

中国沪深 300 指数尾部风险与收益率预测*

——基于个股高频横截面数据的研究

陈思，任宇

内容摘要：本文借鉴 Kelly & Jiang (2014) 的方法，构建了中国沪深 300 指数尾部风险的代理变量。我们选取 1995 年 1 月至 2022 年 12 月期间个股高频股票指数横截面数据，根据个股发生尾部风险的经验分布估计市场周度尾部风险指数，并研究该尾部风险对资产收益率的预测能力。我们使用了中国市场沪深 300 等指数投资组合，并构造了根据公司行业、规模划分的各种成分投资组合，实证结果显示，尾部风险与投资组合预期回报之间存在显著的负相关关系。我们选取了股票收益率的已实现波动率，已实现偏度等高阶矩指标为预测因子进行对比，使用滚动窗口回归方法以评估尾部风险的样本外预测能力。样本外结果显示使用尾部风险模型进行样本外预测可以得到比基于历史平均值的简单预测更好的结果。资产配置方面，研究使用尾部风险因子作为筛选变量能够提高投资组合的表现。该策略通过每周尾部指标的预测值筛选掉尾部风险较大的行业投资组合，进而实现投资组合中集中配置低尾部风险行业的目标。采用尾部风险策略的投资组合的平均回报率和夏普比率比 1/N 策略高，同时风险也更小。

关键词：尾部风险 样本内预测 样本外预测 资产配置

一、引言

风险和收益的关系是资产管理领域的重要主题。党的十八大以来，习近平总书记要求坚决打好防范化解重大风险攻坚战，“防范化解金融风险，特别是防止发生系统性金融风险，是金融工作的根本性任务，也是金融工作的永恒主题。”尾部风险即极端事件风险，是重要的系统性风险之一。其发生概率虽然很低，然而一旦发生却损害巨大、甚至会造成灾难性后果。资本市场出现金融危机等极端事件时，收益率往往呈现暴涨暴跌的趋势，并非遵循收益率正态分布的传统假设。市场下跌风险比上涨风险更加严重，出现极端事件的概率比正态分布高得多。因此，度量尾部风险需要考虑收益率分布的厚尾性。传统的“均值-方差模型”、“波动率”等风险度量方法不能完全反映尾部风险，资产定价模型应当考虑尾部风险（极端损失）。

在经典的资产定价领域中，左尾风险预测收益理想的框架应该是：高风险高收益。目前已有相关研究表明，系统性的左尾部风险能够正向预测股票的横截面收益，即尾部风险产生风险溢价。然而，实际情况与上述研究结论一致吗？

实际上，国内外学者已有研究证明股票市场上存在“尾部风险异象”。Bali et al.

(2014) 基于下偏距 (LPM) 的思想分别构建了系统性、特异性和混合性的左尾部风险指标，实证研究发现以下偏距度量的特异性左尾部风险会负向影响股票的未来收益，这与正

* 陈思，经济学硕士，中南财经政法大学文澜学院，邮政编码：430073，电子信箱：chensi199923@163.com。

任宇（通讯作者），经济学博士，金融学教授，中南财经政法大学文澜学院，邮政编码：430073，电子信箱：yren@vip.163.com。

风险回报权衡理论相反，作者发现在美国市场上以 VaR 和 ES 度量的左尾部风险对股票横截面收益具有显著的负向影响，并且在加入其他定价因子进行回归之后这种负相关关系依旧存在，即发现了“左尾风险异象”；同样发现在美国股票市场中，个股的 VaR 和 ES 指标与股票截面收益之间存在负相关关系。

类似于美国股票市场的尾部风险研究，中国学者也开始关注“尾部风险异象”的存在和影响。在中国股票市场中，左尾风险的增加可能对个股的未来表现产生负面影响。然而，由于中国股市的特殊性，包括市场结构、投资者行为和监管制度等方面的差异，对于“尾部风险异象”在中国股市中的具体影响和资产定价机制的研究仍然具有挑战性。进一步的研究将有助于深入了解中国股票市场中的风险特征，并为投资者和决策者提供更准确的风险评估和资产定价工具。

二、国内外研究现状

左尾风险指的是由低概率事件带来的重大损失风险，通过资产回报的概率密度函数在均值左侧远离的部分表示。左尾风险的概念在金融市场中非常重要，有助于投资者评估投资的潜在下行风险。Glasserman & Young (2016) 将金融业尾部风险视为金融系统性风险 (Systematic Risk)，基于所有金融机构的股票日收益率数据，使用 VaR 方法进行度量并探讨了该风险对实体经济的预测能力。在金融建模中，通常使用正态分布对资产收益进行假设，但实际证明其无法捕捉金融市场回报的真实分布。Mandelbrot (1963) 和 Fama (1963) 最早提出无条件股票收益分布具有厚尾性。这个重要突破放弃了事实不符的正态分布冲击的假设，将严重左偏的非正态分布冲击作为尾部风险建模的出发点。金融市场回报的厚尾特性意味着极端事件比在正态分布下预期发生的更频繁。

由于极端事件的罕见性，很难像 GDP 增长率时间序列预测中构建尾部风险动态指标通过 Garch 模型估计动态波动，所以传统动态尾部风险估计方法应用较为困难。建模左尾风险具有挑战性，同时金融市场中尾部风险的时变性也需要考虑。关于尾部风险，尤其是很可能导致较大损失的左尾部风险，学术界对此话题兴趣颇深且做了较多的研究。其研究重点大致可以分为两类：第一类主要聚焦于研究左尾部风险的度量问题；第二类研究的重心为该风险与资产横截面收益之间的关系，从而研究其对资产收益率的定价作用。

(一) 尾部风险度量相关研究

国内外文献基于不同的理论构建了多种股票市场尾部风险的代理指标，具体概括为以下四类：

第一类是基于股票收益分布的高阶矩特征。考虑到股票收益率分布的高阶矩（偏度和峰度）特征可以抓取其非对称及“厚尾”信息，部分学者直接使用偏度或峰度衡量股票的尾部风险。负偏程度越深、峰度越高的股票，尾部风险越大。Amaya et al. (2015) 使用日内数据计算周度频率的股票已实现偏度和峰度，以度量其尾部风险。他们发现股票已实现偏度与预期收益的负向关系在经济和统计意义上均非常显著，但已实现峰度同预期收益的正向关系则并不稳健。另外，Habib et al. (2018) 提出负收益偏态系数也常被用来度量股价被高估后的崩盘风险。

第二类是体现系统性尾部风险敏感度的尾部风险贝塔。从股票收益率的左侧分布出发，通过计算收益率与市场尾部收益率的条件协方差来构建股票尾场（系统性）尾部风险的敏

感度。Kelly & Jiang (2014) 认为当受到同样的冲击时，个股的尾部分布也会表现出同向的变化，这种个股的同向运动反映了市场整体的尾部风险。基于此，他们首先混合个股的收益率数据，构造了市场的时变尾部风险，然后在截面上通过滚动回归得到个股的尾部风险贝塔，以表征个股对市场尾部风险的敏感度，结果显示尾部风险贝塔高的股票也有更高预期收益。

第三类是直接测度损失可能性及大小的在险价值和预期损失。在险价值 (Value at risk, VaR) 是指预测在给定的时间期限和某种置信水平下，金融资产或投资组合可能损失的最大损失金额，因而是一种具有边界性质的风险测度，衡量尾部风险的经济意义也较为明确。VaR 测量方法实际上抛弃了收益分布的左尾部分，不能直接测量尾部风险。因此，在使用 VaR 测量方法来测量尾部风险时，通常需要限制或分割分布函数的形状或数学特性。Bali et al. (2009) 研究了使用参数法和非参数法估计在险价值的典型过程，为以在险价值和预期损失度量的股票尾部风险效应研究奠定了重要基础和方法示例。期望损失 (ES) 是指当投资组合的损失超过 VaR 门槛时产生的平均损失。由于 ES 在 VaR 基础上考虑了极端情况下的平均损失水平，因此可以更全面地衡量投资组合的极端损失风险，作为更全面地测量尾部风险的 VaR 补充。

第四类是基于极值理论的尾部指数。极值理论 (Extreme value theory, EVT) 专以随机过程的极值分布为研究对象，极值理论只侧重分析股票收益分布的尾端情况。不同于收益高阶矩特征需要关注股票收益的整体分布，该方法不需要预设分布具体类型，也避免了模型预设错误。Huang et al. (2012) 在美股市场股票收益率数据计算残差收益后，使用极值理论构造了个股的极端下行风险指标，该指标同时包含了股票特质层面的极端下行和肥尾风险信息。在此基础上，他们探究了极端下行风险的公司特征，发现极端下行风险高的公司通常杠杆率较高、盈利能力较弱，面临着较大的财务困境风险，同时公司的股票价格也会较低且波动性较大。因此，以极端下行风险代理的股票特质性尾部风险也较好地反映了公司的基本面和市场风险。Kelly & Jiang (2014) 提出了能够度量时变尾部风险的方法，假设个股收益率序列的条件尾部分布服从动态幂指数过程，采用 Hill (1975) 提出的估计量来推断尾部事件的幂律概率 (power law probabilities)。使用个股横截面数据，应用面板估计方法，有效规避收益率数据有限的缺陷，而且不需要基于特定的极端经济金融事件去衡量尾部风险。陈国进等 (2015) 从中国 A 股市场个股横截面数据提取尾部风险，以此作为隐含的灾难风险的代理指标。但与 Kelly 等选用固定阈值 (5%) 不同，由于灾难风险的度量对阈值的大小具有很强的敏感性，他们通过 Patie 提出的峰度法动态地确定阈值。

