

目录索引

一、 引言	5
二、 因子构建	8
三、 实证分析	9
数据说明	9
策略构建	9
因子特征	9
实证分析—全市场、中证 500 因子选股分档表现	12
实证分析—全市场选股	13
实证分析—中证 500 选股	16
四、 总结	19
五、 风险提示	20

图表索引

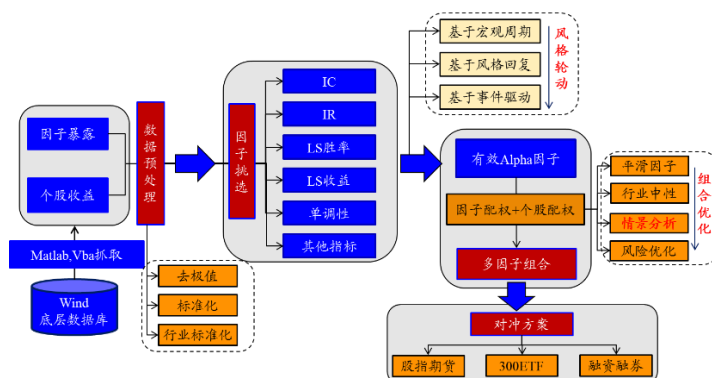
图 1: 广发金融工程多因子选股框架一览	5
图 2: 广发金融工程多因子选股平台一览	5
图 3: 广发金融工程多因子选股平台框架一览	6
图 4: 全市场三个月股价反转因子历史多空收益率表现一览	6
图 5: 全市场三个月股价反转因子 IC 表现一览	7
图 6: 全市场流通市值因子历史多空收益率表现一览	7
图 7: 全市场流通市值因子 IC 表现一览	7
图 8: 全市场个股波动率分布一览	10
图 9: 全市场个股波动率百分位走势一览	10
图 10: 全市场个股偏度分布一览	10
图 11: 全市场个股偏度百分位走势一览	10
图 12: 全市场个股峰度分布一览	10
图 13: 全市场个股峰度百分位走势一览	10
图 14: 中证 500 成分股波动率分布一览	11
图 15: 中证 500 成分股波动率百分位走势一览	11
图 16: 中证 500 成分股偏度分布一览	11
图 17: 中证 500 成分股偏度百分位走势一览	11
图 18: 中证 500 成分股峰度百分位走势一览	11
图 19: 中证 500 成分股峰度百分位走势一览	11
图 20: 因子指标 $RVol$ 全市场选股分档表现	12
图 21: 因子指标 $RSkew$ 全市场选股分档表现	12
图 22: 因子指标 $RKurt$ 全市场选股分档表现	12
图 23: 因子指标 $RVol$ 中证 500 选股分档表现	13
图 24: 因子指标 $RSkew$ 中证 500 选股分档表现	13
图 25: 因子指标 $RKurt$ 中证 500 选股分档表现	13
图 26: 全市场 $RSkew$ 因子 IC 值走势一览	14
图 27: 全市场 $RSkew$ 因子选股多-中证 800 策略净值走势表现一览	15
图 28: 中证 500 $RSkew$ 因子 IC 值走势一览	16
图 29: 中证 500 指数成分股 $RSkew$ 因子选股多-空策略净值走势表现一览	17
图 30: 中证 500 指数成分股 $RSkew$ 因子选股多-中证 500 策略净值走势表现一览	18
表 1: 广发金融工程大数据研究报告一览	8
表 2: 全市场选股-IC 表现	13
表 3: 全市场选股-IC 分年度表现一览	14
表 4: 全市场多-中证 800 分年度表现一览	15
表 5: 全市场选股换手率分年度统计一览	16
表 6: 中证 500 指数内选股-IC 表现	16
表 7: 中证 500 选股-IC 分年度表现一览	17

表 8: 中证 500 数成分股 $RSkew$ 因子选股多-空策略分年度表现.....	18
表 9: 中证 500 选股多头-500 策略分年度表现一览.....	18
表 10: 中证 500 选股换手率分年度统计一览.....	19

