

《我国量化私募基金的业绩和系统性尾部风险》附录

附录 1 TML因子构建方法

参考 Agarwal et al. (2017), TML因子的具体构造方法如下: 首先对每个基金—月份样本点构造新指标*TailRiskGroup*, 并在*TailRisk* = 0 的样本点上将其赋值为 0, 可以认为这部分基金的个体尾部风险与市场尾部风险并无溢出关系。然后, 将*TailRisk* ≠ 0 的样本点逐月按升序分为 5 组, 并为其*TailRiskGroup*赋值 1—5。接着, 我们按照上月*TailRiskGroup*取值, 用等权重加权的方式计算每月每组基金的收益率。最后, 用第 5 组(系统性尾部风险最高)收益率减去第 0 组(系统性尾部风险最低)收益率, 得到系统性尾部风险因子TML的时序收益率。

附录 2 alpha 的滚动估计方法

以基金(或基金组合)*i*在*t*月为例, 我们基于其从*t* - 23 到*t*月的收益率, 估计得到其对于各因子的风险敞口, 再用其*t*月(*t* + 1 月)实际收益率减去该风险敞口下的因子组合收益率, 得到时变的样本内(外)alpha。以基准模型为例, 样本内、外估计 alpha 的方法如下:

$$\hat{\alpha}_{i,t,J} = r_{i,t} - r_{f,t} - \sum_{j=1}^J \hat{\beta}_{j,i,t} f_{j,t} \quad (A1)$$

$$\hat{\alpha}_{i,t+1,J} = r_{i,t+1} - r_{f,t+1} - \sum_{j=1}^J \hat{\beta}_{j,i,t} f_{j,t+1} \quad (A2)$$

其中, $J \in \{1,3,4,5,6\}$ 对应上述五个因子模型, $\hat{\alpha}_{i,t,J}$ 为*J*因子模型中基金(或基金组合)*i*在*t*时的样本内 alpha, $\hat{\alpha}_{i,t+1,J}$ 为*J*因子模型中基金(或基金组合)*i*在*t* + 1 时的样本外 alpha, $\hat{\beta}_{j,i,t}$ 为基金(或基金组合)*i*在*t*时按各因子模型在 24 个月滚动窗口中估计出的对第*j*个风险因子的敞口。 $f_{j,t}$ 为*t*时的第*j*个风险因子收益率。例如, $J = 4$ 对应四因子模型, $\hat{\alpha}_{i,t,4}$ 和 $\hat{\alpha}_{i,t+1,4}$ 分别为*t*时样本内四因子 alpha 和*t* + 1 时样本外四因子 alpha。

附录 3 变量定义

文章第五部分的分析还使用了一系列基金特征作为控制变量。参考 Agarwal et al. (2017), 控制变量分为以下三类: (1) 基金年龄 (Age); (2) 基金收益率相关特征: 包括过去 24 个月窗口中滚动估计出的偏度 (Skewness)、峰度 (Kurtosis)、95% VaR (VaR_0_05)、95% 置信水平的期望损失值 (ES)、波动率 (StdDev)、对市场因子的敞口 (BetaCAPM); (3) 管理费比例 (MgtFee) 和业绩报酬比例 (IctFee)。其中, 控制上述 (1) 和 (2) 类变量不影响样本数。(3) 类变量观测数较少, 对其进行控制会对样本量产生较大影响, 故我们仅将其用于稳健性分析。本文回归分析所用的主要变量如附表 1 所示。为直观, VaR_0_05 和 ES 均为收益率原值。

附表 1 变量定义

变量名称	变量定义	变量描述
Panel A 业绩收益表现		
<i>ExcessReturn</i>	基金(组合)超额收益率(%)	基金收益率减去无风险利率
<i>Alpha</i>	基金(组合)alpha(%)	详见三(二)正文
Panel B 系统性尾部风险		

