

doi:10.19920/j.cnki.jmsc.2026.04.004

基于高维因子模型的A股市场系统性风险研究^①

魏照昊^{1,5}, 李秀婷^{1,2}, 卢全莹^{3,4}, 董纪昌^{1,2}, 董志^{1,2*}

(1. 中国科学院大学经济与管理学院, 北京 100190; 2. 中国科学院大数据挖掘与知识管理重点实验室, 北京 100190; 3. 北京工业大学经济与管理学院, 北京 100124; 4. 中国科学院预测科学研究中心, 北京 100190; 5. 申万宏源证券有限公司博士后科研工作站, 上海 200031)

摘要: 面临高维特征的挑战以及特设稀疏性假设的诸多弊端, 本文参照高维因子统计推断领域的前沿研究进展, 从隐性因子结构出发构建了针对于A股市场具有更强资产定价能力且更具一般代表性的系统性风险因子组合, 并对其相应的因子特征及补偿定价结构进行了全面拆解及深入分析. 研究结果显示: 1) 与CH-4以及Fama-French多因子定价模型相比, 基于高频的RP-PCA因子组合捕获了更多的系统性风险特征, 拥有着更好的定价表现以及更为稳健的时变载荷结构; 2) A股市场中蕴含着五个稳定的系统性风险因子. 除市场因子外, 其余四个因子分别可由以金融为代表的六个特定细分行业的投资组合近似表示; 3) A股市场的系统性风险未被完全定价, 风险敞口的非对称性特征显著存在. 整体来看, 投资者或更加关注于系统上涨而个股下跌以及系统及个股同时下跌的风险.

关键词: 系统性风险; 高维协方差矩阵; 高频数据; 因子模型; SDF

中图分类号: F832.5 **文献标识码:** A **文章编号:** 1007-9807(2026)04-0066-19

0 引言

自资本资产定价模型^[1,2]及套利定价理论^[3]提出以来, 如何科学的测度系统性风险及其补偿结构一直是资产定价领域的核心问题^[4,5]. 近年来, 随着金融市场的全面开放以及直接融资渠道的加速改革, A股市场的风险结构呈现出了愈发复杂的特征. 基于此, 众多学者对Fama-French等^[6-8]传统多因子定价模型的现实适用性提出了质疑并对由此所产生的海量异象展开研究, 以期改进因子的构造方式或补充其所遗漏的风险特征, 从而获取基准定价模型中无法解释的超额收益^[9-21].

类似研究虽然在较大程度上对已有因子进行了改进, 但却暴露出了两方面的关键问题. 一方面, 在针对个别因子的研究中, 以往研究均大多以

Fama-French多因子模型为基准, 其额外的定价能力也仅限于以此为基础的个别异象解释. 近年来, 随着更多的异象被不断挖掘, 产生了众多具有额外定价能力的新因子. 但是, Feng等^[19]却指出这些所谓的新因子中的大多数均为伪因子或是对于特定风险的冗余描述. Chordia等^[20]、Harvey和Liu^[21]的研究也进一步印证了这一观点. 针对这一“因子动物园”乱象, Pukthuanthong等^[22]及Kozak等^[23]强调系统性因子的判别标准应重点聚焦在其与资产收益协方差矩阵间的关系. 与此同时, Fama和French^[24]也强调了类似的问题并进一步指出应基于因子组合而非单个因子的视角以捕获系统性风险的重要性. 另一方面, 在对因子组合的研究中, 以往的经济因子模型中往往存在着因子结构的先验特征设定偏误、遗漏变量以及模

① 收稿日期: 2022-08-01; 修订日期: 2024-09-15.

基金项目: 国家自然科学基金资助项目(71850014; 72334006; 72004214); 上海市“超级博士后”激励计划资助项目(2023066).

通讯作者: 董志(1988—), 男, 山东烟台人, 博士, 副教授. Email: dongzhi@ucas.ac.cn

型不确定性等严重缺陷^[25]。尤其是在高维环境中,以期通过以往经验直接给定各因子的先验结构将愈发困难,类似问题或将更加突出。

为了更好的解决上述诸多问题,近年来基于高维(high-dimensional)统计技术的因子构造方式愈发流行。首先,面临投资组合维度的大幅提升,Bai^[27]基于样本协方差矩阵提出了一种针对于静态高维因子模型的统计推断理论。在此基础上,Fan等^[28]及Nagel^[29]指出在高维特征环境中,利用个别经济特征并不能够实现风险结构的有效降维。即从统计视角再次明确了以往经济因子定价模型在先验特征结构假设上的现实缺陷。随后,Fan等^[30]还进一步改进了以往模型中所依赖的严格因子模型假设。即通过对特质成分(idiosyncratic component)协方差矩阵引入稀疏结构从而给出了多种范数下的统计因子求解范式。随后,其更是在理论上证明了PCA方法在高维协方差矩阵中可获取与多因子定价模型渐进一致估计结果的研究结论^[31]。这一发现不仅解决了在高维特征环境中面临截面相关、时序相关及异质性问题时传统因子定价模型难以测度的关键缺陷,与此同时其也进一步推动了PCA方法在高维因子模型中的广泛运用^[32-34]。其次,考虑到静态因子结构假设的局限性,部分学者还进一步的探讨了在因子定价过程中所可能存在的动态演变特征,并进一步优化了相应因子的构造方式^[35-39]。

除以上针对于高维因子的研究外,近年来面临金融数据的爆发式增长,部分学者还进一步结合高频(high-frequency, HF)交易信息对系统性风险的定价过程进行了更为细致的探讨。目前虽然在这一研究方向上的相关成果仍相对较少,但是综合利用高频计量及高维统计推断技术的因子定价研究已然成为了这一领域的前沿方向。具体而言,首先,考虑到低频样本中的噪声干扰,Pelger^[5]基于PCA理论提出了一种更为通用的因子测度方法并证明了适度增加采样频率或可更为有效的捕捉各定价因子的微观结构,从而改善其对于系统性风险的解释能力。类似的,Ait-Sahalia等^[40]也得到了较为一致的研究结论。其次,Ait-Sahalia和Xiu^[41]在高维高频的交易环境中指出资产收益的协方差矩阵中或将存在着明显的低秩稀疏结构,并且这一结构特征会随着资产维度的

增加而愈发显著。随后其还进一步给出并发展了连续时间过程下的高频PCA因子模型理论^[42]。在此基础上,Pelger^[43]通过对高频跳变成分的分,进一步细化了系统性风险的定价过程并给出了不同风险成分的识别条件及其相应的因子构造方法。值得注意的是,以往的类似研究虽然证明了在因子定价过程中所存在的不同定价成分,但是在不同成分间的定价能力比较上目前却还暂未有统一论^[5, 44-47]。

综上所述,依托于特设稀疏性(ad-hoc sparsity)假设前提,近年来众多学者多致力于对个别特征或行为现象的定价研究,尤其是针对于以CAPM或Fama-French为基准的新因子构造或测度。但是,随着投资组合维度的不断攀升,以往经济因子模型不仅在因子的先验结构、载荷构建方式、评价标准以及特征不确定性上暴露出了诸多关键问题,更为重要的是以此为基础而所构建的各类新风险因子其自身所存在的较高的伪发现概率常常致使各系统性成分难以在此框架下得以科学解释。因此,为摒弃对于特征结构的主观设定,近年来依托于隐性结构的统计因子模型愈发流行。尤其是针对于高维高频统计推断这一前沿领域,目前虽然形成了较多的理论成果,但是其相关理论的现实应用探讨仍相对较少。特别是针对于各因子及其载荷结构的经济含义、机理解释及实际可投资性等方面的讨论上,仍具有较大的争议和改进空间。基于此,本文从全局视角出发,依托于Pelger^[43]及Lettau和Pelger^[4, 39]等在高维因子统计推断领域的前沿研究成果,在SDF(stochastic discount factor)的一般资产定价框架下详细探讨了A股市场的系统性风险结构及其补偿定价过程,并与以往各类代表性资产定价模型进行全面对比。

本文的主要贡献在于:1)从全局视角构建了具有更强资产定价能力的系统性风险因子组合。区别于以往研究,各因子的构造过程仅依托于隐性结构,即摆脱了对于传统特设稀疏性的强假设依赖。在SDF框架下给出了针对于A股市场更具一般代表性的因子结构;2)综合时序和截面定价误差,分频度、成分、分因子的系统考察了A股市场的系统性风险演变过程及其在不同交易场景下的定价特征及结构差异,多角度揭示了A股市场的潜在风险来源及

其自身所具有的非对称风险敞口、时变定价结构等复杂特征. 通过更贴合于现实交易规则的因子处理框架以及基于高维特征矩阵的低秩稀疏结构处理方法有效弥补了以往研究的不足, 为相关研究提供了一种新颖的分析框架及研究思路.

1 模型设定及因子提取

1.1 因子模型设定

同 Pelger^[43] 的研究, 本文假设资产定价的一般过程可由下式所述的因子模型所表示

$$P_i(t) = \Lambda_{i,s}^T F(t) + e_i(t), \quad i=1, \dots, N, t \in [T_s, T_{s+1}], s=1, \dots, S \quad (1)$$

其中 $P(t)$ 代表股票对数价格. $F(t)$ 和 Λ 分别代表系统性因子及其相应的载荷结构, $e(t)$ 为特质成分. 依据模型设定, $\Lambda_{i,s}$ 描述了第 i 支股票在第 s 个固定时间区间中的风险暴露情况, 并且这一风险暴露还将在 $[T_s, T_{s+1}]$ 的局部窗口中维持恒定. 考虑到在不同时间区间中可能存在的因子结构差异, 因此在局部时间窗口的选取上, 本文分别设定了月度、季度、年度以及全样本时期下的多种不同场景. 式(1)整体为不同因子模型间的对比提供了统一的理论框架, 针对于给定的局部时间窗口 $[0, T]$, 其可进一步简化为如下形式

$$P_i(t_j) = \Lambda_i^T F(t_j) + e_i(t_j), \quad i=1, \dots, N, t \in [0, T], j=0, \dots, M \quad (2)$$

其中 N 为截面观测的样本数量, M 为多维随机过程 $P(t)$ 时序观测频度. $t_0 = 0, t_1 = \Delta_M, t_2 = 2\Delta_M, \dots, t_M = M\Delta_M$ 代表了 $P(t)$ 的多个离散时间采样窗口, 并有 $\Delta_M = t_{j+1} - t_j = \frac{T}{M}$. 考虑到在现实交易中因市场震荡或其他外部冲击下所可能带来的价格非连续性, 因此在 $P(t)$ 的设定上, 本文假设其在不包含微观结构噪声的前提下服从局部有界的 Itô 半鞅随机过程. 已有研究表明, 这一半鞅结构不仅可以较好的拟合股票价格的现实走势, 同时其也适用于资产定价结构更为复杂的高频交易环境^[48]. 此外, 这一 Itô 半鞅的假设前提同时也是在高维资产定价环境中获取系统因子及其载荷