一、引言

传统的多因子选股策略在国内市场上广泛应用，在过去几年中传统的多因子选股策略在实际运作中取得了较为稳定的超额收益率。在国内市场中，传统的多因子选股框架中，从2007年开始，较为有效的因子主要是反转类以及小市值类的因子。在传统的多因子研究框架中，对因子的挖掘主要集中于上市公司财务报表、分析师预期相关数据以及相对频率较低的价格数据（如开盘价、收盘价、成交额等日频、周频相关的数据维度），调仓的频率也往往集中于月度频率等相对低频的调仓频率。

图 1：广发金融工程多因子选股框架一览



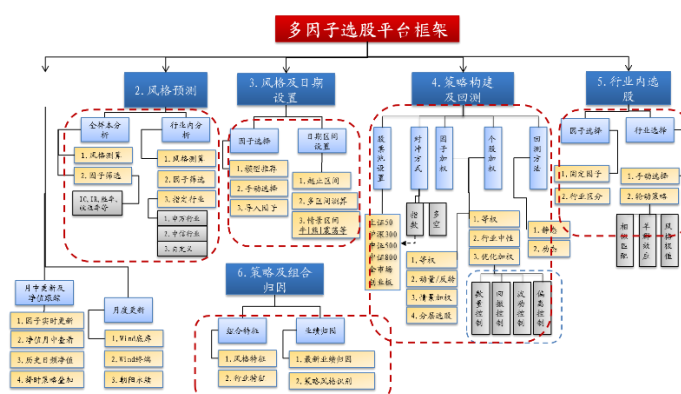
数据来源：广发证券发展研究中心

图 2：广发金融工程多因子选股平台一览



数据来源：广发证券发展研究中心

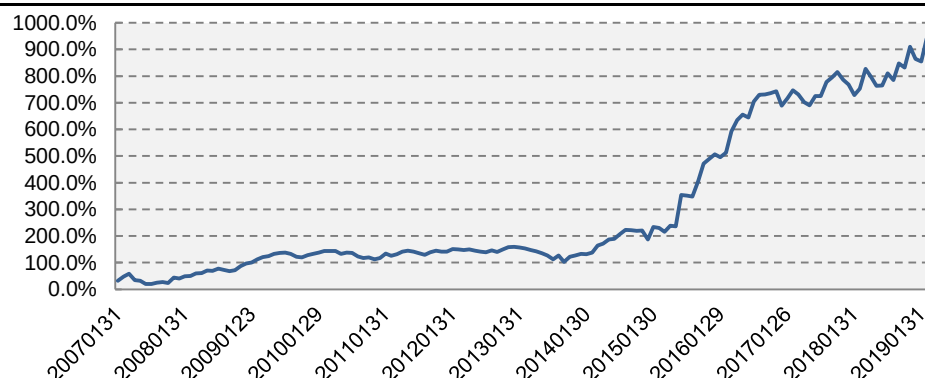
图 3：广发金融工程多因子选股平台框架一览



数据来源：广发证券发展研究中心

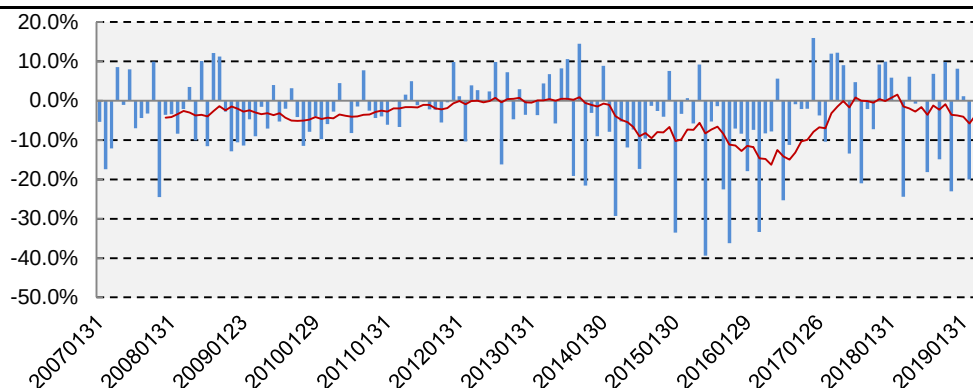
随着国内市场对传统多因子选股的应用越来越广泛，以往有效的因子逐渐失效，而且对中低频率价量相关的数据以及财务报表等数据的因子挖掘已经逐渐饱和，已有的数据维度上增量价值信息有限，很难再在当前维度的数据中挖掘出持续有效的新维度的因子，对新的因子的挖掘提出了迫切的需求。