关于尾部风险，中国的研究起步相对较晚，因此大多数国内研究采用了国外学者提出的研究思路和方法，创新性方法鲜有涌现。此外，中国与西方国家国情存在较大差异，因此国外学者在研究中国金融市场的尾部风险问题时，其研究方法并不完全适用。因此，有必要探索适合中国市场特点的尾部风险研究方法，以确保研究的有效性。

(二) 尾部风险与收益率相关理论

关于左尾风险股票定价机制，最早由 Rietz (1988) 开始研究市场崩溃风险对于股票市场的影响，发现发生概率很低的左尾风险可以用于解释股权溢价；Robert J (2006) 进一步研究了系统性左尾部风险对于股票溢价的定价作用。Bali, Demirtas & Levy (2009) 采用了风险价值 (VaR)、预期差额和尾部风险衡量系统性下行风险，发现下行风险与纽约证券交易所、美国证券交易所、纳斯达克股票的投资组合收益之间存在显著的正相关关系。与传统

的风险度量相比，VaR 仍然是一个更好的风险度量。

Gabaix (2012) 研究将灾难风险随时间变化纳入 Rietz (1988) 的假设，即风险溢价是由罕见大灾害发生的可能性产生的。在灾难期间，资产的基本价值会随时间变化而下降。这反过来又产生了时变的风险溢价，从而产生了波动的资产价格和回报可预测性。并且利用灾难风险定量解释了股权溢价之谜资本市场的异象。Allen (2012) 使用三种不同方法构造的 VarR 和 ES 的均值作为衡量系统性尾部风险的指标，并发现该指标和银行借贷活动负相关，并且尾部风险在股票市场中具有一定的定价作用。陈国进等 (2015) 从个股横截面数据提取尾部风险作为时变罕见灾难风险的代理指标，实证分析罕见灾难风险作为定价因子对我国股市收益的预测能力和横截面收益的解释能力。

股票收益可预测性，是金融研究的核心问题之一，众多经济变量被发现可以作为股票收益的预测变量。Campbell (2000) 指出学术界对股票收益有显著的样本内可预测性已基本达成共识。然而，股票收益的样本外可预测性至今还有争议。比如 Welch & Goyal

(2008) 认为历史平均超额股票收益比超额收益对预测变量的回归更能预测未来超额股票收益。但 Campbell & Thompson (2008a) 指出通过附加基于金融理论参数约束，许多预测回归就会超过历史平均收益。尽管样本外解释能力很小，但对均值方差投资者来说仍具有经济意义。Ludvigson & Ng (2009) 等通过渐近主成分法提取多个动态因子，宏观经济因素对债券超额收益具有统计显著的预测能力。并展示了如何对因子增广回归的参数估计进行偏差校正。此外，在中国市场中，姜富伟等 (2011) 研究了投资组合和根据公司行业规模、面值市值比和股权集中度等划分的各种成分投资组合的股票收益的可预测性。选取中国股市估值、货币发行、交易量等 9 个经济变量作为预测变量，中国市场投资组合和各种成分投资组合都存在显著的样本内和样本外可预测性。不同成分投资组合的可预测性存在显著差异，其中金融与保险业、房地产业和制造业等行业投资组合的可预测性特别强，小市值、低面值市值比和低股权集中度的投资组合也非常容易预测。

近年来国外学者越来越关注左尾风险与股票收益之间的关系，中国股票市场散户众多，投资者更容易产生非理性行为，在这种情况下，创新地研究我国 A 股市场个股左尾风险与股票收益之间的关系显得更有意义。值得注意的是，目前股票市场风险价值的研究大多基于低频数据进行风险价值的分析，但由于低频数据取样较少，例如，由于中国股市比较年轻，所以月度尾部风险样本数据较少，容易忽略市场信息，从而低估风险。如市场在相邻两天的收盘价大致相同，而这两天却发生了较大的市场动荡，从而引发严重的市场风险，此时低频模型是无能为力的。而少数考虑高频信息的相关文献大多没有考虑股票市场的厚尾性，且基本集中于成熟股票市场。为了进一步开展中国国情的尾部风险的研究，本文基于“理论基础—测度模型—样本内预测—样本外预测—资产配置”的尾部风险研究框架，使用中国 A 股市场收益率高频数据，研究尾部风险对收益率的预测能力。

三、指标以及投资组合构建方法

(一) 样本选择与数据来源

本文的研究对象为沪深 300 指数尾部风险与收益率，并且具体地针对 2005 年至 2022 年中国 A 股上市股票行业、市值分类投资组合研究对象进行分析。主要使用股票高频日内收益率数据（以 30、15、10、5 分钟为数据频率的报价），数据来源于聚宽量化平台数据库。

为了能够准确获取横截面的尾部风险信息，我们以周度为频率估计尾部风险指数，从 2005 年 1 月到 2022 年 12 月共 906 期（周度），数据集包含了足够大的横截面，提高了极端值估计的准确性。

为保证结果准确性，构造全 A 股投资组合时，我们对样本观测值做了数据清洗处理，依据 Liu et al. (2019) 的做法，从 2005 年 1 月起，在每个月末进行一次如下过滤：

1. 剔除停牌及退市 ST 股票；
2. 以当月末为基准，回溯 180 个自然日，删除在此之后上市的股票（上市不到半年的新股）；
3. 根据上月末市值（根据删除壳价值污染的解释和控股理论，此处应当按总市值删），将处于公司规模最小的 30% 股票进行剔除；

上述程序按月滚动执行，形成股票池。进一步，我们将数据池根据周度分组，得到 906 期周度股票池，即为计算市场收益率和因子收益率的周度股票池。本文展示的预测表现包括不同数据频率下的回归结果，用于评估沪深 300 指数尾部风险的预测能力。数据频率指回归模型使用的数据的时间频率，包括 1 小时、30 分钟、15 分钟、10 分钟和 5 分钟。

（二）投资组合构建方法

参考的方法，并加以改进，形成每月行业投资组合。在每年 6 月底，以申银万国提供的分类为依据，在 20 个行业中形成价值加权投资组合。每月构建行业投资组合方式如下：

1. 将中国全 A 股股票按照行业分类为 31 个数据集。对于每个行业的数据集，使用上个月月末市值计算股票权重，计算该行业投资组合的市值加权收盘价。
2. 计算日度各个行业的投资组合市值加权收益率。
3. 使用周度内行业收益率的加总数值，作为行业的周度收益率。
4. 收益率中位数指标计算方式：使用日收益率，取一周内交易日的沪深 300 指数日收益率的中位数作为周收益率。

每个行业按照市值加权，按照上个月最后一个交易日计算市值，每个月调整一次权重。构建投资组合，并计算出该行业投资组合回报率。依据之前构造沪深 300 指数尾部指标方法，构建每个行业的尾部指标，探讨该指标是否对行业投资组合回报率的各阶矩的预测能力以及样本外预测能力。此外，构建市值投资组合时，每月按照上月末总市值大小排序分组，构造 30 个含相同数量上市公司的市值加权的规模投资组合，并按市值数值大小升序依次命名为 Size1, ..., Size30。

（三）尾部风险因子估计

1. 收益条件尾部分布

估计尾部风险因子，难点在于其具有时变性。如果公司层面的动态尾部风险具有类似的分布，那么单个公司的崩溃事件截面可以用来识别其尾部风险在每个时间点的共同分量。因此，参考 Kelly & Jiang (2014) 设计的尾部风险面板估计方法，基于中国股票市场第 t 周内的横截面收益尾部样本信息，利用大样本性质，能够得出 t 期的尾部风险估计。

本文的实证框架集中在收益尾部分布研究。时间 t 左尾分布被定义为回报率低于某些极端负阈值的事件集合。我们假设资产收益 i 的左尾分布为：

$$P(R_{i,t+1} < r \mid R_{i,t+1} < u_t \text{ and } \mathcal{F}_t) = \left(\frac{r}{u_t}\right)^{-\frac{\alpha_i}{\lambda_t}} \quad (1)$$

其中， $r < u_t < 0$ 。式 1.1 表明，极端回报事件服从幂律。模型的关键参数 $-\frac{\alpha_i}{\lambda_t}$ 决定

了尾部的形状.被称为尾部指数。因为 $r < u_t < 0$ ， $\frac{r}{u_t} > 1$ 。因此 $-\frac{\alpha_i}{\lambda_t} > 0$ 保证概率 $\left(\frac{r}{u_t}\right)^{-\frac{\alpha_i}{\lambda_t}}$ 总是介于0和1之间。 λ_t 的高值对应的是“肥”尾和极端收益的高概率。