结构渐进分布的重要基础^②

$$P(t) = P(0) + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \int_R xv(ds, dx) \quad (3)$$

如式(3)所示, $P(t)$ 的连续价格部分可被拆解为一个可预测的漂移项 μ_t 以及一个服从 N 维标准布朗运动 W_t 的连续时间鞅, 而其价格的跳跃部分则可由一个复合泊松过程所表示. 其中, $v(ds, dx)$ 表示股票价格在局部时间区间 ds 内跳跃幅度为 dx 的跳跃次数^③. 在此基础上, 考虑到在高维特征空间中所面临的“维数灾难”问题, 本文还进一步的引入了由 Chamberlain 和 Rothschild^[49] 所构建的近似因子定价结构. 综上所述, 在 APT 假设下, 资产收益的因子定价模型可总结如下

$$R_{(M \times N)} = F_{(M \times K)} \Lambda_{(K \times N)}^T + e_{(M \times N)} \quad (4)$$

上式中 $F_j = F(t_{j+1}) - F(t_j)$, $e_{j,i} = e_i(t_{j+1}) - e_i(t_j)$, $R_{j,i} = (P_i(t_{j+1}) - P_i(t_j)) - r_j(t_j)$. 其中 r_j 为无风险收益率. 在近因结构假设下, e 中能够存在较弱的截面相关或序列相关并且其二次协变差矩阵 (quadratic covariation matrix) 的最大特征值将依概率有界, 即 $\lambda_1([e, e]_T) = O_p(1)$. 在 $N \rightarrow \infty$ 的高维环境中, 有 $\sum_{i=1}^N \| [e_k, e_i]_T \| = O_p(1)$. 其中, 在对二次协变差矩阵 $[e, e]$ 的测算上, 对于高频交易中所任意指定的 i_1 和 i_2 , $\lim_{M \rightarrow \infty} \sum_{j=1}^M e_{j,i_1} e_{j,i_2} \xrightarrow{P} [e_{i_1}(t), e_{i_2}(t)]_T$. 不同于以往的经济因子构造方式, 在本文的研究中, 式(4)始终保持着隐性结构. 即除 R 外, 其因子载荷 Λ 、系统性因子 F 、特质成分 e 以及因子个数 K 均为待估参数且不可直接观测获取. 此外, 为保证 $F\Lambda^T$ 对于系统成分的充分拟合, 式(4)中还进一步补充添加了 $\sum_{j=1}^M F_j F_j^T \xrightarrow{P} [F, F]_T = : \Sigma_F$ 及 $\left\| \frac{\Lambda^T \Lambda}{N} - \Sigma_\Lambda \right\| \rightarrow 0$ 的额外约束.

特别的, 考虑到不同收益成分间的定价结构差异^[5, 44], 式(4)还可进一步转换为如下形式

② 理论证明部分详见 Pelger^[38].

③ 假设在局部时间窗口中仅存在着有限次的跳跃过程.

$$R(t) = \mathbf{A}^{CT} F^C(t) + \mathbf{A}^{DT} F^D(t) + \mathbf{A}^{NT} F^N(t) + e(t) \quad (5)$$

其中 $R(t)$ 被整体拆分为了日内及隔夜收益两部分. 而日内收益又可被进一步拆分为日内连续和日内跳跃收益两部分. 在以上的因子结构中, 日内连续载荷 \mathbf{A}^C 、日内跳跃载荷 \mathbf{A}^D 及隔夜载荷 \mathbf{A}^N 三者间均具有各自独立的特征结构, 系统性因子 F^C 、 F^D 及 F^N 的设定同理.

1.2 因子识别及模型估计

为保证对于具有较低“信号强度”因子的有效识别, 依托 Lettau 和 Pelger^[4, 39] 的研究, 本文利用可综合考虑时序及截面定价误差的 RP-PCA (Risk-Premium PCA) 模型对 A 股市场的系统性风险特征进行研究. 大量的算例分析结果表明, 一方面, 相较于经济因子定价模型^④, RP-PCA 无需指定先验特征结构, 从而有效避免了对于先验结构的设定偏误、遗漏变量以及模型不确定性等问题. 另一方面, 相较于以高维 PCA 为代表的隐性因子学习方法, RP-PCA 通过对截面定价误差信息的引入, 可实现对于真实 SDF 更为有效的拟合^[4]. 其理论模型如下所示

$$\hat{F}, \hat{\mathbf{A}} = \underset{\mathbf{A}, F}{\operatorname{argmin}} \frac{1}{NM} \sum_{i=1}^N \sum_{j=0}^M (X_{ij} - F_j \mathbf{A}_i^T)^2 + \gamma \frac{1}{N} \sum_{i=1}^N (\bar{X}_i - \bar{F} \mathbf{A}_i^T)^2 \quad (6)$$

其中 $\gamma \geq -1$, 描述了截面定价误差相对于时序误差的权重占比. 针对于式(6)的目标函数, 其求解过程可进一步转化为对于 $\Sigma_{RP} = \frac{1}{M} R^T R + \gamma \bar{R} \bar{R}^T$ 的谱分解过程^[39], 并有 $\hat{F} = R \hat{\mathbf{A}} (\hat{\mathbf{A}}^T \hat{\mathbf{A}})^{-1}$. 其中, 因子载荷 $\hat{\mathbf{A}}$ 与 Σ_{RP} 的前 k 个最大特征值所对应的特征向量成正比, \bar{R} 为股票超额收益的样本均值. 而对于参数 k 的估计, 其则可通过对于扰动特征值比率聚类特征变化的观测从而获取^[43]. 具体而言, 假设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ 为矩阵 Σ_{RP} 从大到小依次排列的非零特征值, 则其扰动特征值比率可被定义为

$$ER_k = \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}}, \quad (7)$$

$$k = 1, \dots, N - 1$$

$$\hat{\lambda}_k = \lambda_k + \sqrt{N} \times \operatorname{median} \{ \lambda_1, \dots, \lambda_N \}.$$

在此基础上, 参数 k 的判断标准可进一步表述为

$$\hat{k}(\iota) = \max \{ k \leq N - 1 : ER_k > 1 + \iota \} \quad (8)$$

其中 ι 为截断参数且 $\iota > 0$ ^⑤. 在 $\lambda_1([e, e]_T) = O_p(1)$ 的约束下, 如若扰动特征值比率在 1 左右停止聚类, 此时因子则将逐步进入系统谱内, 否则其将持续位于非系统谱中. 相较于 Onatski^[50] 及 Ahn 和 Horenstein^[51] 等所提出的高维因子估计方法来说, 这一评判标准不仅可摆脱对于随机矩阵理论的强假设依赖, 而且其还能够更好的适用于强弱系统因子混叠共存的复杂环境^[43].

依托于以上估计结果, 基于 SDF 的一般资产定价框架, 通过对于均值方差有效边界中最优因子投资组合所能获取的最大夏普比率的测度, 即可给出不同因子组合间对于系统性风险捕获能力的实际差异情况及具体判定依据. 其中, $\hat{b}_{MV} = \sum_F^{-1} \mu_F$ 刻画了各因子在最优投资组合中的权重分布情况, 而 $M_i = 1 - \hat{b}_{MV}^T (\hat{F}_i - E[\hat{F}_i])$ 则描述了其所相对应的隐含 SDF^[39]. 在此基础上, 考虑到线性因子组合的旋转不变性, 在本文的研究中进一步引入了由 Bai 和 Ng^[52] 所构建的典型相关 (Canonical correlation) 算法从而系统对比不同因子组合及其载荷结构的向量空间差异情况^⑥并详细分析了各因子的潜在经济含义及其定价逻辑.

2 参数估计及模型定价能力比较

2.1 样本选取与数据处理

为了规避个别特征指数的样本偏差, 在样本

④ 限于篇幅限制, Fama-French 三因子、Carhart 四因子、Fama-French 五因子以及 CH-4 因子模型的详细构建方法及其数据处理流程有需要者可向作者索要.

⑤ Pelger^[38] 基于大量模拟分析的研究表明, 截断参数的合理取值区间为 [0.05, 0.2]. 在本文中, 这一参数取值设定为 0.1. 后续研究结果表明, 本文中的这一参数取值设定较为合理性且具备一定的鲁棒性.

⑥ 高维高频环境下的渐进分布性质及理论证明过程详见 Pelger^[38].

选取上本文选定了更具市场代表性的中证全指为代表^⑦. 综合数据可得性^⑧, 样本区间设定为 2012 年 1 月 1 日—2020 年 12 月 31 日, 样本期内共计 2 188 个交易日. 剔除因长期停牌而存在明显数据缺失的个别样本^⑨, 最终可获取 723 只有效个股. 此外, 在高频观测间隔的设定上, 参照 Ait-Sahalia 和 Jacod^[48] 及 Pelger^[5] 的研究, 在兼顾最小化微观结构噪声以及最大化可用样本数量的前提下, 设定以 5 min 日内交易频度为代表. 日内交易的观测区间均为每日上午 9:30 ~ 上午 11:30 及下午 1:00 ~ 下午 3:00. 样本期内共计约有 7 603.7 万条日内高频交易数据及约 318.6 万个个股特征及低频交易数据. 无风险利率均采用中国人民银行公布的整存整取一年定期存款利率. 所有原始数据均来源于天软数据库. 在具体的数据处理上, 为保证交易连贯性, 本文对所有价格序列均以 2020 年 12 月 31 日

为基准进行前复权调整^⑩. 此外, 不同于以往众多研究, 在对各因子收益的计算中, 本文均进行了更贴合于现实交易规则的一字涨停及停牌调整. 即设定在所有调仓日中发生一字涨停或被停牌的股票均无法买入, 其投资组合收益将递延自股票复牌或非一字涨停日开始计算, 以防止因子收益的高估^⑪.

2.2 参数 γ 选取及系统性因子个数确认

参照 Lettau 和 Pelger^[39] 的研究, 在对参数 γ 的估计中, 本文通过对于大量的不同参数组合及其定价过程的模拟从而给出了其参数选取的具体依据. 如图 1 所示, 其展示了针对于 2012 年 220 组不同参数组合下日内高频交易部分的股票风险收益定价过程^⑫. 图 1 中每一个矩阵块均代表了在 SDF 的一般资产定价框架下, 基于给定的因子数 k 及参数 γ , 各因子组合在 5 min 高频交易窗口中所能捕获的最大夏普比率.