<i>TailRisk</i>	基金系统性尾部风险水平	详见三（二）正文
Panel C 控制变量		
<i>Age</i>	基金年龄（年）	基金净值报告日期与基金成立日间隔
<i>ES</i>	基金收益期望损失值（%）	由基金收益率滚动估计得到，置信水平为 95%
<i>BetaCAPM</i>	基金市场因子敞口	由基金收益率对市场因子滚动回归得到
<i>Skewness</i>	基金收益偏度	由基金收益率滚动估计得到
<i>Kurtosis</i>	基金收益峰度	由基金收益率滚动估计得到
<i>Var_0_05</i>	基金收益在险价值（%）	由基金收益率滚动估计得到，置信水平为 95%
<i>StdDev</i>	基金收益波动率（%）	由基金收益率滚动估计得到
<i>MgtFee</i>	基金管理费（%，年）	朝阳永续数据库直接获得
<i>IctFee</i>	基金业绩报酬激励（%，年）	朝阳永续数据库直接获得
Panel D 风险因子		
<i>MKT</i>	市场风险因子	CSMAR 数据库直接获得
<i>SMB</i>	规模因子	CSMAR 数据库直接获得
<i>HML</i>	账面市值比因子	CSMAR 数据库直接获得
<i>UMD</i>	动量因子	CSMAR 数据库直接获得
<i>RMW</i>	盈利能力因子	CSMAR 数据库直接获得
<i>CMA</i>	投资风格因子	CSMAR 数据库直接获得
<i>TML</i>	系统性尾部风险因子	详见附录 1

附录 4 描述性统计

本文主要回归部分的样本包括 2015 年 7 月至 2024 年 6 月我国 6784 只量化私募基金。上证 50ETF 期权、上证 50 股指期货和中证 500 股指期货于 2015 年陆续出现。这些工具的出现使得量化私募通过卖空实现“策略中性”成为可能，业界也普遍将 2015 年下半年定义为我国量化私募“2.0 阶段”的开始，故本文的基准实证分析基于 2015 年下半年之后的样本展开。

附表 2 汇报了主要变量的描述性统计结果。Panel A 直接对全部基金一月份的样本点进行描述性统计，其中，基金的平均月超额收益率只有 0.070%。Panel B 逐基金（Fund by Fund）对 *ExcessReturn*、四因子 *alpha* 和 *TailRisk* 进行描述性统计，在此之下，基金的平均月超额收益率和 *alpha* 为负。

附表 2 描述性统计

Panel A 全样本统计（单位与附表 1 一致，本文其余部分同）						
变量名	样本数	均值	标准差	25%分位数	中位数	75%分位数
<i>ExcessReturn</i>	140733	0.0700	5.7958	-1.7825	0.0645	2.0172
<i>TailRisk</i>	140733	0.5595	0.6045	0.0000	0.4343	1.0027
<i>Skewness</i>	140733	0.0879	0.8682	-0.3878	0.0023	0.4992
<i>Kurtosis</i>	140733	3.8130	2.2474	2.5998	3.1124	4.0991
<i>ES</i>	140733	-7.6020	5.1069	-10.4739	-7.2558	-3.6397
<i>Age</i>	140733	3.7990	1.7525	2.5517	3.2772	4.5147
<i>Var_0_05</i>	140733	-6.4565	4.5101	-9.2209	-6.1489	-2.7204

<i>Std_dev</i>	140733	4.6539	3.1912	2.4029	4.6204	6.0322
<i>MgtFee</i>	36171	1.3384	0.5646	1.0000	1.5000	1.8000
<i>IctFee</i>	36579	19.8623	9.7092	20.0000	20.0000	20.0000
Panel B 逐基金平均统计						
<i>ExcessReturn</i>	6784	-0.2252	2.4724	-0.8907	-0.1348	0.4571
<i>Alpha</i>	4710 ¹	-0.1501	1.3530	-0.6072	-0.0713	0.3032
<i>TailRisk</i>	6784	0.5856	0.5713	0.0615	0.4379	1.0164

附录 5 系统性尾部风险的渠道分析

基于量化私募在多层次资本市场中的独特功能定位——为高净值投资者提供非传统风险因子暴露，本节将从基金经理能力、客户需求两个角度探究量化私募对于系统性尾部风险进行暴露的机制条件。