与以往的幂律研究不同，式（1）条件收益尾部分布模型。指数中的 $1/\lambda_t$ 项可能会随着条件信息集 \mathcal{F}_t 的变化而变化。虽然不同的资产可能有不同程度的尾部风险（由恒定的常数 α_i 决定），但所有资产的动态都是相同的，因为它们都是由共同的过程 λ_t 驱动的。因此，我们将 λ_t 称为 t 时刻的“尾部风险”，将（1）中的尾部结构称为“动态幂律”。

本文从动态幂律结构（1）中构建了尾部风险测度的估计器（Tail Risk Measure）。因此，在一个足够大的横截面中，每个时期将有足够多的股票经历个别尾部事件，以提供关于当前尾部风险水平的准确信息。将Hill（1975）的幂律估计器应用到时间 t 截面上，可以得到 λ_t 的估计。

2. 提取尾部风险因子

Kelly & Jiang（2014）的尾部风险估计方法依赖于个别资产尾部风险的共性，反过来利用回报横截面中相对丰富的尾部风险信息。我们允许不同资产之间存在不同水平的公司特定尾部风险，但假设所有资产的尾部风险波动都由一个单一过程控制。这种结构意味着公司有不同的无条件尾部风险，但它们的尾部风险动态是相似的。

在超过某些极低的尾部阈值 u_t 和给定信息 \mathcal{F}_t 的条件下，我们假设资产的回报服从尾部概率分布(1.1)。通过将Hill（1975）幂律估计器应用于 t 期所有股票回报观测集，我们估计了回报尾的时变分量 λ_t ，逐期每个月应用到集合截面上，它的形式基于收益次序统计量的左偏尾部样本，构造Hill统计量如下：

$$\lambda_t^{Hill} = \frac{1}{K_t} \sum_{k=1}^{K_t} \ln \frac{R_{k,t}}{u_t} \quad (2)$$

其中， u_t 为第 t 期周度数据池所选取的阈值，是收益率序列从传统正态分布区别于尾部幂率分布的门槛点。参考Kelly & Jiang（2014）的做法，采用5%的固定阈值水平。 $R_{k,t}$ 是第 t 周期间低于阈值 u_t 的第 k 只股票的高频收益率， K_t 为构造尾部风险因子的尾部样本数。基于（2）式，即可得到 t 期尾部风险 λ_t^{Hill} 。

具体地，我们以周度为频率，将每周内所有A股股票的高频收益率序列汇总成周度数据池。在一个足够大的全A股横截面中，每个时期将有足够多的股票经历个别尾部事件，以提供关于当前尾部风险水平的准确信息。对于第 t 月样本数据池，当样本数据量足够大时，尾部风险度量值期望损失，即为收益率尾部样本均值。在周度横截面数据池内，对高频收益率从低到高进行排序，得到第 t 期次序高频收益率统计量。将Hill（1975）的幂律估计器应用到时间 t 截面上，能够得到 λ_t 的估计。

（四）其余指标构建方法

1. 周度收益率及未来持有期收益率

针对不同的数据频率（60分钟、30分钟和15分钟），我们计算了相应频率下的收益率。以每60分钟频率数据为例，我们使用每个交易日上午9:30到11:30，下午13:00到15:00的连续竞价时段的60分钟频率的报价计算日内对数收益率：

$$r_{t,i} = p_{t,i} - p_{t,(i-1)} \quad (3)$$

其中，其中， $r_{t,i}$ 表示第 t 日 i 时刻的对数收益率， $p_{t,i}$ 表示第 t 日 i 时刻的对数价格。 n 为日内高频收益率的个数。

我们定义预期1周收益率为 $t+1$ 期的一周收益率。预期2周收益率则是 $t+1$ 至 $t+2$ 周收益率的累加。以此类推，我们可以计算未来 n 期的预期收益率。

2. 已实现波动率

参考 Andersen et al. (2003) 的研究，计算已实现波动率的公式如下：

$$RV_t^d = \sum_{i=1}^M r_{t,i}^2 \quad (4)$$

其中 r 表示 t 期的第 i 个收益率， p 表示 t 期的第 i 个收盘价， M 表示采样频率。计算周波动率时，直接对一周内的每 M 分钟收益率进行公式应用。

VR_30min: (Volatility Rate)表示从每30分钟数据中提取的波动率。

VS_30min: 表示从每30分钟数据中提取的波动标准差。波动标准差为波动率的根号二分之一。其余变量名称以此类推。

3. 已实现偏度

高频收益率数据最典型的特征之一是其分布往往具有尖峰厚尾的特征。这些特征可以通过变量的高阶矩特征来反映，其中收益率的三阶矩是收益率偏度。Bali et al. (2016) 提出由于投资者追求具有正偏度的股票，导致其价格容易被高估，预期收益率相应地较低。参照 Amaya et al. (2015)，我们利用高频数据构造沪深300指数等投资组合的收益率已实现偏度。计算日度已实现偏度：

$$RSkew_t = \frac{\sqrt{n} \sum_i r_{t,i}^3}{RDVar_t^{\frac{3}{2}}} \quad (5)$$

本文使用周度数据的预测，因此我们将日度数据转换为周度平均数据：

$$RSkew_j = \frac{1}{N_j} \sum_t RDSkew_t \quad (6)$$

其中， i 表示周数， N 表示第 i 周的天数。

三、实证模型

(一) 样本内预测检验方法

参照姜富伟等 (2011) 与陈坚和张轶凡 (2018) 相关文献，本文通过单变量预测回归模型来检验尾部风险对股票市场收益率的预测能力：

$$R_{t+1} = \alpha + \beta R_{Tail,t} + \epsilon_{t+1} \quad (7)$$

其中， R_{t+1} 为 $t+1$ 时刻的股票市场超额收益率， $R_{Tail,t}$ 是 t 时刻的尾部风险， ϵ_{t+1} 是残差项。判断尾部风险指标对股票超额收益率的预测能力，主要是通过最小二乘法 (OLS) 估计模型(5)中的 β ，得到估计值 $\hat{\beta}$ 以及对应的 t 统计量。检验的原假设是 $\beta = 0$ ，即尾部风险没有预测能力；备择假设是 $\beta \neq 0$ ，即尾部风险可以预测未来股票超额收益率。

为了对比尾部风险与经济变量的预测能力，姜富伟等 (2011) 提出了经济变量可以预

测未来股票市场超额收益率。本文使用收益的各阶距作为尾部风险样本内预测能力的对比。

$$R_{t+1} = \alpha + \beta R_{Tail,t} + \sum_{i=1}^n \theta_i X_{i,t} + \epsilon_{t+1} \quad (8)$$

其中， $X_{i,t}$ 代表了第 i 个预测因子。如果模型中的 β 依然统计显著，就说明尾部风险可以提供额外的预测信息，从而提高了模型的预测能力。那么，在预测中国股票市场收益率时，应该在预测模型中加入尾部风险，从而提高模型的预测能力。

(二) 样本外预测检验方法

根据 Campbell & Thompson (2008) 以及 Welch & Goyal (2008)，我们对尾部风险等预测变量进行样本外检验方法。首先需要把总样本观测值分为由 n_1 个观测值组成的样本内初始估计期和 n_2 个观测值组成的样本外检测期两部分。在样本内初始估计期内，根据预测模型(5) 获得样本外第一个股票收益率预测值：

$$\hat{R}_{n_1+1} = \hat{\alpha}_{n_1} + \hat{\beta}_{n_1} Tail_{n_1} \quad (9)$$

其中 $\hat{\alpha}_{n_1}$ 和 $\hat{\beta}_{n_1}$ 是根据模型(5) 把 $\{Tail_t\}_{t=1}^{n_1-1}$ 对常数项和 $\{R_t\}_{t=1}^{n_1-1}$ 做线性回归得到的参数估计值。

我们将样本外预测使用滚动窗口，窗口大小为 52 周。具体地，对于每个样本外检测期，我们将其对应的样本内估计期的时间段向前滚动 52 周 ($n_1=52$)，并以此作为新的样本内估计期。例如，对于第 $n_1 + 2$ 个样本外检测期，我们将其对应的样本内估计期从 $\{Tail_t\}_{t=1}^{n_1}$ 滚动窗口到 $\{R_t\}_{t=1}^{n_1+1}$ ，并以此重新估计模型参数 $\hat{\alpha}_{n_1+2}$ 和 $\hat{\beta}_{n_1+2}$ 。然后使用更新后的模型对第 $n_1 + 2$ 个样本外检测期进行预测，以此类推，直到最后一个样本外检测期。

下一期的样本外预测模型为：

$$\hat{R}_{n_1+2} = \hat{\alpha}_{n_1+1} + \hat{\beta}_{n_1+1} Tail_{n_1+1} \quad (10)$$

其中， $\hat{\alpha}_{n_1+1}$ 和 $\hat{\beta}_{n_1+1}$ 是根据模型(5) 把 $\{R_t\}_{t=1}^{n_1}$ 对常数项和 $\{Tail_t\}_{t=1}^{n_1}$ 做线性回归得到的参数估计值。以此类推，不断扩展估计窗口，共可以得到 n_2 个样本外股票收益率的预测值 \hat{R}_{t+1} 。与基于尾部风险的预测模型(5) 作对比，基准预测模型为历史平均收益率 \bar{R}_{t+1} ：

$$\bar{R}_{t+1} = \frac{1}{t} \sum_{j=1}^t R_j \quad (11)$$

我们分别计算了预期持有期为 1 周的收益率，同时也尝试了预期持有期为 2、3、4 周的收益率。具体而言，未来 4 周的持有期收益率是从当前时间点 $t + 1$ 到 $t + 4$ 周的周度收益率的总和。