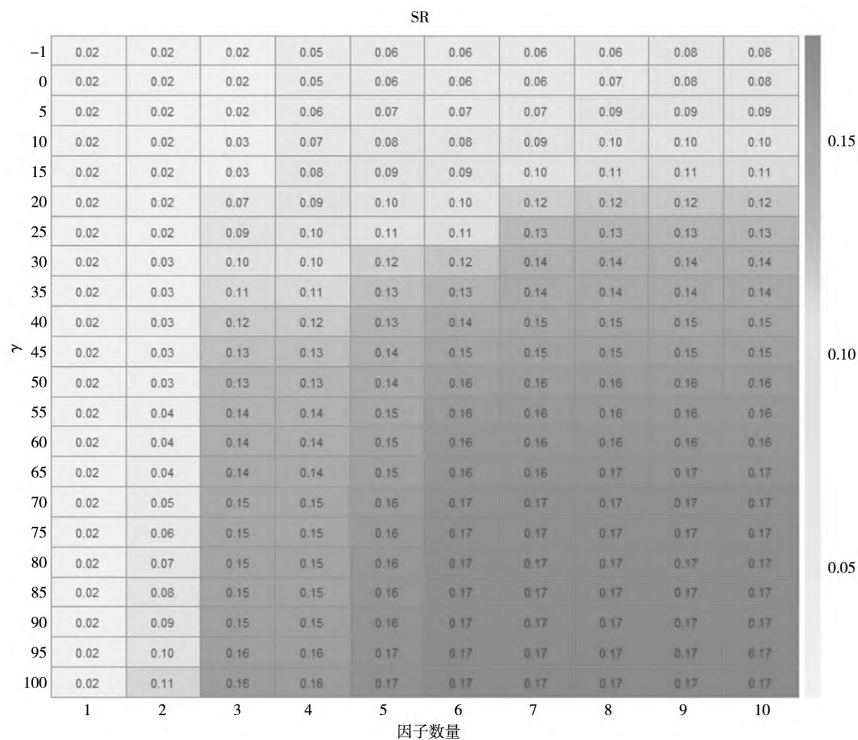


图 1 2012 年高频 RP-PCA 因子组合 SR 热力图

Fig. 1 SR heat map of the combination of HF RP-PCA factors in 2012

⑦ 中证全指: 可系统概括沪深市场上市公司的整体表现. 在本文的研究中, 已剔除 ST、* ST 等不合指数编制条件的股票样本.
 ⑧ 中证全指的高频交易数据自 2011 年 8 月 2 日后可获取, 2012 年是可获取其完整高频交易数据的首个年度.
 ⑨ 为避免因数据缺失过多而带来的小样本影响, 本文剔除了在全年中连续停牌超过 25 个交易日 (约占全年交易日的 1/10) 的个股.
 ⑩ 依据交易所实际发布的数据进行除权, 均采用比例复权方式.
 ⑪ 据统计, 在本文所涉及的 78 084 个调仓节点中, 发生一字涨停或停牌的节点共计 7 214 个, 占比高达 9.24%. 事实证明, 忽略这一因素将会致使各因子收益的明显高估, 因此这一调整是关键且必要的.
 ⑫ 限于文章篇幅, 其他年份、部分的风险收益定价模拟结果 (共计 5 720 组) 有需要者可向作者索要.

具体来看,当参数 γ 固定时,随着因子数 k 的不断增多,夏普比率普遍呈现出逐步增大的趋势.即随着越来越多的风险特征被不断纳入,模型的定价能力也随之呈现出了同步增强的趋势.而当因子数 k 固定时,其同样也表现出了类似的特征.但是,值得注意的是,无论是从纵向还是横向来看,这一定价能力的增长都并非是线性可持续的.即通过观察可以发现,在参数 k 任意给定的情况下, $\gamma = 70$ 时模型定价能力的增长开始逐步趋于停滞,即其定价能力的绝大部分增量均已被现有因子组合所捕获^⑬.这一结论在对其他年份的研究中以及在对日频和隔夜交易的研究中依然适用.因此,在本文的后续研究中,均设定测度参数 $\gamma = 70$.

进一步的,在因子数量的判断上,参照式(8)的评判标准,本文对于系统谱及非系统谱的不同成分间进行了逐步分离.具体而言,首先,本文基于RP-PCA对样本期内每一年的全部高频交易数据的二次协变差矩阵进行单独估计,并相对应的记录其在不同因子个数假设下的扰动特征值比率的详细变化情况.其次,通过对扰动特征值比率聚类现象的观察,从而给出各年份中系统性因子个

数的实际判断依据.如图2所示,其展示了在不同年份中扰动特征值比率的具体变化情况.以2012年为例,可以发现,扰动特征值比率在 $k = 3$ 时首次实现了对于临界值的突破并成功进入到系统谱中.即在近因结构假设下,2012年中共存在着3个系统性因子.相对应的,可以发现,在2013年中同样存在着3个系统性因子.而在2014年—2016年以及2018年—2019年中则各具有4个系统性因子.在2017年及2020年中各具有5个系统性因子.可以发现,各个年份中的系统性因子数量并不固定,存在着明显的时变特征.这一发现明显不同于以Fama-French多因子模型为代表的一系列具有固定因子数量假设的资产定价模型.同时这一发现也相对应的印证了Giannone等^[25]有关于以往因子模型中所存在的稀疏性设定偏误以及模型不确定性等问题的研究结论.此外,结合图1可以发现,在参数 γ 任意给定的前提下,3个~5个系统性因子的这一设定已然捕获了绝大部分的定价增量^⑭.即进一步证明了本文对于截断参数以及因子个数结果判断的合理性.

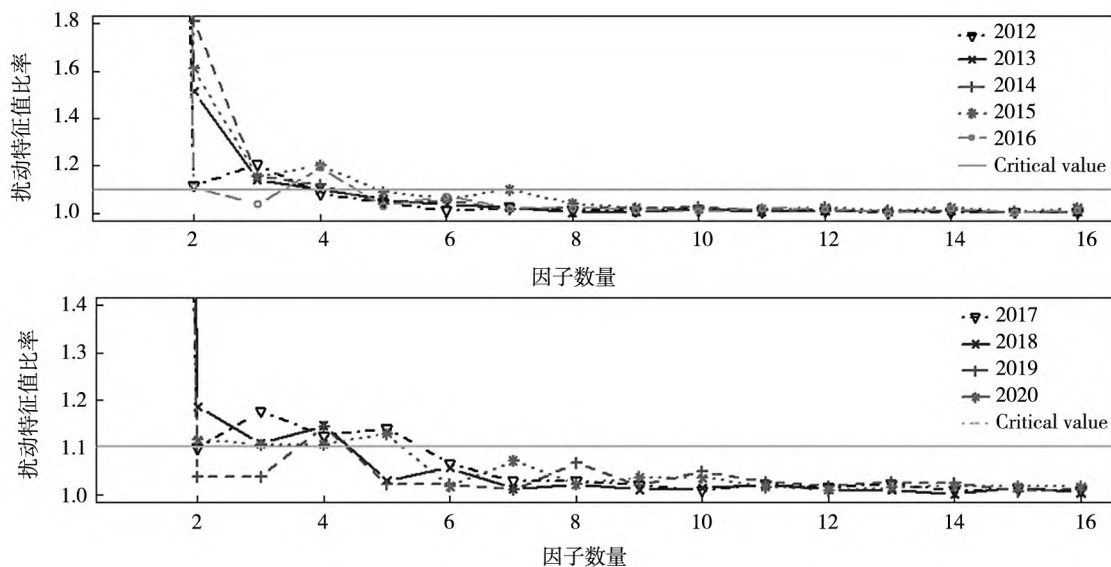


图2 扰动特征值比率变化情况

Fig. 2 Changes in the ratio of perturbed eigenvalues

⑬ 需要特别说明的是,在 $k = 1$ 时,定价能力的增长似乎仍还具有一定的提升空间.但是本文的后续研究即将表明,不存在仅有两个系统性因子的情况.

⑭ 在其他年份中这一结论依然成立.限于篇幅,相关结果有需要者可向作者索要.

2.3 模型定价能力比较

在 APT 假设下, SDF 将完全由系统性因子的投资组合所构成. 因此, 对各因子模型间的定价能力对比即可进一步转化为对其投资组

合有效边界中最优因子组合夏普比率的分析过程. 如表 1 所示, 其展示了在均值方差有效边界上各相切因子组合所能获取的最大年化夏普比率.

表 1 各因子模型在不同交易频度下的资产定价能力比较

Table 1 Comparison of asset pricing power of different factor models at various transaction frequencies

	2012	2013	2014	2015	2016	2017	2018	2019	2020
k	3	3	4	4	4	5	4	4	5
日内 (Intraday)									
RP-PCA	16.20	10.68	15.18	10.82	14.07	15.15	15.12	9.73	10.80
CH-4	2.51	4.36	5.05	3.23	3.03	5.37	1.39	2.35	3.04
Fama-3	2.30	5.00	6.90	4.51	5.04	5.17	1.56	2.46	2.92
Fama-4	5.15	5.68	6.91	4.77	5.12	6.04	2.82	2.64	3.19
Fama-5	2.32	5.60	6.84	4.29	5.25	5.01	3.44	3.00	2.95
隔夜 (Overnight)									
RP-PCA	1.93	1.76	1.87	0.87	1.79	1.37	1.35	1.21	1.76
CH-4	0.50	0.51	0.55	0.24	0.33	0.92	0.45	0.32	0.14
Fama-3	0.52	0.48	0.41	0.15	0.42	0.63	0.35	0.26	0.18
Fama-4	0.53	0.50	0.53	0.16	0.45	0.65	0.35	0.27	0.21
Fama-5	0.55	0.50	0.47	0.42	0.50	0.72	0.47	0.29	0.19
日间 (Daily)									
RP-PCA	0.53	0.52	0.56	0.49	0.52	0.46	0.54	0.57	0.48
CH-4	0.11	0.15	0.20	0.23	0.13	0.19	0.15	0.14	0.14
Fama-3	0.12	0.17	0.30	0.18	0.17	0.20	0.14	0.11	0.10
Fama-4	0.19	0.22	0.30	0.30	0.17	0.21	0.15	0.12	0.14
Fama-5	0.14	0.27	0.34	0.21	0.27	0.31	0.15	0.19	0.18

首先, 从风险补偿结构来看, 无论是对于统计因子模型还是经济因子模型, 其风险补偿均主要源于日内交易. 即相对于日间交易^⑤及隔夜交易, 日内高频交易信息中呈现出了更多的有关于系统性风险的补偿定价信息. 这一发现也与 Pelger^[5]及 Ait-Sahalia 等^[40]有关于美股的研究结论较为相似. 即对于高频交易数据的深入研究或将更有利于改善对于高维投资组合协变结构的估计, 从而构建出更为有效的定价因子. 其次, 如图 3 所示, 通过对补偿结构的进一步拆解可以发现, 各因子模型虽均在日内交易中获取了较高的风险补偿, 但是这种风险补偿经隔夜后却普遍呈现出了

明显下降的趋势^⑥. 并且这一规律恒定的存在于各年份之中且不受定价因子种类影响. 针对于 A 股这一特殊现象的原因解释, 其则可归结于 $t+1$ 交易制度下所带来的次日卖出权利成本^[53]. 隔夜负收益效应的持续存在致使了自日内至日间风险补偿的逐步下降现象. 最后, 无论是对于日内、日间还是隔夜交易来说, 相较于以往的各类经济因子模型, 基于 RP-PCA 的统计因子组合均具有更好的定价表现. 特别地, 其定价过程最多仅依赖于 RP-PCA 前五个隐性因子的投资组合, 即其在摆脱特设稀疏性假设的前提下依然实现了对于高维协变关系的有效降维.