同时，在国内市场中，由于小市值效应的长期较为显著的影响，传统的多因子选股策略往往受其影响，如在2017年，市场的风格较以往几年发生了急剧的变化，风格上主要集中于价值蓝筹类个股，传统的反转类、市值类等因子指标失效，市场上的传统多因子选股策略产品经历了较大的回撤。

图 4：全市场三个月股价反转因子历史多空收益率表现一览



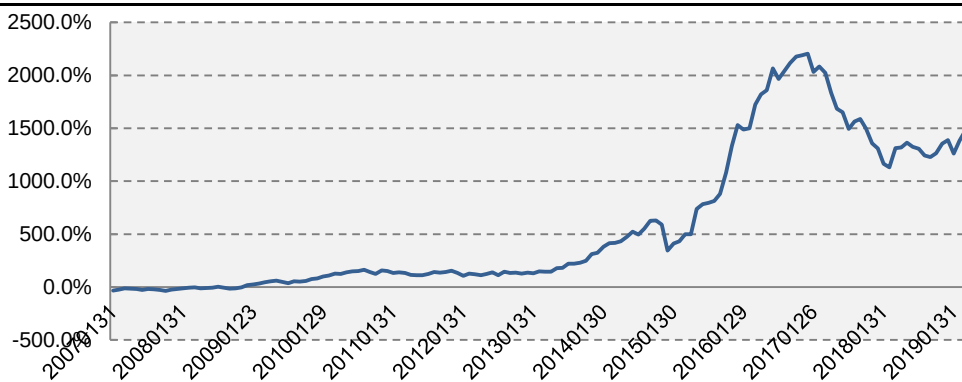
数据来源：广发证券发展研究中心

图 5：全市场三个月股价反转因子IC表现一览



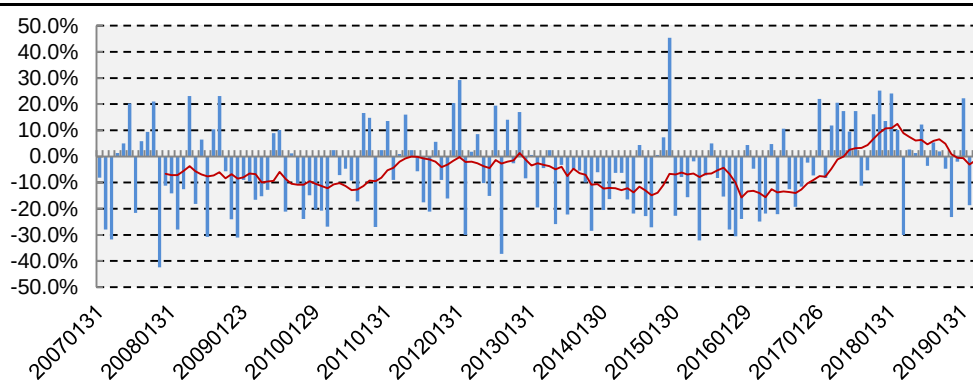
数据来源：广发证券发展研究中心

图 6：全市场流通市值因子历史多空收益率表现一览



数据来源：广发证券发展研究中心

图 7：全市场流通市值因子IC表现一览



数据来源：广发证券发展研究中心

传统的多因子选股框架在因子指标维度上需要进一步丰富和扩展，挖掘出新的有效的因子指标，从当前可切入的数据维度上看，新的因子指标的挖掘主要集中在两大方向，一块为对另类数据的因子指标挖掘，如对股吧、社交媒体、搜索引擎、新闻等另类数据的挖掘，关于另类数据在市场的运用，可以参考广发证券金融工程