如果尾部风险包含对中国股票超额收益率的样本外预测信息，那么它的预测误差应该小于历史平均这一简单的预测，因为后者忽略了有用的预测信息。根据 Campbell & Thompson (2008a)，样本外 R^2 (R_{OS}^2) 的定义为：

$$R_{OS}^2 = 1 - \frac{\sum_{k=1}^{n_2} (R_{n_1+k} - \hat{R}_{n_1+k})^2}{\sum_{k=1}^{n_2} (R_{n_1+k} - \bar{R}_{n_1+k})^2} \quad (12)$$

当 $R_{OS}^2 > 0$ 时，表明 R_{t+1} 的样本外预测误差小于 \hat{R}_{t+1} 的样本外预测误差，即尾部风险具有更好的样本外预测能力。反之，说明尾部风险的样本外预测能力弱于历史平均预测。

我们使用 Clark & West (2007) 提出的 MSFE-adjusted 统计量对 $R_{os}^2 > 0$ 进行假设检验。原假设是预测性回归模型预测的均方预测误差和历史平均收益的均方预测误差相等，即 $R_{os}^2 < 0$ ；备择假设（单侧）是预测性回归模型预测的均方预测误差比历史平均收益预测的均方预测误差小，即 $R_{os}^2 > 0$ 。

（三）资产配置检验指标

夏普比率衡量的是每承担一单位风险，会获得多少超额收益，对投资者来说每增加一分的风险，能够增加几分的投资回报。夏普比率越高则策略效果越好。

夏普比率的临界值为 1，若夏普比率大于 1，则组合的收益率高于波动风险。夏普比率计算公式为：

$$\text{Sharpe Ratio} = \frac{E(R_p) - R_f}{R_f} \quad (13)$$

其中， $E(R_p)$ 与 R_f 、 R_f 分别代表预期年化报酬率，年化无风险利率与年化报酬率。

四、样本内预测表现

利用个股日内高频收益率信息提取的周度尾部风险能否较好预测股市的收益？依据厚尾冲击的资产定价理论，尾部风险同边际效用负相关且具有持续性。文中股市收益率指市场超额收益率，即为去除无风险利率的市场指数收益率。如无特别说明，以下变量均标准化，且为周度频率。我们计算发现，尾部风险因子同预期一周市场收益率两者的相关系数约为-0.23。整体上说，尾部风险同股市收益率之间有着较强的负相关关系。

（一）尾部风险因子单变量预测结果

1. 沪深 300 指数

首先，我们将每周内所有沪深指数 300 成分股的高频收益率序列汇总成周度数据池。在周度横截面数据池内，对高频收益率从低到高进行排序，得到第 t 期次序高频收益率统计量。将 Hill (1975) 的幂律估计器应用到时间 t 截面上，能够得到沪深 300 指数尾部风险因子 $Tail$ 。参照姜富伟等 (2011) 与陈坚和张轶凡 (2018) 相关文献，本文通过单变量预测回归模型来检验尾部风险对股票市场收益率的预测能力。首先，我们使用估计得到的尾部风险因子对沪深 300 指数未来持有期收益率（中位数方式计算）进行样本内预测。

据表 1 可知，尾部风险对不同期限的预期收益率中位数的回归系数显著为负。在所有不同持有期内，尾部风险（使用每 30、15、10、5 分钟的收益率计算）对预期收益率的回归系数在 1% 的置信水平下均显著，这意味着尾部风险可以作为一个有效的预测因子。

表 1 沪深 300 指数回报可预测性：单变量预测因子表现

自变量	未来 1 周持有期			未来 2 周持有期		
	Coeff.	t-value	R ²	Coeff.	t-value	R ²
Tail_{30min}	-0.0996***	-2.998	0.01	-0.1009***	-3.037	0.01
Tail_{15min}	-0.0822**	-2.469	0.01	-0.1227***	-3.704	0.02
Tail_{10min}	-0.1133***	-3.415	0.01	-0.1547***	-4.689	0.02
Tail_{5min}	-0.0601*	-1.802	0.00	-0.0957***	-2.878	0.01
ret_lag1	0.0618*	1.854	0.00	0.1296***	3.915	0.0168

(续表)

	未来 3 周持有期			未来 4 周持有期		
Tail_{30min}	-0.1545***	-4.683	0.02	-0.1524***	-4.62	0.02
Tail_{15min}	-0.1814***	-5.525	0.03	-0.1825***	-5.559	0.03
Tail_{10min}	-0.2047***	-6.263	0.04	-0.2106***	-6.451	0.044
Tail_{5min}	-0.1257***	-3.793	0.02	-0.1315***	-3.974	0.02
ret_lag1	0.1631***	4.952	0.03	0.1809***	5.508	0.03

注：该表格展示了沪深 300 指数市场尾部风险因子对未来 N 期持有期收益率（中位数）的单变量预测回归结果。

(1) Tail_30min 表示使用每 30 分钟数据频率下沪深 300 成分股收益率提取出的尾部风险因子。其余不同数据频率（15、10、5 分钟）下的尾部风险因子命名以此类推。

(2) ret_lag1 为行业投资组合滞后一期周度（中位数）收益率。

当预测的未来收益率持有期变长时，尾部风险对预期收益率的预测效果更为显著。例如，在预测未来四周持有期收益率，使用每 30 分钟、15 分钟、10 分钟、5 分钟的收益率计算而成的尾部风险对预期收益率的回归系数均显著，而且 R^2 也更大，拟合效果更好。尾部风险与未来股票收益呈显著的负相关关系，可以被用作预测预期收益率的指标，因为它捕捉到了市场可能出现的极端波动。

2. 行业分组投资组合

表 2 行业投资组合回报可预测性：单变量预测因子表现

行业投资组合	Panel A：预测因子为 Tail_15min			Panel B：预测因子为 Tail_10min		
	Coef.	t-value	R^2	Coef.	t-value	R^2
交通运输 I	-0.0062	-1.505	0.0025	-0.0067	-1.429	0.0022
休闲服务 I	-0.0125***	-2.614	0.0074	-0.0143***	-2.641	0.0076
医药生物 I	-0.0191***	-3.783	0.0154	-0.0187***	-3.516	0.0134
商业贸易 I	-0.0125***	-2.836	0.0087	-0.0086*	-1.73	0.0033
国防军工 I	-0.0065	-1.113	0.0014	-0.0053	-0.847	0.0008
家用电器 I	-0.0111***	-2.881	0.009	-0.0137***	-3.296	0.0118
建筑材料 I	-0.0116***	-2.621	0.0075	-0.0142***	-3.041	0.01
机械设备 I	-0.0106**	-2.247	0.0055	-0.0144***	-2.815	0.0086
汽车 I	-0.01**	-2.211	0.0053	-0.0153***	-3.1	0.0104
环保 I	-0.0102**	-2.237	0.0055	-0.0138***	-2.772	0.0084
电子 I	-0.0109**	-2.12	0.0049	-0.0164***	-2.975	0.0096
纺织服装 I	-0.0126***	-2.805	0.0086	-0.0132***	-2.67	0.0078
美容护理 I	-0.0078**	-2.381	0.0062	-0.0012	-0.361	0.0001
计算机 I	-0.0175***	-3.191	0.011	-0.021***	-3.622	0.0142
轻工制造 I	-0.0114**	-2.528	0.007	-0.0138***	-2.843	0.0088
通信 I	-0.0076*	-1.851	0.0037	-0.0097**	-2.193	0.0052
银行 I	-0.0038	-1.375	0.0021	-0.0022	-0.72	0.0006
非银金融 I	-0.0079*	-1.672	0.0031	-0.0065	-1.243	0.0017
食品饮料 I	-0.0076*	-1.815	0.0036	-0.0104**	-2.273	0.0056

注：该表格展示了行业投资组合尾部风险因子对未来 N 期持有期收益率（中位数）的单变量预测回归结果。

(1) Tail_30min 表示使用每 30 分钟数据频率下行业投资组合成分股收益率提取出的尾部风险因子。其余不同数据频率下的尾部风险因子命名以此类推。

(2) ret_lag1 为行业投资组合滞后一期周度（中位数）收益率。

参考姜富伟（2011）的方法，并加以改进，形成每月行业投资组合。我们将中国市场全 A 股股票以申银万国提供的分类为依据，在 20 个行业中形成价值加权投资组合。使用上个月月末市值计算股票权重，计算该行业投资组合的市值加权收盘价。依据构造沪深 300

指数尾部指标方法，构建每个行业尾部风险指标，探讨该指标是否对行业投资组合回报率的预测能力。

表 3 显示，在 15 分钟、10 分钟和 5 分钟频率计算下的尾部风险回归系数中，医药生物、建筑材料、休闲服务、商业贸易、家用电器、钢铁、计算机、轻工制造、通信和采掘行业的系数显著地小于零。