⑤ 日间交易 (Daily) = 日内交易 (Intraday) + 隔夜交易 (Overnight).

⑥ 限于篇幅, 这里仅以 2012 年 RP-PCA 因子的累计收益情况进行举例. 其他模型或其他年份中的收益分布规律均具有相似特征, 相应测算结果有需要者可向作者索要.

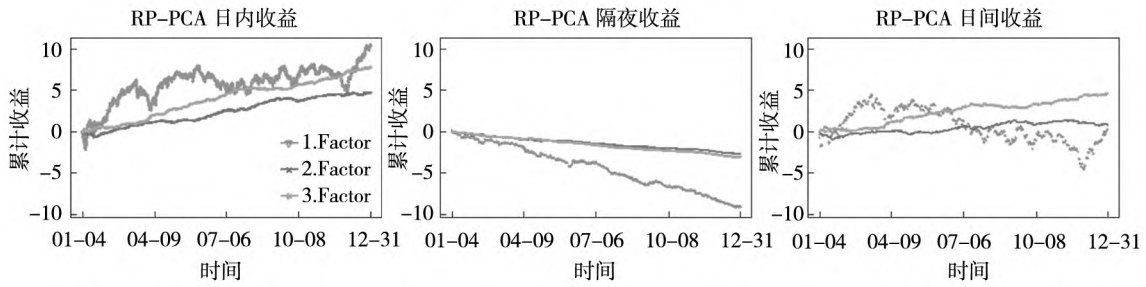


图 3 2012 年 RP-PCA 三因子累计收益分布情况

Fig. 3 Cumulative return distribution of the three RP-PCA factors in 2012

综上所述,针对 RP-PCA 的高频因子研究将更有助于进一步深入理解 A 股市场的系统性风险.基于此,本文将重点围绕高频 RP-PCA 前五个隐性因子的风险定价结构展开研究,并对其因子空间、载荷结构及其潜在的定价机理等进行多方面详细探讨.

3 理解 A 股市场系统性风险

3.1 高频跳跃成分识别

考虑到在以往研究中有关于高频连续风险

$$TOD_j^{(i)} = \frac{n \sum_{t=1}^T |\Delta_{j_t}^n p|^2 l(|\Delta_{j_t}^n p^{(i)}| \leq \tau \sqrt{BV_t^{(i)} \wedge RV_t^{(i)}} n^{-\varpi})}{\sum_{s=1}^{nT} |\Delta_s p|^2 l(|\Delta_s p^{(i)}| \leq \tau \sqrt{BV_{[s/n]}^{(i)} \wedge RV_{[s/n]}^{(i)}} n^{-\varpi})} \quad (9)$$

其中 $j_t = (t-1)n + j$, n 为高频交易频数, l 为指示函数. τ 及 ϖ 为阈值参数并有 $\tau > 0, \varpi \in (0, 0.5)$. RV_t 及 BV_t 分别刻画了股票高频波动的已实现波动率以及双幂次变差(bipower variation, BV),其具体定义分别如下

$$RV_t^{(i)} = \sum_{j=tn+1}^{tn+n} |\Delta_j p^{(i)}|^2, \quad (10)$$

$$BV_t^{(i)} = \frac{\pi}{2} \sum_{j=tn+2}^{tn+n} |\Delta_j p^{(i)}| |\Delta_{j-1} p^{(i)}|$$

在 $n \rightarrow \infty$ 时,基于如式(3)所示的 Itô 半鞅假设, RV_t 及 BV_t 将分别收敛于平方变差 QV_t 和积分方差 IV_t [56-58], 即

$$RV_t^{(i)} \xrightarrow{P} \int_t^{t+1} (\sigma_s^{(i)})^2 ds + \int_t^{t+1} \int_R x^2 v^{(i)}(ds, dx),$$

$$BV_t^{(i)} \xrightarrow{P} \int_t^{t+1} (\sigma_s^{(i)})^2 ds \quad (11)$$

由上式可知,利用 RV_t 及 BV_t 间的取值差异即可剥离高频收益的跳跃波动扩散过程.其中,作

及跳跃风险的因子溢价争议[5, 44, 45], 本文首先对 A 股市场的高频收益结构进行详细探讨,从而更为全面的展示 A 股市场的风险定价结构特征.具体来看,在对不同收益成分的划分中,考虑到在高频交易结构中的日内效应影响,参照 Bollerslev 等[54] 及 Todorov 和 Bollerslev[55] 的阈值截断方法,本文对各原始交易数据均进行了 TOD (time-of-day) 调整,以消除在跳跃成分识别过程中日内各时点的周期性交易特征影响

为波动率的一致估计结果, BV_t 需要依赖于相邻高频时间窗口内非同时发生价格跳跃的强假设.但鉴于在有限样本下这一假设可能带来的跳跃性波动低估现象,参照 Mancini[59] 的研究,本文利用 CV_t 统计量对其进行了进一步的修正,即

$$CV_t^{(i)} = \sum_{j=tn+1}^{tn+n} |\Delta_j p^{(i)}|^2 l(|\Delta_j p^{(i)}| \leq \alpha_j^{(i)} n^{-\varpi}) \quad (12)$$

$\alpha_j^{(i)} = \tau \sqrt{(BV_{[j/n]}^{(i)} \wedge RV_{[j/n]}^{(i)}) \times TOD_{j-[j/n]n}^{(i)}}$, $j = 1, \dots, nT$. 在 ϖ 和 τ 均给定的前提下, $\alpha_j^{(i)} n^{-\varpi}$ 即构成了划分连续成分及跳跃成分的临界阈值.在具体的参数选取上,依托 Bollerslev 等[54] 的大量数值模拟估计结果,本文最终设定 $\varpi = 0.49$ 以保证跳跃成分及连续成分的有效分离.而在参数 τ 的设定上,本文给出了多组不同取值以系统展示其在不同阈值标准下的跳跃波动扩散过程.在此基础上,日内高频跳跃波动可被具体定义为

$$JV_t^{(i)} = RV_t^{(i)} - CV_t^{(i)} \xrightarrow{P} \int_t^{t+1} \int_R x^2 v^{(i)}(ds, dx) \quad (13)$$

其中 $JV_t^{(i)}/RV_t^{(i)}$ 描述了股票的跳跃性波动对其整体波动变化的贡献比例。

表2 跳跃检验结果

Table 2 Jump test results

	2012	2013	2014	2015	2016	2017	2018	2019	2020
	跳跃次数占比 (%)								
$\tau = 3.5$	0.70	0.65	0.70	0.72	0.63	0.63	0.58	0.55	0.55
$\tau = 4.0$	0.37	0.34	0.38	0.40	0.34	0.34	0.31	0.29	0.30
$\tau = 4.5$	0.21	0.19	0.22	0.23	0.19	0.20	0.18	0.17	0.18
$\tau = 5.0$	0.12	0.11	0.13	0.14	0.11	0.12	0.11	0.10	0.11
	总二次变差中的跳跃平均贡献 (%)								
$\tau = 3.5$	14.10	12.85	13.42	12.76	13.61	15.10	13.86	13.72	12.55
$\tau = 4.0$	9.76	8.68	9.22	8.33	9.42	11.01	10.17	10.11	8.98
$\tau = 4.5$	6.88	5.98	6.52	5.62	6.68	8.25	7.71	7.74	6.76
$\tau = 5.0$	5.00	4.22	4.69	3.85	4.88	6.41	6.06	6.03	5.30

表2的估计结果显示,在日内高频收益结构的划分上,样本区间中有超过99%的收益均为连续成分.相比之下,跳跃成分的占比不足1%.并且随着阈值参数的不断增大,跳跃成分的占比进一步缩小.结合波动特征来看,以2012年为例,在阈值参数 $\tau = 3.5$ 时,99.3%的收益均为高频连续收益成分,但是其却仅能够解释同年总二次变差的85.9%.相比之下,占比仅0.7%的高频跳跃成分却在全年中有着14.1%的波动平均贡献.即不同收益成分间存在着明显的风险特征差异,对于不同成分间定价特征的进一步对比是关键且必要的.参照Pelger^[5, 43]的划分标准^①,在以下研究中本文均采用 $\tau = 3.5$ 的阈值参数.

3.2 高频因子分析

在对各系统性因子的研究中,表3的上半部分展示了基于全样本区间的高频RP-PCA因子集合与以往特征因子集合所覆盖(span)的向量空间差异.结果表明,以往经济因子的不同组合均无法捕获由RP-PCA因子组合所构造的向量空间.以Fama-3因子为例, $\{0.99, 0.88, 0.56\}$ 的估计结果显示其最多可以捕获高频RP-PCA中的两个统计因子.即使在进一步添加动量、投资、盈利及换手率等不同风险特征后,这一估计结果也未能得到明显改善.即除市场风险外,RP-PCA所刻画各类系统性风险特征均难以在现有因子中得以完整解释.这与Pelger^[5]针对于美股的研究结论也较为相似.

表3 各因子集合向量空间的广义相关性比较

Table 3 Comparison of generalized correlation of vector space of each factor set

	1. GC	2. GC	3. GC	4. GC	5. GC
高频 RP-PCA 五因子					
Fama-3	0.99	0.88	0.56	0.00	0.00
CH-4	0.99	0.83	0.34	0.15	0.00
Fama-4	0.99	0.89	0.59	0.18	0.00
Fama-5	0.99	0.88	0.58	0.11	0.03
Market	0.99	0.00	0.00	0.00	0.00
高频 PR-PCA 五因子投资组合权重					
ω Continuous	1.00	1.00	0.99	0.98	0.60
ω Jump	0.99	0.75	0.60	0.29	0.10
ω Overnight	0.99	0.86	0.51	0.16	0.06
ω Daily	0.99	0.97	0.87	0.84	0.77
ω Week	0.99	0.94	0.87	0.73	0.27
ω Month	0.99	0.91	0.69	0.34	0.11

① 以往大量研究表明,针对于5 min交易频度的高频收益结构,跳跃成分平均贡献的合理区间范围为10%~20%.