相关的系列研究成果；第二块为对高频价量相关的数据的因子指标挖掘，如对个股日内盘口数据、日内分钟、秒钟等级别的价量数据的因子挖掘。本篇专题报告将从第二个方向的角度出发，利用个股日内高频的相关数据进行因子指标的挖掘研究。

表 1：广发金融工程大数据研究报告一览

基于网络新闻热度的择时策略—互联网大数据挖掘系列之一
公告披露背后隐藏的投资机会—互联网大数据挖掘系列之二
倾听股吧之声，洞察大盘趋势—互联网大数据挖掘系列之三
那些年一起追过的财经小编选股策略—互联网财经频道文本挖掘策略
基于互联网挖掘的热点选股策略—互联网大数据挖掘系列之五
基于大数据挖掘的关联个股投资机会—互联网大数据挖掘系列之六
基于大数据挖掘的Smart Beta策略—互联网大数据挖掘系列之七
多维数据下的大数据择时策略研究—互联网大数据挖掘系列之八
基于大数据挖掘的概念轮动策略—互联网大数据挖掘系列之九
基于大数据挖掘的行业轮动策略—互联网大数据挖掘系列之十
基于舆情的资产配置模型—互联网大数据挖掘系列之十一
基于网络舆情的指数轮动策略研究—互联网大数据挖掘系列之十二
基于网络舆情再探指数轮动策略—互联网大数据挖掘系列之十三

数据来源：广发证券发展研究中心

二、因子构建

在个股高频数据中，主要包括开盘价、收盘价、最高价、最低价、成交量、成交额等指标以及分笔的盘口相关的数据。本篇专题报告主要是对个股的分钟级别的成交相关的数据进行因子挖掘，希望能从中挖掘出有效的因子指标。

具体因子指标构建如下：

- 对于每个个股在交易日 t ，首先计算个股在特定分钟频率下第 i 个的收益率 $r_{t,i}$ ， $r_{t,i} = p_{t,i} - p_{t,i-1}$ ，其中 $p_{t,i}$ 表示在交易日 t ，个股在第 i 个特定分钟频率下的对数价格， $p_{t,i-1}$ 表示在交易日 t ，个股在第 $i-1$ 个特定分钟频率下的对数价格。
- 对于每个个股，根据 $r_{t,i}$ 分别计算个股在交易日 t 下的已实现方差（Realized Variance） $RDVar_t$ 、已实现偏度（Realized Skewness） $RDSkew_t$ ，已实现峰度（Realized kurtosis） $RDKurt_t$ 。其中：

$$RDVar_t = \sum_{i=1}^N r_{t,i}^2$$

$$RDSkew_t = \frac{\sqrt{N} \sum_{i=1}^N r_{t,i}^3}{RDVar_t^{3/2}}$$

$$RDKurt_t = \frac{N \sum_{i=1}^N r_{t,i}^4}{RDVar_t^2}$$

N 表示个股在交易日 t 中特定频率的分钟级别数据个数，如在5分钟级别下，交

易日t下共有的数据个数N为48(60*4/5=48)。

- 对于每个个股,在交易日t计算累计已实现波动(Realized Volatility) $RVol_t$, 已实现偏度(Realized Skewness) $RSkew_t$ 、已实现峰度(Realized Kurtosis) $RKurt_t$, 其中:

$$RVol_t = \left(\frac{242}{n} \sum_{i=0}^n RDVar_{t-i} \right)^{1/2}$$

$$RSkew_t = \frac{1}{n} \sum_{i=0}^n RDSkew_{t-i}$$

$$RKurt_t = \frac{1}{n} \sum_{i=0}^n RDKurt_{t-i}$$

- 在每期调仓日截面上,按照上述公式计算每个个股的已实现波动(Realized Volatility) $RVol_t$, 已实现偏度(Realized Skewness) $RSkew_t$ 、已实现峰度(Realized Kurtosis) $RKurt_t$ 指标,针对每个由高频数据计算得到的因子指标在历史上的分档组合表现,试图寻找出相对有效的因子指标。