3. 市值分组投资组合

表 3 市值投资组合单变量分组特征（每 5 组平均值）

市值投资组合	Tail_30min	Tail_15min	Tail_10min	Tail_5min
size1-5	0.3799	0.3905	0.3898	0.3846
size6-10	0.3724	0.3849	0.3870	0.3822
size11-15	0.3699	0.3838	0.3843	0.3803
size16-20	0.3690	0.3800	0.3802	0.3761
size21-25	0.3682	0.3783	0.3787	0.3750
size26-30	0.3653	0.3779	0.3778	0.3687

注：该表格报告了每 5 组市值投资组合单变量分组特征（平均值）。Tail_30min 表示使用每 30 分钟数据频率下行业投资组合成分股收益率提取出的尾部风险因子。其余不同数据频率下的尾部风险因子命名以此类推。

表 1 为市值投资组合根据股票市值大小分组后的单变量分组特征表格（每 5 组平均值）。表格中列出了不同市值组合的分组尾部收益率的平均值。这个表格提供了一个对股票市值与尾部收益率之间关系的初步认识，在不同收益率频率下，随着投资组合的市值特征增大，尾部风险相应的降低。小市值的股票尾部风险较高。市值小的股票通常交易量较小，流动性较差，因此价格更容易受到个别交易者的影响，从而导致价格波动较大。通常小市值公司更具有创新性和成长性，但同时也面临着更大的经营风险和财务风险，这可能导致其尾部风险较高。

表 4 市值投资组合回报可预测性：单变量预测因子表现

市值投资组合	Panel A: Tail_15min			Panel B: Tail_10min		
	Coeff.	t-value	R ²	Coeff.	t-value	R ²
size1	-0.0091	-1.647	0.003	-0.0127**	-2.099	0.0048
size2	-0.0118**	-2.077	0.0047	-0.0158**	-2.585	0.0073
size3	-0.012**	-2.126	0.0049	-0.0136**	-2.214	0.0053
size4	-0.0089*	-1.765	0.0034	-0.0152***	-2.814	0.0086
size5	-0.0114**	-2.204	0.0053	-0.0184***	-3.317	0.0119
size6	-0.014***	-2.812	0.0086	-0.0194***	-3.653	0.0144
size7	-0.0139***	-2.61	0.0074	-0.0163***	-2.87	0.009
size8	-0.011**	-2.2	0.0053	-0.0175***	-3.219	0.0112
size9	-0.0115**	-2.006	0.0044	-0.0236***	-3.773	0.0154
size10	-0.0086*	-1.676	0.0031	-0.016***	-2.893	0.0091
size11	-0.0157***	-2.999	0.0098	-0.0181***	-3.243	0.0114
size12	-0.012**	-2.317	0.0059	-0.0174***	-3.042	0.01
size13	-0.0153***	-3.005	0.0098	-0.0169***	-2.97	0.0096
size14	-0.0112**	-2.271	0.0056	-0.0147***	-2.765	0.0083
size15	-0.0131**	-2.595	0.0073	-0.0124**	-2.322	0.0059
size16	-0.0128***	-2.603	0.0074	-0.0149***	-2.839	0.0088
size17	-0.0088*	-1.743	0.0033	-0.0129**	-2.378	0.0062
size18	-0.0136***	-2.808	0.0086	-0.0136**	-2.585	0.0073
size19	-0.0116**	-2.356	0.006	-0.0142***	-2.689	0.0079

(续表)

市值投资组合	Panel A: Tail_15min			Panel B: Tail_10min		
	Coeff.	t-value	R ²	Coeff.	t-value	R ²
size20	-0.0155***	-3.09	0.0104	-0.0132**	-2.471	0.0066
size21	-0.0126***	-2.604	0.0074	-0.0154***	-2.976	0.0096
size22	-0.0133***	-2.73	0.0081	-0.0129**	-2.467	0.0066
size23	-0.0122**	-2.409	0.0063	-0.0148***	-2.815	0.0086
size24	-0.008	-1.62	0.0029	-0.0108**	-2.021	0.0045
size25	-0.0107**	-2.139	0.005	-0.0124**	-2.292	0.0057
size26	-0.0107**	-2.246	0.0055	-0.013**	-2.55	0.0071
size27	-0.0145***	-3.157	0.0108	-0.0173***	-3.566	0.0138
size28	-0.0084*	-1.93	0.0041	-0.0136***	-2.835	0.0087
size29	-0.0049	-1.183	0.0015	-0.0106**	-2.415	0.0064
size30	-0.006*	-1.806	0.0036	-0.0063*	-1.849	0.0037

注：该表格展示了市值投资组合尾部风险因子对未来 N 期持有期收益率（中位数）的单变量预测回归结果。表示使用。Panel A 展示了使用每 15 分钟数据频率下市值投资组合成分股收益率提取出的尾部风险因子（Tail_15min）预测未来持有期一期收益率的样本内预测表现。Panel B 展示了使用 Tail_10min 作为预测因子的样本内预测表现。

表 3 展示了针对不同市值划分的投资组合的尾部风险因子对未来持有期中位数收益率单变量回归分析结果。表格列出了每个组合的回归系数、t 值和 R²。在每个月度股票数据池中，我们根据市值大小分为 30 组，从小到大排列(Size 1 - Size 30)。在不同市值投资组合内，使用上个月月末总市值作为权重形成投资组合。表格结果表明：市值投资组合尾部风险与组合回报之间存在负相关关系。具体而言，对于所有组合，尾部风险对未来一期周度收益率（中位数）的回归系数都显著为负，表明随着尾部风险的增加，投资组合收益率会下降。然而，在不同的组合之间尾部风险因子对中位数收益率的解释力度有所不同。在 10 分钟的频率下，Size 5-Size 10 的回归结果中 R² 的数值大于 1%，而 Size 20-Size 30 的 R² 的数值相比较小，一定程度上说明在小市值的投资组合的尾部风险对于预期收益率的拟合效果更好。

(二) 双变量预测表现

1. 尾部风险与滞后期收益率预测收益率中位数

为了对比尾部风险与经济变量的预测能力，姜富伟等（2011）提出了经济变量可以预测未来股票市场超额收益率。本文使用收益的各阶距作为尾部风险样本内预测能力的对比。表 5 显示了使用尾部风险和滞后期收益率作为预测变量，对中位数收益率进行双变量样本内预测的结果。表格中列出了不同预测持有期（一周、两周、三周和四周）内的系数、t 值和 R 平方值。

表 5 沪深 300 指数回报可预测性：双变量预测因子表现（尾部风险与滞后期收益率）

预测变量	Panel A: 未来 1 周持有期			Panel B: 未来 2 周持有期		
	Coeff.	t-value	R ²	Coeff.	t-value	R ²
Tail_30min	-0.0992***	-2.991	0.0137	-0.1001***	-3.037	0.0268
ret_lag1	0.0612*	1.845		0.129***	3.915	
Tail_15min	-0.0754**	-2.247	0.0094	-0.1076***	-3.238	0.0282
ret_lag1	0.0519	1.547		0.1155***	3.477	
Tail_10min	-0.1057***	-3.14	0.0147	-0.1362***	-4.087	0.0348
ret_lag1	0.0433	1.287		0.1059***	3.176	
Tail_5min	-0.0533	-1.589	0.0066	-0.081**	-2.436	0.0233
ret_lag1	0.0553	1.648		0.1197***	3.599	

(续表)

预测变量	Panel C: 未来 3 周持有期			Panel D: 未来 4 周持有期		
	Coeff.	t-value	R ²	Coeff.	t-value	R ²
Tail_30min	-0.1535***	-4.715	0.0502	-0.1514***	-4.663	0.0556
ret_lag1	0.1622***	4.982		0.18***	5.544	
Tail_15min	-0.1628***	-4.963	0.0527	-0.1615***	-4.939	0.0584
ret_lag1	0.1417***	4.321		0.1597***	4.882	
Tail_10min	-0.1818***	-5.522	0.0586	-0.1846***	-5.63	0.0658
ret_lag1	0.1314***	3.992		0.1487***	4.533	
Tail_5min	-0.1073***	-3.251	0.038	-0.1111***	-3.378	0.0449
ret_lag1	0.15***	4.544		0.1673***	5.086	

注：该表格展示了沪深 300 指数尾部风险因子、滞后期收益率对未来 N 期持有期收益率（中位数）的双变量预测回归结果；Tail_30min 表示使用每 30 分钟数据频率下沪深 300 指数成分股收益率提取出的尾部风险因子。其余不同数据频率下的尾部风险因子命名以此类推。

从表格中可以看出，滞后期收益率（ret_lag1）都对中位数收益率有显著的预测能力。此外，尾部风险（Tail_30min、Tail_15min、Tail_10min 和 Tail_5min）也对中位数收益率有预测作用，尤其是在较短的预测期限内。在所有预测期限内，尾部风险的系数都为负数，这表明尾部风险与中位数收益率之间存在负相关关系。尾部风险和滞后期收益率具有一定能力预测未来的收益率中位数。