表3的下半部分展示了针对于不同收益成分及交易频度下各RP-PCA因子投资组合权重结构的差异情况。一方面,针对不同的收益成分,测算结果表明,高频RP-PCA因子与高频连续收益RP-PCA因子的投资组合权重结构间存在着较高的相似性,即二者间最多可共享四组因子的权重空间。而对于高频跳跃收益成分而言,其与高频RP-PCA因子间却仅具有以市场风险为代表的一组公共权重空间。即从权重结构来看,连续成分及

跳跃成分二者均具有明显的风险溢价特征。不同于仅需补偿高频跳跃风险的研究结论^[44, 45],本文的测算结果表明,连续风险同样是A股系统性风险结构中不可或缺的关键构成成分。另一方面,从不同的观测频度来看,相较于高频RP-PCA因子组合,在日度、周度以及月度的交易中其各自依次缺失了1个~3个具有共同投资组合权重结构的因子。即随着样本观测频度的逐渐下降,风险补偿的定价信息也在同步减少。

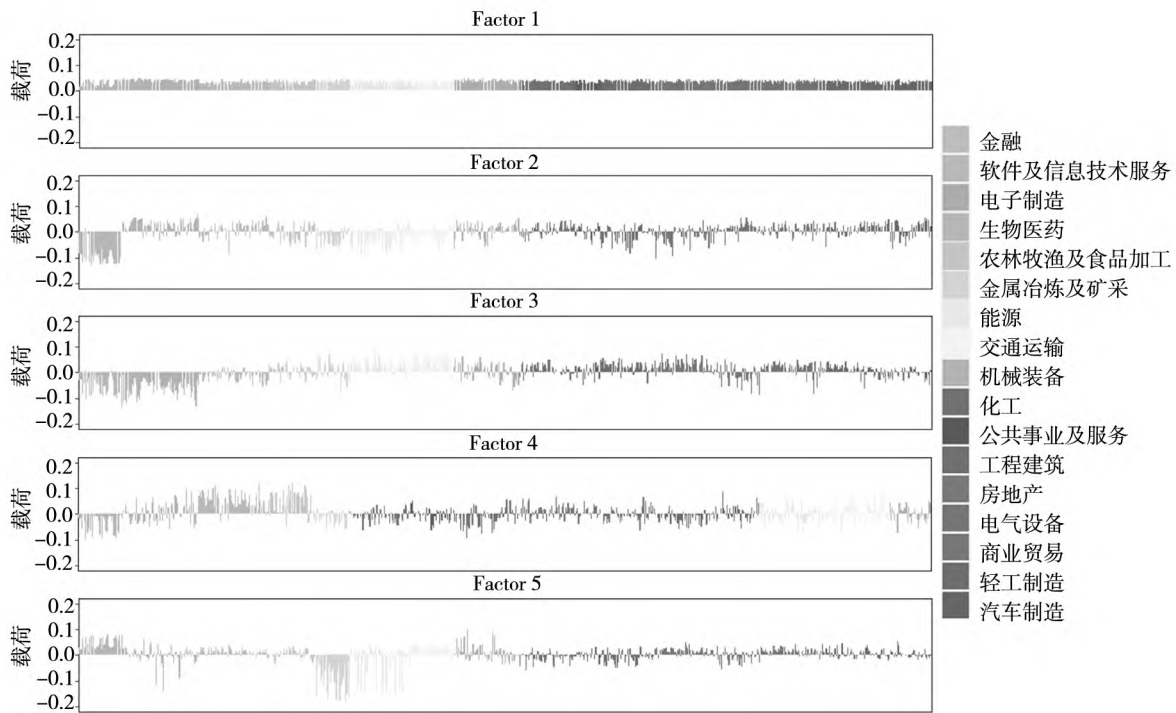


图4 高频RP-PCA五因子的投资组合权重分布情况

Fig. 4 Weight distribution of the high-frequency RP-PCA five-factor portfolio

为了更加清晰的展示在不同观测频度下各系统性因子的权重结构差异及其风险收益定价过程的内在机理,本文基于行业分类特征对各个RP-PCA因子进行了全面拆解。具体来看,在综合参照申银万国以及证监会行业分类标准的前提下,本文将723支样本股票划分为了17个细分行业^⑩。如图4所示,其展示了高频RP-PCA五因子投资组合权重结构在各细分行业的依次分布情况。可以看到,第一个因子是一个纯多头的市场投资组合。整体上近似于一个等权重分布的市场因子,并且这一结构特征在表3中同样得到了印证。而第二个因子则是一个主要聚焦于金融行业的空头

投资组合。相比之下,其在其他行业中的权重占比则较为分散且平均载荷相对较小,并未产生明显的风险暴露。类似的,第三个因子是一个由金融、软件和信息技术服务以及电子制造业所构成的空头投资组合。而第四个因子则是一个由金融、生物医药和农林牧渔及食品加工业所组成的多空投资组合。而第五个因子的权重结构则主要集中于金属冶炼及矿采行业。即整体来看,A股市场的系统性风险均可由以上各细分行业的投资组合所近似表示,并且这一结论在对高频连续成分以及跳跃成分的分析中依然稳健^⑩。此外,对比月度RP-PCA五因子的投资组合权

⑩ 有需要者可向作者索要。

重分布情况可以发现^⑩, 依托于低频观测样本所提取的因子权重分布结构更为离散并且其行业分类特征及因子含义也更加模糊. 即再次印证了自高频至低频所致使的风险补偿定价信息缺失现象. 总体来说, 以上结果不仅揭示了各高频因子的潜在经济含义并且其还进一步的印证了 RP-PCA 在不同交易频度下的定价结构差异情况.

3.3 载荷结构分析

考虑到因子风险敞口的不确定性^[5, 35-37], 在

对因子载荷结构的分析中本文采取了针对于局部时间区间的滚动窗口估计方法, 以期更为全面地考察其载荷结构的动态演变过程. 具体来看, 在保持各高频因子投资组合权重结构不变的前提下, 本文对比了基于全样本时间区间以及基于局部时间区间中所获取的因子载荷结构差异情况. 如图 5 所示, 其展示了在 21 个连续交易日的月度时间窗口中各因子载荷结构空间跨度的动态变化^⑩

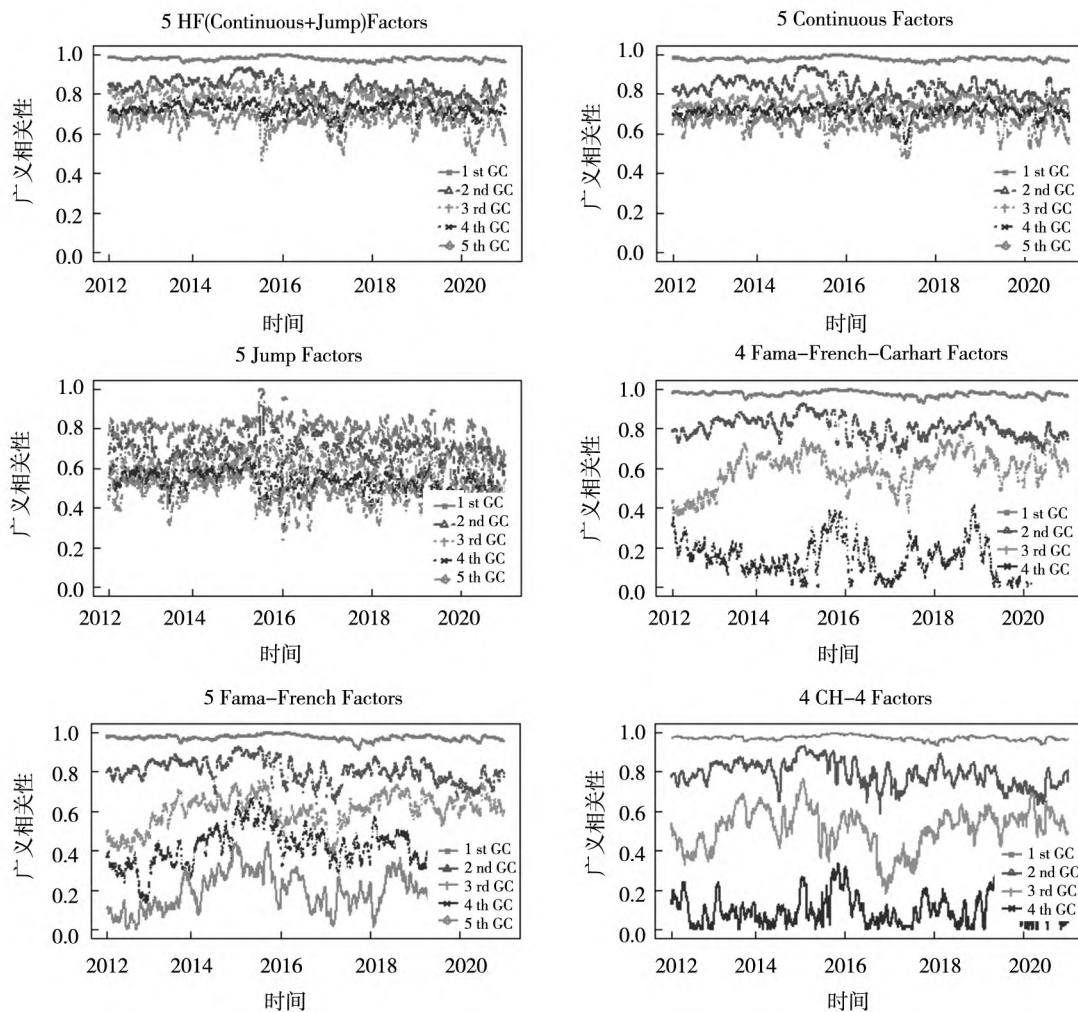


图 5 各因子模型载荷结构空间跨度的动态演变过程

Fig. 5 Dynamic evolution process of spatial span of loading structures with different factor models

首先, 无论是对于 RP-PCA 高维高频统计因子模型还是以 CH-4 及 Fama-French 为代表的经济因子定价模型, 其载荷结构均具有明显的时变特征. 但从载荷结构的时空跨度来看, 二者间却存

在着较大差别. 具体来看, 相较于以往的因子模型, 基于 RP-PCA 的高频因子组合拥有更为稳健的载荷结构. 即在不同的时间区间中, 各投资标的对其因子的投影空间始终保持着高度相似的特

⑩ 在季度以及年度的不同时间窗口中, 研究结论依然稳健. 限于篇幅, 相关结果可向作者索要.

征,截面关系更为稳定. 相比之下,CH-4 以及 Fama-French 多因子组合则具有较大的载荷空间跨度. 也就是说,其在全样本时间区间以及局部时间区间中并不具备统一的因子结构. 因此在面临不同的时间区间时,其因子的投资组合权重结构需实时进行调整以保证其风险敞口估计的无偏性. 但是,鉴于部分财务指标的更新频率,显然这一结构的短期调整将会明显受限. 其次,进一步分析不同风险收益成分的载荷结构可以发现,高频 RP-PCA 因子组合的稳定敞口主要来源于高频连续

成分而非跳跃成分. 相比之下,仅部分跳跃因子中存在着较为一致的载荷结构特征,这一发现也与表 3 的结果相互契合. 此外,需要特别说明的是,因子载荷结构的稳健性与因子数量的时变特征之间并不矛盾. 其原因在于,在特定的时间区间中,局部相对过小的风险敞口或因因子波动率均可致使个别因子暂时失效而无法被有效识别. 因此,只有在因子数量变动的关键原因是源于其载荷结构的大幅调整时,二者间才必须保持着相似或相同的特征结构.