三、实证分析

数据说明

- 样本区间: 2007年1月1日至2019年3月27日(以下如无特别说明,2019年至今指的是2019年1月1日至2019年3月27日)
- 样本范围: 全市场个股、中证500历史成分股
- 数据频率: 个股每个交易日5分钟频率的收盘价、成交量、成交额等数据

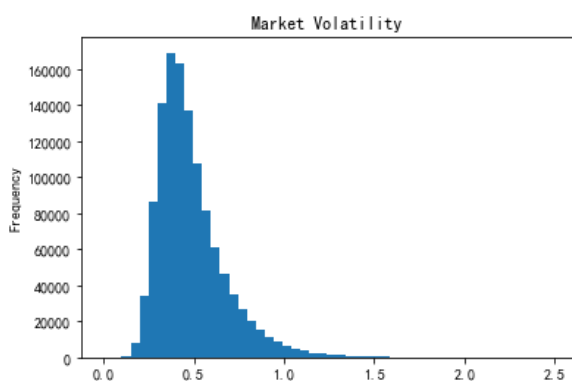
策略构建

- 实证区间: 2007年1月1日至2019年3月27日
- 选股范围: 全市场、中证500历史成分股,剔除上市不满一年的股票,剔除ST股票、*ST股票,剔除交易日停牌的股票
- 分档方式: 根据当期个股计算的因子值: 已实现波动(Realized Volatility) $RVol_t$, 已实现偏度(Realized Skewness) $RSkew_t$ 、已实现峰度(Realized Kurtosis) $RKurt_t$, 从小到大分为5档
- 调仓周期: 周频换仓, Q1档为因子值最小的, Q5档为因子值最大的。
- 参数说明: $N=48$, $n=5$

因子特征

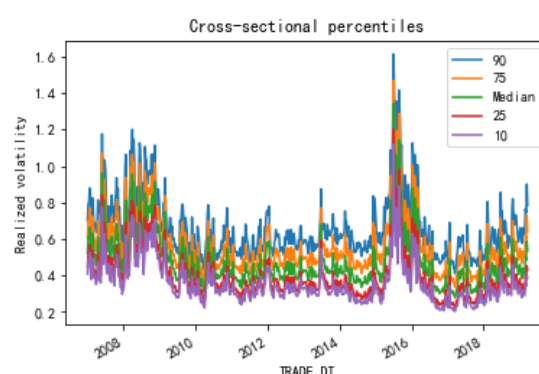
首先,分别统计已实现波动(Realized Volatility) $RVol$, 已实现偏度(Realized Skewness) $RSkew$ 、已实现峰度(Realized Kurtosis) $RKurt$ 在历史上的特征。