2. 波动率与尾部风险因子预测中位数收益率

表 6 显示了使用尾部风险因子（Tail）和波动率（VR）作为预测变量，对不同预测期限（一周、两周、三周和四周）内中位数收益率进行双变量样本内预测的结果。

表 6 沪深 300 指数回报可预测性：双变量预测因子表现（波动率与尾部风险）

预测变量	Panel A: 未来 1 周持有期			Panel B: 未来 2 周持有期		
	Coeff.	t-value	R ²	Coeff.	t-value	R ²
Tail_30min	-0.0938***	-2.75	0.0105	-0.0961***	-2.817	0.0106
VR_30min	0.0257	0.754		0.0213	0.625	
Tail_15min	-0.0799**	-2.398	0.0081	-0.1199***	-3.613	0.0171
VR_15min	0.0351	1.054		0.0447	1.346	
Tail_10min	-0.1152***	-3.474	0.0162	-0.1572***	-4.775	0.0296
VR_10min	0.056*	1.689		0.0756**	2.296	
Tail_5min	-0.0652*	-1.956	0.0087	-0.1014***	-3.05	0.0153
VR_5min	0.056*	1.689		0.0756**	2.296	
预测变量	Panel C: 未来 3 周持有期			Panel D: 未来 4 周持有期		
	Coeff.	t-value	R ²	Coeff.	t-value	R ²
Tail_30min	-0.162***	-4.786	0.0249	-0.1636***	-4.834	0.0256
VR_30min	-0.0336	-0.991		-0.0497	-1.467	
Tail_15min	-0.1824***	-5.54	0.0331	-0.1845***	-5.609	0.0343
VR_15min	-0.0147	-0.448		-0.032	-0.973	
Tail_10min	-0.2056***	-6.287	0.0426	-0.211***	-6.457	0.0445
VR_10min	0.0274	0.838		0.0126	0.385	
Tail_5min	-0.1275***	-3.837	0.0164	-0.1323***	-3.984	0.0174
VR_5min	0.0249	0.748		0.0101	0.304	

注：该表格展示了沪深 300 指数尾部风险因子、波动率对未来 N 期持有期收益率（中位数）的双变量预测回归结果。Tail_30min 表示使用每 30 分钟数据频率下行业投资组合成分股收益率提取出的尾部风险因子。VR_30min 表示使用每 30 分钟数据频率下沪深 300 指数收益率计算的已实现波动率。其余不同数据频率（15、10、5 分钟）下的因子命名以此类推。

从表格中可以看出，波动率与尾部风险因子对中位数收益率的预测效果不理想。在所有预测期限内，尾部风险因子的系数都为负数，这表明增加的尾部风险与中位数收益率的下降之间存在正相关关系。然而，与尾部风险因子相比，波动率对中位数收益率的预测能力更弱。在所有预测期限内，VR的系数都不显著，这表明波动率本身并不能很好地解释收益率的变化。综上所述，尾部风险因子比波动率更适合用于预测收益率中位数。

五、样本外预测结论

相对于样本内预测，样本外预测对于现实世界投资者的实际投资决策更有意义。因此，这部分将考察尾部风险的样本外预测能力。借鉴 Campbell & Thompson (2008) 以及 Welch & Goyal (2008) 的研究方法，本文选择 2005 年 1 月至 2022 年 12 月作为样本外检验期，计算出样本外 R_{os}^2 及其 MSFE-adjusted 统计量和 p-value 统计量。

(一) 沪深 300 指数

首先，我们进行关于沪深 300 指数的样本外预测的研究，研究中使用了不同的滚动窗口大小和不同的预测期限，以评估尾部风险的样本外预测能力。表 7 中的结果显示，在大部分情况下，使用尾部风险模型进行样本外预测可以得到比基于历史平均值的简单预测更好的结果。

首先，在滚动窗口为 52 周，尾部风险样本外 R_{os}^2 达到 0.081%，且在 10% 显著性水平下统计显著。这说明尾部风险的样本外预测能力要显著优于基于历史平均值的简单预测。

表 7 沪深 300 指数样本外预测结论 (滚动窗口 = 52 周)

预期收益率持有期	预测变量	R_{os}^2	MSFE-adjusted	p-value
未来 1 周	Tail_10min	0.00081*	1.3801	0.08
	Tail_15min	0.001858*	1.3285	0.09
未来 2 周	Tail_5min	0.004176*	1.5256	0.06
	Tail_5min	0.010207*	1.4604	0.07

注：(1) 表格报告了基于沪深 300 市场指数在 1 周、2 周、4 周期间的预测回归的样本外预测表现。在每周 t (从 t = 52 开始，以允许足够大的初始估计期)，我们估计尾部风险序列和替代预测因子的周度市场回报的滚动单变量预测回归。

(2) “ R_{os}^2 为正”意味着预测变量的表现优于样本历史均值的预测。估计值旁边的星号(*)表示根据 the Clark and McCracken (2001) ENC-NEW 样本外检验，在 5% 或更高的水平上具有统计显著性。

此外，表格中还显示了不同滚动窗口大小和预测期限下的结果。根据表格中的数据，在不同的滚动窗口和预测时段下，尾部风险模型的预测能力存在显著差异。在滚动窗口为 12 周时，对于预测收益率持有期为 1 周和 2 周， R_{os}^2 的值均为正，但是 MSFE-adjusted 统计量的 p 值仍然小于 0.1，表明尾部风险模型的预测能力仍然优于历史平均值预测。而对于预测持有期为 3 周和 4 周， R_{os}^2 和 MSFE-adjusted 统计量的值都相对较高，表明尾部风险模型的预测能力更有效。

表 8 沪深 300 指数样本外预测结论 (滚动窗口 = 12、8 周)

预期收益率持有期	预测变量	R_{os}^2	MSFE-adjusted	p-value
Panel A: 滚动窗口 = 12 周				
未来 1 周	Tail_30min	0.0102	1.87	0.03
未来 2 周	Tail_10min	0.0251	1.36	0.09
未来 3 周	Tail_10min	0.0063	2.21	0.01
未来 4 周	Tail_5min	0.0401	1.59	0.06

(续表)

预期收益率持有期	预测变量	R_{os}^2	MSFE-adjusted	p-value
Panel B: 滚动窗口 = 8 周				
未来 1 周	Tail_30min	0.0194	6.33	0
未来 2 周	Tail_30min	0.0698	6.33	0
未来 3 周	Tail_30min	0.0804	6.33	0
未来 4 周	Tail_5min	0.0647	1.77	0.04

注：(1) 表格报告了基于滚动窗口（12 周、8 周），沪深 300 市场指数在 1 周、2 周、4 周期间的预测回归的样本外预测表现。在每周 t ，我们估计尾部风险序列和替代预测因子的周度市场回报的滚动单变量预测回归。

(2) R_{os}^2 意味着预测变量的表现不如设置等于样本均值的预测。估计值旁边的星号(*)表示根据 the Clark and McCracken (2001) ENC-NEW 样本外检验，在 5% 或更高的水平上具有统计显著性。

在滚动窗口为 8 周时 R_{os}^2 均为正，而且 MSFE-adjusted 统计量的 p 值都小于 0.05，意味着尾部风险模型在这些情况下的预测能力优于历史平均值预测。总的来说，这些结果表明使用尾部风险模型进行样本外预测可以提高预测能力，但具体效果受到滚动窗口大小和预测期限等因素的影响。

(二) 市值分组投资组合

表 9 行业分组投资组合可预测性：单变量预测因子样本外表现

预测变量	R_{os}^2	MSFE-adjusted	p-value
Panel A: 未来 2 周持有期			
Tail_10min	0.003782*	6.30	0
Tail_5min	0.002198*	7.66	0
Panel B: 未来 3 周持有期			
Tail_30min	0.009440*	3.64	0.0001
Tail_15min	0.010937*	3.39	0.0003
Tail_10min	0.014294*	2.60	0.0046
Tail_5min	0.012554*	4.31	0.0000
Panel C: 未来 4 周持有期			
Tail_30min	0.012015*	3.46	0.0003
Tail_15min	0.012610*	2.73	0.0031
Tail_10min	0.010658*	4.13	0.0000
Tail_5min	0.010817*	5.25	0.0000

注：(1) 表格报告了基于尾部风险因子对行业投资组合在 1 周、2 周、4 周期间的预测回归的样本外预测 R_{os}^2 的百分比。表格中 4 个面板分别显示了四个回归模型，分别为 Tail_30min、Tail_15min、Tail_10min 和 Tail_5min。在每周 t (从 $t = 52$ 开始，以允许足够大的初始估计期)，我们估计尾部风险序列和替代预测因子的周度市场回报的滚动单变量预测回归。

(2) R_{os}^2 为正意味着预测变量的表现优于样本历史均值的预测。估计值旁边的星号(*)表示根据 the Clark and McCracken (2001) ENC-NEW 样本外检验，在 5% 或更高的水平上具有统计显著性。

该表格显示了行业组合未来持有期回报的回归结果。共有三个面板，分别对应不同期限的未来持有期回报预测结果。每个面板中列出了多个回归模型，这些模型分别使用不同的尾部度量进行预测。表格的关键指标包括 R_{os}^2 、MSFE-adjusted 和 p-value。

从面板 A 开始，该面板显示了两个回归模型，Tail_10min 和 Tail_5min，这些模型用于预测未来 2 期持有期回报。它们的 R_{os}^2 分别为 0.003782 和 0.002198，这表明它们的样本外预测误差小于基准模型。面板 B 是关于未来 3 期持有期回报的样本外预测结果。该面板中显示了四个回归模型，分别为 Tail_30min、Tail_15min、Tail_10min 和 Tail_5min。这些模型预测因子的 R_{os}^2 值最高达到 0.014，显示出较好的样本外预测能力。所有模型的 MSFE-adjusted 值显示“ R_{os}^2 为正”是显著的， p 值都接近于 0。