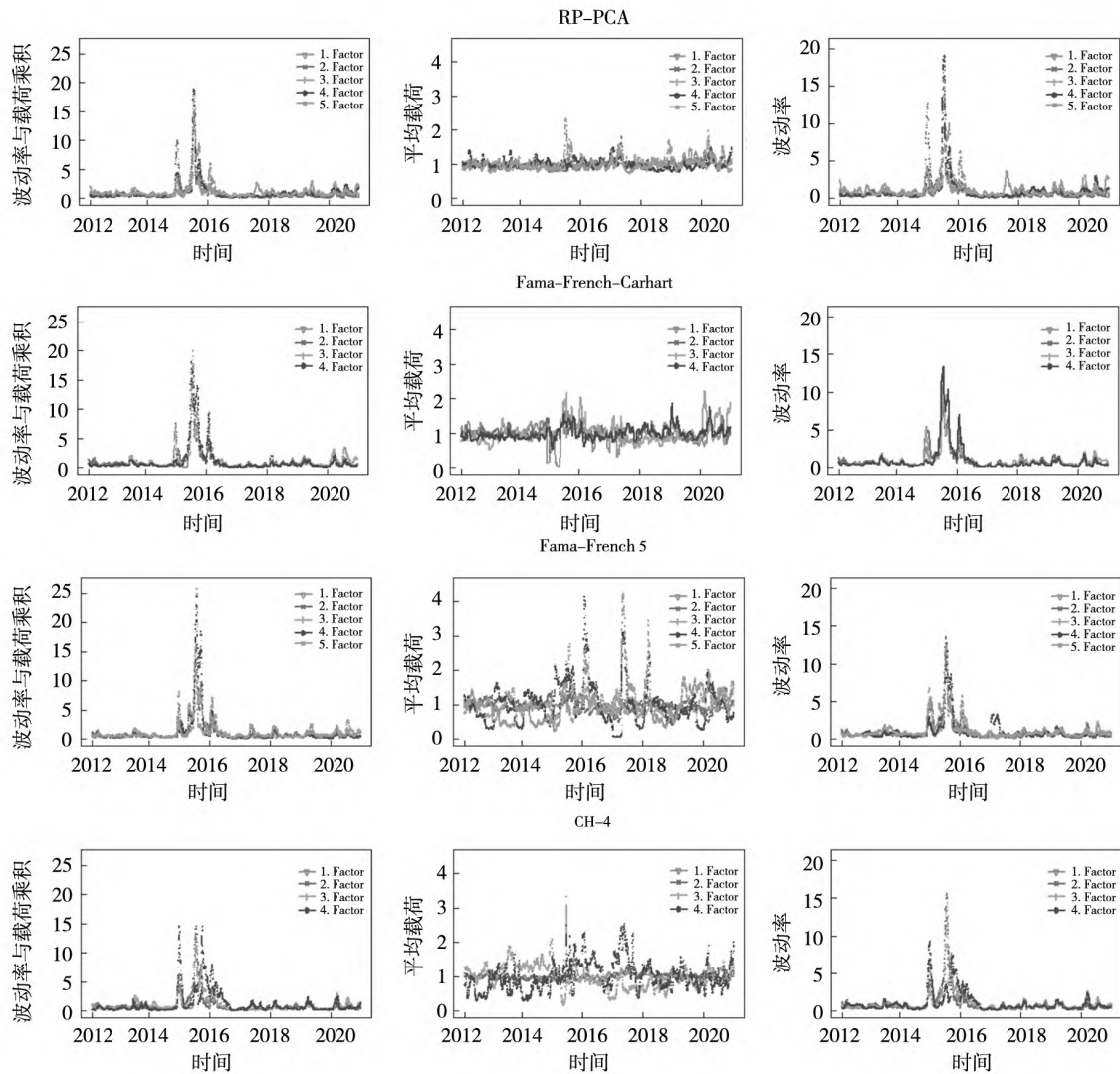


图 6 各因子集合的时变特征结构分解

Fig. 6 Time-varying eigenstructure decomposition of each factor set

为了更为清晰的解释上述问题,图 6 中进一步展示了各高频因子载荷 $\Lambda_k(t)$ 及其波动率 $\sigma_k^2(t)$ 的变化过程. 其中,左图通过构造

$\frac{\Lambda_k(t)^T \Lambda_k(t)}{N} \sigma_k^2(t)$ 展示了 $\Lambda_k(t)$ 与 $\sigma_k^2(t)$ 的乘积变化,其描绘了因子识别过程中第 k 个因子所

对应的特征值波动变化情况. 具体测算上,同以上研究, $\Lambda_k(t)$ 及 $\sigma_k^2(t)$ 均采用基于全样本区间所确定的因子投资组合权重结构并且在月度时间窗口中进行滚动估计. 其中 $\sigma_k^2(t)$ 为第 k 个因子在第 t 个月度时间窗口中的二次变差. 此外,为保持各因子间的可比性并进一步凸显 $\Lambda_k(t)$ 及 $\sigma_k^2(t)$ 在因子识别过程中的差异性贡献,在实际测算中本文对 $\Lambda_k(t)$ 和 $\sigma_k^2(t)$ 均进行了基于均值的归一

$$\text{Variation Explained} = \frac{\text{tr} \left[\sqrt{\text{diag} (R_t^T R_t)^{-1}} \widehat{R}_{t,F(t)}^T \widehat{R}_{t,F(t)} \sqrt{\text{diag} (R_t^T R_t)^{-1}} \right]}{\text{tr} \left[\sqrt{\text{diag} (R_t^T R_t)^{-1}} R_t^T R_t \sqrt{\text{diag} (R_t^T R_t)^{-1}} \right]} \quad (14)$$

基于以上测算结果,图7进一步展示了在时变载荷结构下各因子组合对于股票风险收益的解释能力及差异情况. 如式(14)所示,其测度了在局部样本时间区间中因子组合所捕获的风险收益占全

化处理. 结果显示,在不同时期中呈现出不同数量系统性因子的关键原因在于各因子自身预期收益的大幅波动,而非其载荷结构的大幅调整. 因此,即使在各个年份中检测出了不同数量的系统性因子,其载荷结构也依然可以保持稳定. 特别的,从各因子载荷的实际变化来看,以上结果再次印证了对于 RP-PCA 高频因子组合所具有的稳定风险敞口的判断.

部风险收益的比重情况. 其中, $\widehat{R}_{t,F(t)} = F \widehat{\Lambda}_t^T = R_t \widehat{\Lambda}_t \times (\widehat{\Lambda}_t^T \widehat{\Lambda}_t)^{-1} \widehat{\Lambda}_t^T$, $t = 1, \dots, 2168$. R_t 和 $\widehat{R}_{t,F(t)}$ 分别表示在第 t 个局部时间窗口中的原始收益矩阵以及时变载荷 $\widehat{\Lambda}_t$ 下各因子模型所刻画的局部收益矩阵.

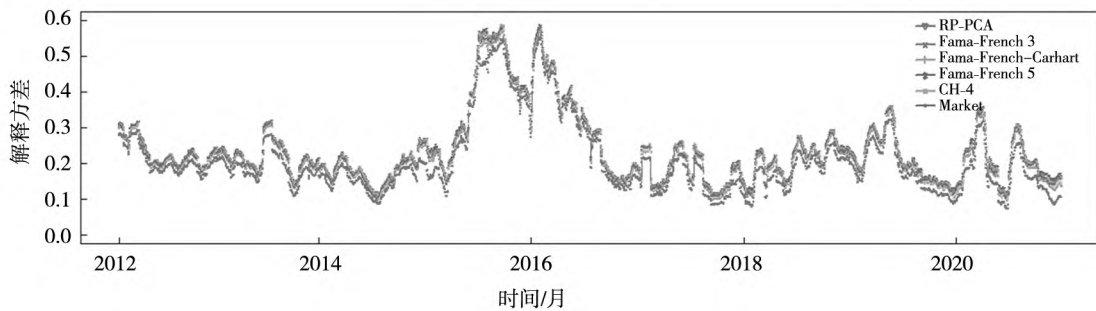


图7 各因子模型所捕获的风险收益占比情况及其时变过程

Fig. 7 Risk-return shares captured by different factor models and their time-varying processes

估计结果显示,相比之下,基于 RP-PCA 的高频因子组合拥有着更强的风险收益解释能力. 其中,大部分的系统性风险均已在市场因子中得以解释. 各因子模型的测算结果表明,在不同的时间区间中平均拥有着 10% ~ 30% 的系统性风险. 特别的,在 2015 年—2016 年间的“杠杆牛市”中,这一比例得到了显著提升且呈现出快速演变的趋势. 总体来说,本节的研究结果表明,相较于以往的经济因子模型,RP-PCA 提供了一组更为稳健的系统性风险因子组合,且无论是在载荷结构的稳定性还是对风险收益的整体解释能力上,其均拥有着更好的表现.

3.4 风险补偿结构分析

以往研究表明,相比于资产价格的上涨风险,投资者或将更加关注于价格下跌风险,因此系统性风险的补偿结构往往并非完全对称^[46, 60]. 考虑到 A 股的单边交易机制以及较高的个人投资者

占比,这一现象或将更加突出^[26]. 基于此,在风险补偿结构的分析中,本文对各个系统性因子的风险敞口进行了详细拆解以探讨其所可能存在的非对称定价结构

$$\hat{\beta}_{t,i,k}^+ = \frac{\sum_{s=1}^n r_{t,s,i} f_{t,s,k}^+}{\sum_{s=1}^n (f_{t,s,k}^+)^2}, \hat{\beta}_{t,i,k}^- = \frac{\sum_{s=1}^n r_{t,s,i} f_{t,s,k}^-}{\sum_{s=1}^n (f_{t,s,k}^-)^2} \quad (15)$$

$$\hat{\beta}_{t,i,k}^P = \frac{\sum_{s=1}^n r_{t,s,i}^+ f_{t,s,k}^+}{\sum_{s=1}^n f_{t,s,k}^+{}^2}, \hat{\beta}_{t,i,k}^{f+} = \frac{\sum_{s=1}^n r_{t,s,i}^- f_{t,s,k}^+}{\sum_{s=1}^n f_{t,s,k}^+{}^2},$$

$$\hat{\beta}_{t,i,k}^N = \frac{\sum_{s=1}^n r_{t,s,i}^- f_{t,s,k}^-}{\sum_{s=1}^n f_{t,s,k}^-{}^2}, \hat{\beta}_{t,i,k}^{f-} = \frac{\sum_{s=1}^n r_{t,s,i}^+ f_{t,s,k}^-}{\sum_{s=1}^n f_{t,s,k}^-{}^2} \quad (16)$$