图 8：全市场个股波动率分布一览



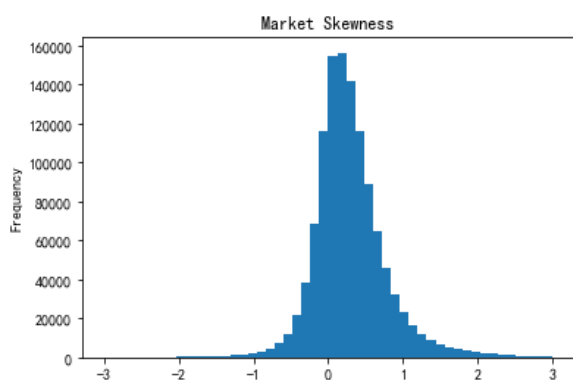
数据来源：天软，Wind，广发证券发展研究中心

图 9：全市场个股波动率百分位走势一览



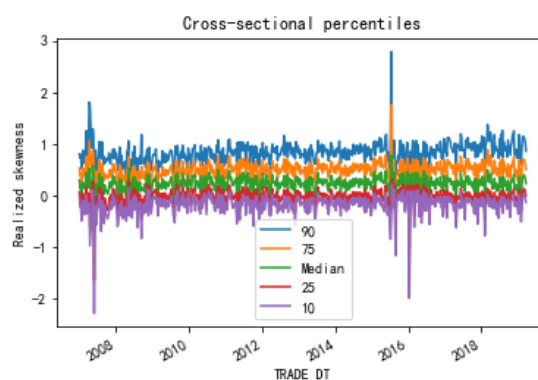
数据来源：天软，Wind，广发证券发展研究中心

图 10：全市场个股偏度分布一览



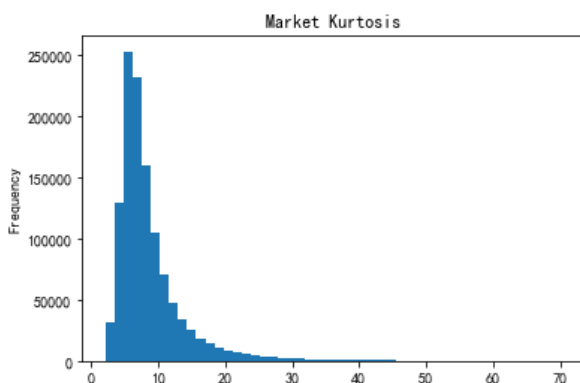
数据来源：天软，Wind，广发证券发展研究中心

图 11：全市场个股偏度百分位走势一览



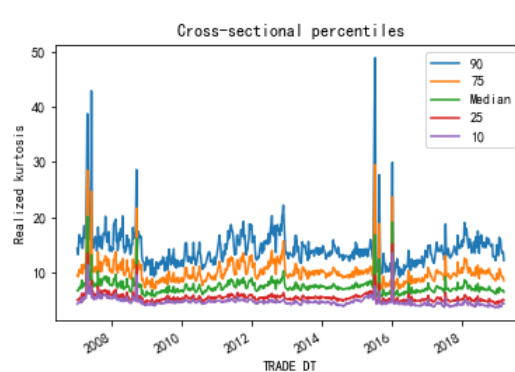
数据来源：天软，Wind，广发证券发展研究中心

图 12：全市场个股峰度分布一览



数据来源：天软，Wind，广发证券发展研究中心

图 13：全市场个股峰度百分位走势一览

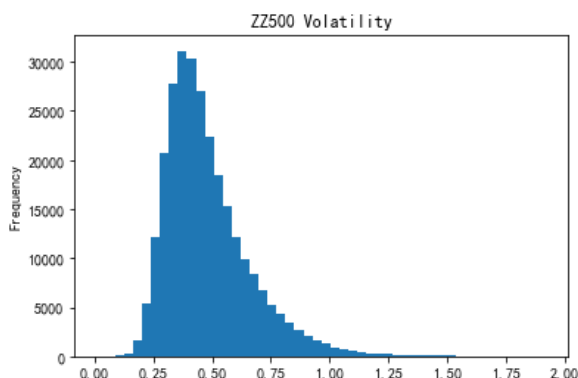


数据来源：天软，Wind，广发证券发展研究中心

个股在分钟级别的数据计算得到的关于个股的波动率、偏度、峰度的结果中可以看出(上图8至图13)，个股的波动率在不同的时间维度上变化较大，从波动率分布上可以看出，整个A股市场个股的波动率分布整体上呈现右偏分布，从时间维度上看，当市场趋势行情较明显时候，个股波动率水平整体上呈现上升的趋势；个股的偏度分布上，整体偏度水平保持在零附近，呈现较为明显厚尾状态，从个股偏度不同百分位时间序列走势上可以看出，个股偏度水平整体较为稳定；从个股的峰度分布