最后，面板 C 是尾部风险对未来 4 期持有期回报的样本外预测结果。这些模型的

R^2_{os} 值相对较高，并且 MSFE-adjusted 和 p 值显示该结果是显著的，表明这些模型可以更好地拟合数据，并且可用于预测未来持有期回报。

综上所述，这个表格提供了有关行业组合未来持有期回报预测的关键信息。面板 B 和 C 的结果（即尾部风险对未来 4 期持有期回报的样本外预测）呈现出更好的样本外预测能力和更好的数据拟合，可以用于制定投资策略。

（三）市值分组投资组合

根据上月末的市值进行分组，根据大小分为 30 组，构建不同市值大小的加权投资组合，并进行样本外预测。

根据回归结果，我们可以看出，在控制了市值大小之后，不同市值大小的加权投资组合对未来持有期的收益表现有了不同的影响。为了更准确地评价这些市值组合的预测能力，我们使用了样本外 R^2_{os} 和 MSFE-adjusted 统计量。

表 10 市值分组投资组合可预测性：单变量预测因子样本外表现

预测因子	R^2_{os}	MSFE-adjusted	p-value
Panel A: 未来 2 周持有期			
Tail_15min	0.001119*	6.65	0
Tail_10min	0.006632*	4.58	0
Tail_5min	0.000784*	6.83	0
Panel B: 未来 3 周持有期			
Tail_30min	0.010538*	4.24	0
Tail_5min	0.008153*	4.72	0
Panel D: 未来 4 周持有期			
Tail_30min	0.012572*	4.11	0
Tail_10min	0.014751*	2.49	0.0063
Tail_5min	0.008020*	4.83	0

注：（1）表格报告了基于尾部风险因子对行业投资组合在 1 周、2 周、4 周期间的预测回归的样本外预测表现。在每周 t (从 t = 52 开始，以允许足够大的初始估计期)，我们估计尾部风险序列和替代预测因子的周度市场回报的滚动单变量预测回归。

（2） R^2_{os} 为正意味着预测变量的表现优于样本历史均值的预测。估计值旁边的星号(*)表示根据 the Clark and McCracken (2001) ENC-NEW 样本外检验，在 5% 或更高的水平上具有统计显著性。

根据市值分组投资组合可预测性（单变量预测因子样本外表现）的结果，我们可以看到，在未来收益持有期为 2 期时，所有市值组合的尾部风险样本外预测 R^2_{os} 值大于 0，表明它们的样本外预测误差小于历史平均预测误差，也就是说，市值组合的尾部风险能提供较好的样本外预测能力。然而，当持有期为 3 期和 4 期时，市值尾部风险的组合的 R^2_{os} 值均大 0，表明它们的样本外预测误差小于历史平均预测误差，也就是说，这些组合具有较好的样本外预测能力。

此外，我们使用 MSFE-adjusted 统计量对 $R^2_{os} > 0$ 进行了假设检验，以确认其统计显著性。根据检验结果，在持有期为 3 期和 4 期时，MSFE-adjusted 为 4.11，表明它们的预测能力达到了统计上的显著水平。而市值组合采用频率为 5min 的数据时，MSFE-adjusted 值为 4.83，说明尾部风险因子的样本外预测表现优于历史均值预测表现。

因此，从以上分析来看，在控制了市值之后，市值组合的尾部风险对未来收益的预测能力随持有期的增加而增强，而 10min 频率下的尾部风险对未来收益的表现相对较好。同时，通过样本外 R^2_{os} 和 MSFE-adjusted 统计量的评价，我们可以更加准确地评估这些市值组合的预测能力，以指导我们的投资决策。

六、资产配置

样本内和样本外检验已经表明尾部风险具有较强的预测能力，接下来进一步分析尾部风险因子对投资组合收益的提升作用。证券投资组合物为构造的各个行业投资组合。我们将下面三种方法构造证券投资组合：

1. 1/N 投资策略：该投资策略将资金平均分配到各个行业，每期回报率是所有行业回报率的算术平均值。

2. HS300 指数投资策略：该投资策略是一种简单而常见的投资策略，它的基本原理是将资金全部投入沪深 300 指数。沪深 300 指数是中国 A 股市场中规模较大的 300 家上市公司的股票组成的指数，代表了中国股市的整体表现。该投资策略的每期回报率就是沪深 300 指数的投资收益率。

3. 尾部指标筛选投资策略：基于市场行业投资组合尾部指标，通过每周尾部指标的预测值筛选掉尾部风险较大的行业投资组合，进而实现投资组合中集中配置低尾部风险行业的目标。具体来说，选择尾部指标最小的 n 个行业进行等权投资，剩余行业投资组合不进行投资。

表 11 行业分组投资组合资产配置策略结果

配置策略	Sharpe Ratio	Mean Return	Std Deviation
1/N	0.043649	0.00038	0.008735
HS300	0.039278	0.0014	0.03565
Panel A: n=15			
Tail_30min	0.041685	0.0004	0.009478
Tail_15min	0.046211	0.00041	0.008769
Tail_10min	0.048973	0.00043	0.008785
Tail_5min	0.047706	0.00042	0.008768
Panel B: n=10			
Tail_30min	0.051898	0.00046	0.008789
Tail_15min	0.053679	0.00048	0.008935
Tail_10min	0.044828	0.00039	0.008739
Tail_5min	0.049964	0.00044	0.008752
Panel C: n=5			
Tail_30min	0.054084	0.00047	0.008764
Tail_15min	0.051965	0.00046	0.008757
Tail_10min	0.05921	0.00053	0.008949
Tail_5min	0.043777	0.00038	0.008722

注：(1) 该表格展示了行业分组投资组合尾部风险因子资产配置策略及传统的 1/N 策略、HS300 指数投资策略的资产配置表现，并计算了夏普比率、平均回报率及风险（标准差）。表格中 Panel A - D 展示了筛选行业股票池不同数量的尾部指标筛选投资策略。该策略选择尾部指标最小的 n 个行业进行等权投资，剩余行业投资组合不进行投资。

(2) 表格每个分面板中包含四种不同频率尾部风险因子的投资表现。Tail_30min 表示使用每 30 分钟数据频率下行业投资组合成分股收益率提取出的尾部风险因子。其余不同数据频率（15、10、5 分钟）下的因子命名以此类推。

我们展示了行业分组投资组合资产配置策略结果。每种策略的表现通过计算 Sharpe 比率、平均回报率和波动率三项指标得出，这些指标可以帮助我们理解每种投资策略的风险和回报特征。并和同期沪深 300 指数回报率的投资表现做比较。这有助于投资者在实践中选择合适的投资策略，并优化资产配置，提高投资组合的风险收益特征。

与和同期沪深 300 指数回报率的均值，方差及 Sharpe ratio 比较，我们可以发现，沪深 300 指数的夏普比率 0.0529，平均回报率为 0.0079，标准偏差为 0.0004。相比于尾部风险

策略和 1/N 策略，沪深 300 的表现稍逊一些，说明尾部风险策略和 1/N 策略的表现相对更好。数据表明，在所有样本量和尾部指标预测时间范围下，使用尾部指标策略构造的投资组合的表现都要好于使用 1/N 策略构造的投资组合。此外，在较小的样本量下，尾部指标预测时间范围较短（5 分钟）的策略表现相对较好，而在较大的样本量下，尾部指标预测时间范围较长（30 分钟）的策略表现更为优秀。

具体而言，尾部指标策略的平均回报率、Sharpe Ratio 均优于 1/N 策略。在样本量 $n=25$ 时，尾部指标策略的 Sharpe Ratio 约为 0.063 大于 1/N 策略，说明尾部指标策略可以在降低风险的同时获得更高的收益。尾部风险策略的平均回报率和夏普比率都比 1/N 策略高，同时标准偏差也较低。这表明，尾部风险策略可以获得更高的收益，同时风险也更小，因此尾部风险资产配置方案表现更好。

此外，根据表格，随着面板数量的增加，投资组合的表现有所下降。这是因为投资组合中不同的行业之间可能存在相关性，当面板数量太多时，相关性可能会变得更明显，从而影响整个投资组合的表现。最后，我们可以发现，尾部风险策略在不同的尾部时间间隔中表现的差异很小。

表 12 市值分组投资组合资产配置策略结果

配置策略	Sharpe Ratio	Mean Return	Std Deviation
1/N	0.0984	0.0010	0.0098
HS300	0.0393	0.0014	0.0357
Panel A: $n=25$			
Tail_30min	0.0975	0.0010	0.0099
Tail_15min	0.0984	0.0010	0.0100
Tail_10min	0.0975	0.0010	0.0100
Tail_5min	0.0928	0.0009	0.0099
Panel B: $n=20$			
Tail_30min	0.0908	0.0009	0.0101
Tail_15min	0.0998	0.0010	0.0098
Tail_10min	0.1029	0.0010	0.0098
Tail_5min	0.1025	0.0010	0.0098
Panel C: $n=15$			
Tail_30min	0.1050	0.0010	0.0100
Tail_15min	0.1109	0.0011	0.0099
Tail_10min	0.0990	0.0010	0.0098
Tail_5min	0.1018	0.0010	0.0098
Panel D: $n=10$			
Tail_30min	0.1052	0.0010	0.0099
Tail_15min	0.1104	0.0011	0.0098
Tail_10min	0.1046	0.0010	0.0094
Tail_5min	0.1003	0.0010	0.0098