具体而言,如式(15)及式(16)所示,其分别展示了 Ang 等^[61]及 Bollerslev 等^[46]所提出的两

种不同的系统性风险敞口分解方法. 整体来看, 二者间的关键差别在于是否额外考虑个股风险收益的变化. 其中 $r_{t,s,i}$ 及 $f_{t,s,k}$ 分别表示在第 t 天中的第 s 个 5 min 交易窗口内第 i 支股票的风险收益以及第 k 个因子的收益情况. 其中 $r_{t,s,i}^+ = \max(r_{t,s,i}, 0)$, $r_{t,s,i}^- = \min(r_{t,s,i}, 0)$, $f_{t,s,k}^+$ 及 $f_{t,s,k}^-$ 的计算同理

$$r_{t+1,i} = \lambda_{t+1,k}^P \widehat{\beta}_{t,i,k}^+ + \lambda_{t+1,k}^N \widehat{\beta}_{t,i,k}^- + \varepsilon_{t+1,i} \quad (17)$$

$$r_{t+1,i} = \lambda_{t+1,k}^P \widehat{\beta}_{t,i,k}^P + \lambda_{t+1,k}^{f+} \widehat{\beta}_{t,i,k}^{f+} + \lambda_{t+1,k}^N \widehat{\beta}_{t,i,k}^N + \lambda_{t+1,k}^{f-} \widehat{\beta}_{t,i,k}^{f-} + \varepsilon_{t+1,i} \quad (18)$$

基于以上各因子风险敞口的分解结果, 结合 Fama-MacBeth 估计方法即可确定其各自的风险溢价情况. 其中, 式(17)和式(18)分别对应于两种风险敞口分解方式下的截面回归过程, 并有

$$\widehat{\lambda} = \frac{1}{T-1} \sum_{t=2}^T \widehat{\lambda}_t.$$

表4 各风险敞口的 Fama-MacBeth 估计结果

Table 4 Fama-MacBeth estimation results of each risk exposure

	β^+	β^-	β^P	β^{f+}	β^N	β^{f-}
Factor 1	0.08 (0.23)	-2.20*** (-5.83)	—	—	—	—
	—	—	2.19 (1.44)	12.77*** (7.54)	-8.19*** (-9.33)	0.84 (0.45)
Factor 2	0.23** (2.31)	-0.18* (-1.69)	—	—	—	—
	—	—	0.09 (0.41)	2.31*** (7.75)	-1.81*** (-5.45)	0.25 (1.08)
Factor 3	0.12** (2.10)	-0.23*** (-3.40)	—	—	—	—
	—	—	-0.75*** (-4.53)	1.75*** (8.61)	-1.95*** (-7.41)	-0.36 (-1.48)
Factor 4	0.23*** (4.34)	-0.05 (-0.87)	—	—	—	—
	—	—	-0.24* (-1.66)	1.61*** (8.63)	-1.32*** (-8.05)	-0.25 (-0.77)
Factor 5	0.11** (2.38)	-0.05 (-1.04)	—	—	—	—
	—	—	-0.58*** (-3.96)	1.59*** (9.41)	-1.19*** (-8.57)	0.17 (1.39)

注：括号内为各影响系数(%)经 Newey-West 调整后的 t 值, 最大滞后期为 7; ***, **, * 分别表明在 1%, 5%, 10% 的水平下显著.

表4的估计结果表明, 非对称的系统性风险敞口在A股市场中显著存在, 并且在不同因子间具有一定的相似性特征. 具体而言, 以因子1为例, 基于式(15)的半 β 分解结果显示, 仅 β^- 存在着明显的风险溢价. 即从市场因子的收益变化来看, 投资者或更加关注于A股的市场资产价格下跌风险, 而忽略其价格的上涨风险. 然而, 在进一步考虑了如式(16)所示的个股收益方向变化时, 可以发现市场资产价格的下跌风险并未被完全定价. 其中, 仅 β^N 存在着明显的风险溢价. 类似的,

市场资产价格的上涨风险也并未被完全忽略, 暴露如 β^{f+} 所描述的市场收益上涨而个股收益下跌的风险敞口时依然需要额外的风险补偿. 即从整体来看, A股市场的投资者厌恶于市场和个股同时发生下跌以及市场上涨但个股下跌时的风险, 并且这一现象普遍的存在于所有因子之中. 究其原因, 面临大量的个人投资者, 大量的非理性交易常常致使各类资产价格远高于其自身的内在价值^[62], 从而显著降低了各类资产的远期回报. 但受限于严格的卖空限制以及高昂的卖空成本, 这

种预期未来收益下跌的风险却常常不能被投资者有效规避,因此也就造成了这种非对称系统性风险敞口的显著存在.而对于个别因子中 β^p 的部分依然会被显著定价的这一特殊现象,其则可归因于在A股市场中投资者对于特定行业的短期炒作行为所带来的价值高估现象.但同样是受限于做空限制,投资者无法拥有较好的手段以规避这一风险.即从整体的风险补偿结构来看,A股市场的系统性风险并未被完全定价并且存在着一定的套利空间.但随着未来投资者结构、市场交易机制以及监管体系的不断完善,这一风险敞口的非对称性特征或将逐步收缩直至消失.

4 结束语

金融市场的全面开放加速行业出清的同时也为我国资本市场的体系结构带来了明显冲击.面临跨境资本流动规模的逐步扩大、股票市场结构的加速演变以及可投资标的数量的大幅提升,海量的金融交易数据以及高维的资产特征结构使得以往依赖于特设稀疏性假设的因子定价模型暴露出了诸多关键问题.基于此,本文基于SDF的一般资产定价框架,依托于近因结构下的高维高频RP-PCA因子模型统计推断理论,通过对不同交易频度、不同收益成分以及不同因子模型间的系统对比,给出了A股市场更具一般代表性且具有更强资产定价能力的系统性因子组合.在此基础上,通过对因子组合的成分结构、载荷空间及其风险补偿过程的详细拆解,全面展示了A股市场的系统性风险特征及其演变过程.通过以上分析,以期相关政策及风险管理策略的制定提供重要参照.

研究发现:1)相较于以CH-4以及Fama-French为代表的多因子定价模型,基于高维RP-PCA的因子测度框架拥有着更强的风险收益定价能力.相比之下,其在高频交易环境中可捕获更多的系统性风险特征及资产定价信息,从而获取更为有效的因子组合;2)基于高频交易数据的高维协方差矩阵存在着明显的低秩稀疏结构,并且

这一低秩稀疏结构可由RP-PCA前五个特征因子的投资组合所近似表示;3)从因子成分来看,除市场因子外,各RP-PCA因子所刻画的各种系统性风险特征均难以被现有因子组合所完整解释,但是其却可由以金融行业为代表的六个特定细分行业的投资组合所近似概括,并且这一结构特征随着交易观测频度的逐步提高将愈发清晰;4)不同于以往因子组合,基于高维高频的RP-PCA因子组合拥有着更为稳定的载荷结构,给出了A股市场系统性风险在局部区间以及全局区间中动态演变过程的一致估计结果;5)因子数量时变原因在于因子预期收益的大幅波动而非其载荷结构的明显调整;6)A股市场的系统性风险并未被完全定价,非对称的敞口特征显著存在.投资者或更加关注于系统上涨而个股下跌以及系统和个股同时下跌时的风险.

科学地测度系统性风险及其补偿定价结构是合理制定相应监管调控政策的前提和基础.从上述测算结果来看,在相应调控及监管政策的制定上,相关部门应严格监管对于特定行业的过度炒作,在引导投资者理性预期的同时有选择的适度开放做空机制,提高资产定价效率并有效抑制系统性风险敞口的非对称结构,从而进一步收缩套利空间并降低系统性风险的整体水平.而对于相应投资策略的构建来说,一方面,基于现有的交易制度及市场运行特征,与其管控各系统性因子的全部风险敞口,不如仅交易其中已被定价的部分风险敞口,从而在同等收益下有效降低对于特定风险的暴露水平.而在额外考虑现实交易成本影响的前提下,这一策略或将更具优势.另一方面,对于仅在错误定价部分寻求套利机会的投资者来说,系统对比不同交易频度或不同风险收益成分的因子特征结构是至关重要的.原因在于,仅针对于低频交易或部分风险收益的研究往往并不能够完整捕获全部的系统性风险特征,因此忽略这一关键因素将致使其相应的投资策略不可避免的承受特定风险的潜在影响.这些发现对于金融投资者、政策研究机构以及监管部门等均具有重要的参照意义.

参考文献:

- [1] Markowitz H. Portfolio selection[J]. *Journal of Finance*, 1952, 7(1): 71–91.
- [2] Sharpe W F. Capital asset prices: A theory of market equilibrium under conditions of risk[J]. *Journal of Finance*, 1964, 19(3): 425–442.
- [3] Ross S A. The arbitrage theory of capital asset pricing[J]. *Journal of Economic Theory*, 1976, 13(3): 341–360.
- [4] Lettau M, Pelger M. Factors that fit the time series and cross-section of stock returns[J]. *The Review of Financial Studies*, 2020, 33(5): 2274–2325.
- [5] Pelger M. Understanding systematic risk: A high-frequency approach[J]. *Journal of Finance*, 2020, 75(4): 2179–2220.
- [6] Fama E F, French K R. Common risk factors in the returns on stocks and bonds[J]. *Journal of Financial Economics*, 1993, 33(1): 3–56.
- [7] Carhart M M. On persistence in mutual fund performance[J]. *Journal of Finance*, 1997, 52(1): 57–82.
- [8] Fama E F, French K R. A five-factor asset pricing model[J]. *Journal of Financial Economics*, 2015, 116(1): 1–22.
- [9] 陈国进, 张润泽, 谢沛霖, 等. 知情交易、信息不确定性与股票风险溢价[J]. *管理科学学报*, 2019, 22(4): 53–74.
- Chen Guojin, Zhang Runze, Xie Peilin, et al. Informed trading, information uncertainty and stock risk premium[J]. *Journal of Management Sciences in China*, 2019, 22(4): 53–74. (in Chinese)
- [10] 李双琦, 陈其安, 朱 沙. 考虑消费与投资者情绪的股票市场资产定价[J]. *管理科学学报*, 2021, 24(4): 86–108.
- Li Shuangqi, Chen Qian, Zhu Sha. Asset pricing in the stock market with consumption and investor sentiment[J]. *Journal of Management Sciences in China*, 2021, 24(4): 86–108. (in Chinese)
- [11] Yuan Y, Ling A, Hu Z. Ambiguity, limited market participation, and the cross-sectional stock return[J]. *Journal of Management Science and Engineering*, 2024, 9(4): 522–534.
- [12] 王 鹏, 梁鑫垚. 股票收益率能反映三阶矩信息吗? ——基于我国股票市场的实证研究[J]. *管理科学学报*, 2023, 26(8): 94–116.
- Wang Peng, Liang Xinyao. Can stock returns reflect third-order moment information?: Evidence from China's stock market[J]. *Journal of Management Sciences in China*, 2023, 26(8): 94–116. (in Chinese)
- [13] Liu Y, Yang B, Su Y. Changes in firm profitability, heterogeneous investor beliefs, and stock returns[J]. *Journal of Management Science and Engineering*, 2023, 8(2): 258–272.
- [14] 朱红兵, 张 兵. 价值性投资还是博彩性投机? ——中国 A 股市场的 MAX 异象研究[J]. *金融研究*, 2020, (2): 167–187.
- Zhu Hongbing, Zhang Bing. Investment or gambling?: The MAX anomaly in China's A-share stock market[J]. *Journal of Financial Research*, 2020, (2): 167–187. (in Chinese)
- [15] 何诚颖, 陈 锐, 薛 冰, 等. 投资者情绪, 有限套利与股价异象[J]. *经济研究*, 2021, 56(1): 58–73.
- He Chengying, Chen Rui, Xue Bing, et al. Investor sentiment, limited arbitrage and stock price anomalies[J]. *Economic Research Journal*, 2021, 56(1): 58–73. (in Chinese)
- [16] 王 磊, 孔东民. 应计信息、机构投资者反应与股票错误定价[J]. *管理科学学报*, 2017, 20(3): 80–98.
- Wang Lei, Kong Dongmin. Accrual information, institutional investor reaction and stock mispricing[J]. *Journal of Management Sciences in China*, 2017, 20(3): 80–98. (in Chinese)
- [17] 朱小能, 陈俊坪, 朱 杰. 生产还是消费——中国股市生产资本资产定价模型实证检验[J]. *管理科学学报*, 2017,