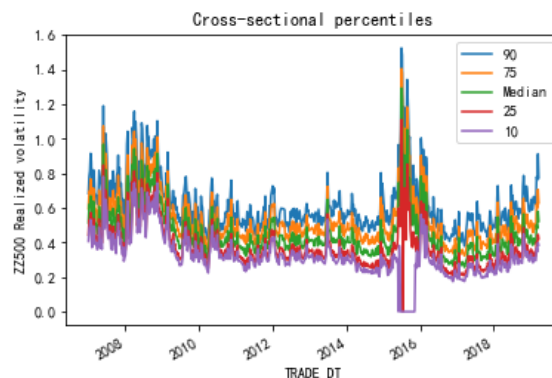
上看，与个股波动率水平类似，分布整体上右偏，且样本内个股的峰度水平大部分大于3，呈现厚尾的现象。

图 14: 中证500成分股波动率分布一览



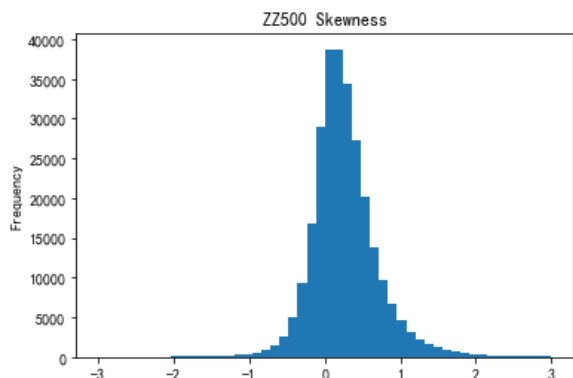
数据来源：天软，Wind，广发证券发展研究中心

图 15: 中证500成分股波动率百分位走势一览



数据来源：天软，Wind，广发证券发展研究中心

图 16: 中证500成分股偏度分布一览



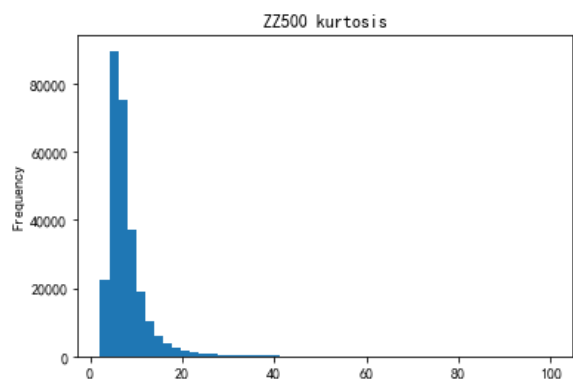
数据来源：天软，Wind，广发证券发展研究中心

图 17: 中证500成分股偏度百分位走势一览



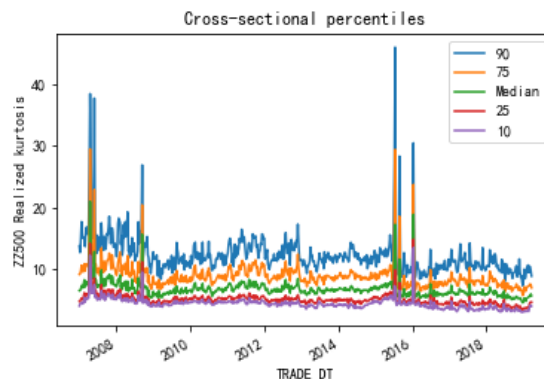
数据来源：天软，Wind，广发证券发展研究中心

图 18: 中证500成分股峰度百分位走势一览



数据来源：天软，Wind，广发证券发展研究中心

图 19: 中证500成分股峰度百分位走势一览



数据来源：天软，Wind，广发证券发展研究中心

利用中证500指数成分股在分钟级别的数据计算得到的关于个股的波动率、偏度、峰度的结果，从中可以看出(上图14至图19)，个股的波动率在不同的时间维度上变化较大，从波动率分布上可以看出，中证500指数成分股波动率分布呈现右偏分布，

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/216105124241010233>