注：（1）该表格展示了市值分组投资组合尾部风险因子资产配置策略及传统的 1/N 策略、HS300 指数投资策略的资产配置表现，并计算了夏普比率、平均回报率及风险（标准差）。表格中 Panel A - D 展示了筛选市值投资组合股票池不同数量的尾部指标筛选投资策略。该策略选择尾部指标最小的 n 个行业进行等权投资，剩余行业投资组合不进行投资。

（2）表格每个分面板中包含四种不同频率尾部风险因子的投资表现。Tail_30min 表示使用每 30 分钟数据频率下行业投资组合成分股收益率提取出的尾部风险因子。其余不同数据频率（15、10、5 分钟）下的因子命名以此类推。

表 21 是关于市值分组投资组合资产配置策略结果。表格中给出了五个面板（Panel A 至 Panel E），每个面板中包含不同数据频率（从 30 分钟到 5 分钟）的投资组合表现。对于每个投资组合，表格提供了三个指标：夏普比率、平均回报和标准差。表格中还列出了两个

基准策略：1/N 和 HS300 策略。使用尾部风险因子作为预测变量的投资组合策略的 Sharpe 比率比使用 1/N 策略和 HS300 策略的 Sharpe 比率都要高，这表明使用尾部风险因子作为预测变量可以提高投资组合的表现。随着行业尾部风险策略的选取股票池数量的减少，投资组合的表现也有所上升。例如，在 Panel A 中，使用 25 个行业分组的投资组合的 Sharpe 比率比使用 5 个行业分组的投资组合的 Sharpe 比率低。

七、结论

参考 Kelly & Jiang(2014)的面板估计方法，本文研究了中国 A 股市场“左尾风险异象”在中国股市中的对股票未来回报影响机制的研究。具体而言，本文首先选取中国 A 股市场中 2005 年至 2022 年上市公司股票的日内高频收益数据，并以此计算左尾部风险指标 (Tail) 及相应的控制变量。本文通过构建不同种类的投资组合，较为全面地探索了左尾部风险与股票截面收益之间的关系，并且还考察了左尾部风险的样本外预测能力。综合上述实证分析过程，本文得出如下研究结论与建议。

样本内预测：在沪深 300 指数、及市值、行业划分的投资组合的样本内预测中，尾部风险与未来股票收益呈显著的负相关关系，证明尾部风险可以作为一个有效的预测因子，对不同期限的预期收益率中位数具有一定的预测能力。在预测回报持有期增加时，尾部风险对预期收益率的预测效果更为显著。因此，尾部风险可以被用作预测预期收益率的指标，捕捉到市场可能出现的极端波动。

样本外预测：我们进行关于沪深 300 指数、及市值、行业划分的投资组合的样本外预测，研究中使用了不同的滚动窗口大小和不同的预测期限，以评估尾部风险的样本外预测能力。实证结果显示，使用尾部风险模型进行样本外预测可以得到比基于历史平均值的简单预测更好的结果。

资产配置：使用尾部风险因子作为筛选变量可以提高投资组合的表现。我们使用了中国 A 股行业、市值行业投资组合，并进行尾部风险策略配置，由此投资组合平均回报率和夏普比率比 1/N、HS300 投资基准策略高，同时标准偏差也较低。尾部风险作为股票投资中不可避免的系统性风险之一，无法通过分散化投资进行消除，每位投资者都面临着其带来的不确定性。尾部风险策略可以获得更高的收益，同时风险降低。对尾部风险与股票收益的关系及其影响因素的深入研究，有助于为投资者提供更为准确有效的资产配置策略。

参考文献

Amaya, D., Christoffersen, P., Jacobs, K. and Vasquez, A., 2015, "Does Realized Skewness Predict the Cross-Section of Equity Returns?", *Journal of Financial Economics*, 118(1): 135-167.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P., 2003, "Modeling and Forecasting Realized Volatility", *Econometrica*, 71(2): 579-625.

Bali, T. G., Cakici, N. and Whitelaw, R. F., 2014, "Hybrid Tail Risk and Expected Stock Returns: When Does the Tail Wag the Dog?", *The Review of Asset Pricing Studies*, 4(2): 206-246.

Bali, T. G., Demirtas, K. O. and Levy, H., 2009, "Is there an Intertemporal Relation Between Downside Risk and Expected Returns?", *Journal of Financial and Quantitative Analysis*, 44(4): 883-909.

Bali, T. G., Engle, R. F. and Murray, S., 2016, *Empirical Asset Pricing: The Cross Section of Stock Returns*, John Wiley & Sons.

Campbell, J. Y. and Thompson, S. B., 2008a, "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?", *The Review of Financial Studies*, 21(4): 1509-1531.

Campbell, J. Y. and Thompson, S. B., 2008b, "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?", *The Review of Financial Studies*, 21(4): 1509-1531.

Campbell, J., 2000, "Asset Pricing at the Millennium", *Scholarly Articles*.

Clark, T. E. and West, K. D., 2007, "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models", *Journal of Econometrics*, 138(1): 291-311.

Fama, E. F., 1963, "Mandelbrot and the Stable Paretian Hypothesis", *The Journal of Business*, 36(4): 420-429.

Gabaix, X., 2012, "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance", *The Quarterly Journal of Economics*, 127(2): 645-700.

Glasserman, P. and Young, H. P., 2016, "Contagion in Financial Networks", *Journal of Economic Literature*, 54(3): 779-831.

Habib, A., Hasan, M. M. and Jiang, H., 2018, "Stock Price Crash Risk: Review of the Empirical Literature", *Accounting & Finance*, 58: 211-251.

Hill, B. M., 1975, "A Simple General Approach to Inference About the Tail of a Distribution", *The Annals of Statistics*: 1163-1174.

Huang, W., Liu, Q., Rhee, S. G. and Wu, F., 2012, "Extreme Downside Risk and Expected Stock Returns", *Journal of Banking & Finance*, 36(5): 1492-1502.

Kelly, B. and Jiang, H., 2014, "Tail Risk and Asset Prices", *The Review of Financial Studies*, 27(10): 2841-2871.

Liu, J., Stambaugh, R. F. and Yuan, Y., 2019, "Size and Value in China", *Journal of Financial Economics*, 134(1): 48-69.

Ludvigson, S. C. and Ng, S., 2009, "Macro Factors in Bond Risk Premia", *The Review of Financial Studies*, 12(22): 5027-5067.

Mandelbrot, B., 1963, "New Methods in Statistical Economics", *Journal of Political Economy*, 71(5): 421-440.

Rietz, T. A., 1988, "The Equity Risk Premium a Solution", *Journal of Monetary Economics*, 22(1): 117-131.

Welch, I. and Goyal, A., 2008, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction", *The Review of Financial Studies*, 21(4): 1455-1508.

陈国进、许秀、赵向琴， 2015： 《罕见灾难风险和股市收益——基于我国个股横截面尾部风险的实证分析》，《系统工程理论与实践》，第 9 期。

陈坚、张轶凡， 2018： 《中国股票市场的已实现偏度与收益率预测》，《金融研究》，第 9 期。

姜富伟、涂俊、Rapach David E.、Strauss Jack K.、周国富， 2011： 《中国股票市场可预测性的实证研究》，《金融研究》，第 9 期。

Tail Risk and Return Prediction in Chinese A-Share Market: Evidence from Cross-sectional Data of HS300 Index

CHEN Si and REN Yu

(Wenlan School of Business, Zhong Nan University of Economics and Law)

Abstract: In this paper, we construct a proxy for the tail risk in Chinese A-share market by following the method in Kelly and Jiang (2014). Specifically, we select the daily return of all stocks from January 1995 to December 2022, and estimate the monthly tail risk index of the market according to the empirical distribution of the tail risk occurred to individual stock. We use portfolios such as HS300 index in the Chinese market, and China A-share portfolio sorted on industry and size. The empirical results show that there is a significant negative relationship between tail risk and expected portfolio returns. We select the realized volatility of stock returns, realized skewness and other high-order moment indicators as predictors for comparison, and use the rolling window regression method to evaluate the out-of-sample predictive ability of tail risk. The out-of-sample results show that using the tail risk model for out-of-sample forecasting can yield better results than simple forecasting based on historical averages. In terms of asset allocation, it has been studied that using tail risk factors as screening variables can improve portfolio performance. This strategy screens out the industry portfolio with high tail risk through the predicted value of weekly tail index, and then achieves the goal of concentrating the allocation of low tail risk industry in the portfolio. The average return and Sharpe ratio of the portfolio with the tail risk strategy are higher than that of the 1/N strategy, while also being less risky.

Keywords: Tail Risk, In-sample Return Predictability, Out-sample Return Predictability, Asset Allocation

JEL Classification: C22, C53, G12