- 20(8): 1 – 12.
- Zhu Xiaoneng, Chen Junping, Zhu Jie. Production or consumption: An empirical test of production-based asset pricing model in the Chinese stock market[J]. *Journal of Management Sciences in China*, 2017, 20(8): 1 – 12. (in Chinese)
- [18] Fan J, Furger A, Xiu D. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data[J]. *Journal of Business and Economic Statistics*, 2016, 34(4): 489 – 503.
- [19] Feng G, Giglio S, Xiu D. Taming the factor zoo: A test of new factors[J]. *Journal of Finance*, 2020, 75(3): 1327 – 1370.
- [20] Chordia T, Goyal A, Saretto A. Anomalies and false rejections[J]. *The Review of Financial Studies*, 2020, 33(5): 2134 – 2179.
- [21] Harvey C R, Liu Y. A Census of The Factor Zoo[EB/OL]. Available at SSRN 3341728, 2019, <https://ssrn.com/abstract=3341728>.
- [22] Pukthuanthong K, Roll R, Subrahmanyam A. A protocol for factor identification[J]. *The Review of Financial Studies*, 2019, 32(4): 1573 – 1607.
- [23] Kozak S, Nagel S, Santosh S. Interpreting factor models[J]. *Journal of Finance*, 2018, 73(3): 1183 – 1223.
- [24] Fama E F, French K R. Choosing factors[J]. *Journal of Financial Economics*, 2018, 128(2): 234 – 252.
- [25] Giannone D, Lenza M, Primiceri G E. Economic predictions with big data: The illusion of sparsity[J]. *Econometrica*, 2021, 89(5): 2409 – 2437.
- [26] Wan D, Yang L, Yang X. How mutual funds respond to asymmetric feedback trading in China's stock market[J]. *Journal of Management Science and Engineering*, 2024, 9(2): 143 – 160.
- [27] Bai J. Inferential theory for factor models of large dimensions[J]. *Econometrica*, 2003, 71(1): 135 – 171.
- [28] Fan J, Fan Y, Lv J. High dimensional covariance matrix estimation using a factor model[J]. *Journal of Econometrics*, 2008, 147(1): 186 – 197.
- [29] Nagel S. Empirical cross-sectional asset pricing[J]. *Annual Review of Financial Economics*, 2013, (5): 167 – 199.
- [30] Fan J, Liao Y, Mincheva M. High dimensional covariance matrix estimation in approximate factor models[J]. *Annals of Statistics*, 2011, 39(6): 3320 – 3356.
- [31] Fan J, Liao Y, Mincheva M. Large covariance estimation by thresholding principal orthogonal complements[J]. *Journal of the Royal Statistical Society*, 2013, 75(4): 603 – 680.
- [32] Giglio S, Xiu D. Asset pricing with omitted factors[J]. *Journal of Political Economy*, 2021, 129(7): 1947 – 1990.
- [33] Rapach D, Zhou G. Sparse Macro Factors[EB/OL]. Available at SSRN 3259447, 2021, <https://ssrn.com/abstract=3259447>.
- [34] Kozak S, Nagel S, Santosh S. Shrinking the cross-section[J]. *Journal of Financial Economics*, 2020, 135(2): 271 – 292.
- [35] Kelly B, Pruitt S, Su Y. Instrumented principal component analysis[EB/OL]. Available at SSRN 2983919, 2017, <https://ssrn.com/abstract=2983919>.
- [36] Kelly B, Pruitt S, Su Y. Characteristics are covariances: A unified model of risk and return[J]. *Journal of Financial Economics*, 2019, 134(3): 501 – 524.
- [37] Fan J, Liao Y, Wang W. Projected principal component analysis in factor models[J]. *Annals of Statistics*, 2016, 44(1): 219 – 254.
- [38] Kong X, Lin J, Liu C, et al. Discrepancy between global and local principal component analysis on large-panel high-frequency data[J]. *Journal of the American Statistical Association*, 2021, 118(542): 1333 – 1344.

- [39] Lettau M, Pelger M. Estimating latent asset-pricing factors[J]. *Journal of Econometrics*, 2020, 218(1): 1–31.
- [40] Ait-Sahalia Y, Kalnina I, Xiu D. High-frequency factor models and regressions[J]. *Journal of Econometrics*, 2020, 216(1): 86–105.
- [41] Ait-Sahalia Y, Xiu D. Principal component estimation of a large covariance matrix with high-frequency data[J]. *Journal of Econometrics*, 2017, (201): 384–399.
- [42] Ait-Sahalia Y, Xiu D. Principal component analysis of high frequency data[J]. *Journal of American Statistical Association*, 2019, 114(525): 287–303.
- [43] Pelger M. Large-dimensional factor modeling based on high-frequency observations[J]. *Journal of Econometrics*, 2019, 208(1): 23–42.
- [44] Bollerslev T, Li S Z, Todorov V. Roughing up beta: Continuous vs. discontinuous betas, and the cross section of expected stock returns[J]. *Journal of Financial Economics*, 2016, 120(3): 464–490.
- [45] Alexeev V, Dungey M, Yao W. Time-varying continuous and jump betas: The role of firm characteristics and periods of stress[J]. *Journal of Empirical Finance*, 2017, (40): 1–19.
- [46] Bollerslev T, Patton A J, Quaedvlieg R. Realized semibetas: Disentangling "good" and "bad" downside risks[J]. *Journal of Financial Economics*, 2021, 144(1): 227–246.
- [47] Levi Y, Welch I. Symmetric and asymmetric market betas and downside risk[J]. *The Review of Financial Studies*, 2020, 33(6): 2772–2795.
- [48] Ait-Sahalia Y, Jacod J. *High-Frequency Financial Econometrics*[M]. Princeton: Princeton University Press, 2014.
- [49] Chamberlain G, Rothschild M. Arbitrage, factor structure, and mean-variance analysis on large asset markets[J]. *Econometrica*, 1983, (51): 1281–1304.
- [50] Onatski A. Determining the number of factors from empirical distribution of eigenvalues[J]. *Review of Economics and Statistics*, 2010, 92(4): 1004–1016.
- [51] Ahn S C, Horenstein A R. Eigenvalue ratio test for the number of factors[J]. *Econometrica*, 2013, 81(3): 1203–1227.
- [52] Bai J, Ng S. Evaluating latent and observed factors in macroeconomics and finance[J]. *Journal of Econometrics*, 2006, (131): 507–537.
- [53] 张 兵. 中国资本市场的 T+1 交易制度研究: 隔夜收益率视角[J]. *管理世界*, 2020, 36(12): 26–35.
Zhang Bing. The research of trading mechanism of China's capital market[J]. *Management World*, 2020, 36(12): 26–35. (in Chinese)
- [54] Bollerslev T, Todorov V, Li S Z. Jump tails, extreme dependencies, and the distribution of stock returns[J]. *Journal of Econometrics*, 2013, 172(2): 307–324.
- [55] Todorov V, Bollerslev T. Jumps and betas: A new theoretical framework for disentangling and estimating systematic risks[J]. *Journal of Econometrics*, 2010, 157(2): 220–235.
- [56] Andersen T G, Bollerslev T, Diebold F X, et al. Modeling and forecasting realized volatility[J]. *Econometrica*, 2003, 71(2): 579–625.
- [57] Barndorff-Nielsen O E, Shephard N. Power and bipower variation with stochastic volatility and jumps[J]. *Journal of Financial Econometrics*, 2004, 2(1): 1–37.
- [58] Barndorff-Nielsen O E, Shephard N. Econometrics of testing for jumps in financial economics using bipower variation[J]. *Journal of Financial Econometrics*, 2006, 4(1): 1–30.
- [59] Mancini C. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps[J]. *Scandinavian Journal of Statistics*, 2009, (36): 270–296.
- [60] Bollerslev T, Li J, Patton A J, et al. Realized semicovariances[J]. *Econometrica*, 2020, 88(4): 1515–1551.

- [61] Ang A, Chen J, Xing Y. Downside risk[J]. *The Review of Financial Studies*, 2006, 19(4): 1191 – 1239.
- [62] Liu J, Stambaugh R F, Yuan Y. Size and value in China[J]. *Journal of Financial Economics*, 2019, 134(1): 48 – 69.

A study of systematic risk in the A-share market based on large-dimensional factor models

WEI Zhao-hao^{1,5}, LI Xiu-ting^{1,2}, LU Quan-ying^{3,4}, DONG Ji-chang^{1,2}, DONG Zhi^{1,2*}

1. School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China;
2. Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China;
3. School of Economics and Management, Beijing University of Technology, Beijing 100124, China;
4. Center for Forecasting Science, Chinese Academy of Sciences, Beijing 100190, China;
5. Postdoctoral Research Workstation, Shenwan Hongyuan Securities Co., Ltd., Shanghai 200031, China

Abstract: Facing the challenges of large-dimensional characteristics and the limitations of ad-hoc sparsity assumption, this paper constructs a portfolio of systematic risk factors with stronger asset pricing capability and broader representativeness for the A-share market. Drawing on the implicit factor structure with reference to the cutting-edge research in the field of large-dimensional factor statistical inference, this study conducts a comprehensive deconstruction and analysis of the corresponding factor characteristics and their associated risk compensation structure. The results indicate that: 1) Compared with the CH-4 and Fama-French multi-factor pricing models, the large-dimensional high-frequency RP-PCA factor portfolio captures more systematic risk characteristics and exhibits better pricing performance along with a more robust time-varying loading structure; 2) The A-share market contains five stable systematic risk factors. Apart from the market factor, the other four can be approximated by the portfolios of six specific sub-sectors, predominantly represented by the finance sector; 3) Systematic risk in the A-share market is not fully priced, and there is significant asymmetry in risk exposure. Overall, investors tend to be more concerned about risks when the market rises while individual stocks fall, as well as during simultaneous declines in both the market and individual stocks.

Key words: systematic risk; large-dimensional covariance matrix; high-frequency data; factor model; SDF