首先，基金需具备相应能力才能设计出系统性尾部风险敞口。相较于主观私募经验驱动型决策范式下样本容量受限与认知偏差的双重约束，量化模型可能更能利用海量历史数据和非线性模型精准地识别出相关风险因子，并借助高频交易与杠杆工具动态配置该风险敞口。该技术赋能的风险暴露机制本质上依赖于基金经理的算法建模能力与复杂数据处理技术（如机器学习、特征工程等），而此类技术能力的形成往往与基金经理的数理金融训练强度正相关。于是，我们使用私募基金经理的最高学历（按学位划分为是否是硕士及以上；若有多个基金经理，则取学历最高的那位）作为其数理技术能力的代理变量。我们预期学历越高的基金管理人（即对量化模型的驾驭能力可能越高）能帮助该基金具备更高的系统性尾部风险水平。同时，若该代理变量合理，基于量化与主观方法论的分野，则高学历群体应在更依赖数理能力的量化私募样本中表现出更突出的风险放大效应。

其次，私募基金在多层次资本市场中的主要服务客户为高净值人群，其投资者通常具备更强的风险偏好及承受能力。这类群体一方面可以在极端市场环境下展现出更高的风险容忍度以追求平均的正溢价，另一方面也可能对非传统风险配置存在较高需求。鉴于此，私募基金管理人可能希望通过构建系统性尾部风险敞口，来形成与客户需求相匹配的风险收益特征。考虑到学术界已广泛认可基金投资者具有本地偏好（Home Bias），而地区的高净值人群的聚集程度和财富水平一般与当地房价正相关，故我们使用基金联系地址所在城市的平均房价作为基金高净值客户需求的代理变量。我们预期有更多高净值客户需求的基金有更强大的动机提升系统性尾部风险水平。

此外，由于人才集聚（高学历）地区的房价可能更高，故为分离两个渠道的偏效应，我们将两种机制放在同一回归中进行检验。考虑到同期*TailRisk*估计窗口（2 年）和房价数据（1 年）并非在时间上完美契合，而房价又可能更多呈现横截面上的变异，我们设计了如下所示的横截面回归：

$$TailRisk_i = \beta_0 + \beta_1 Education_i + \beta_2 HousePrice_i + \sum \gamma_k Controls_{i,k} + \varepsilon_i \quad (A3)$$

其中，*Education_i*为*i*基金的基金经理最高学历是否为硕士及以上的哑变量，是则取 1，否则为 0，从朝阳永续数据库中获得。*HousePrice_i*为该地样本期内平均商品房房价（万元），从 CEIC 数据库获得。考虑到*TailRisk*可能与基金其他高阶矩风险、系统性风险以及基金其他方面的特征相关，我们在回归中加入了一系列关于基金特征的控制变量，具体与正文表 3 列（3）的控制变量相同。由于房价原始数据为当年均值，故该方程中所有变量均为每年按

¹ 为提高估计准确性，我们要求每个进入回归分析的基金至少有 10 个样本点，故样本量有所减少。

月平均后再按年平均。

附表 3 列（1）和列（2）分别报告了量化私募和主观私募的系统性尾部风险与当地房价、基金经理最高学历的关系。由 $Education_i$ 的系数可得，量化私募确实因更好的人力资本和更精进的量化模型，而能更成功地配置高系统性尾部风险水平。相比而言，即使主观私募的样本量更大，但高学历对主观私募系统性尾部风险的提升在经济学和统计学意义上也均不显著。上述结果验证了基金经理能力渠道假说。由 $HousePrice_i$ 的系数可得，量化私募的系统性尾部风险确与其所在地房价正相关，这验证了客户需求假说。对比而言，量化私募样本的回归系数为 0.042，几乎为主观私募的 2.30 倍。尽管对于主观私募而言，更高净值的客户同样有更高的风险承受能力，但可能由于基金经理数理能力上的局限性，其对于主观基金的风险暴露的需求增量较小。这也与基金经理能力渠道的逻辑一致。列（3）和列（4）报告了将商品房房价更换为商品房住宅房价的回归结果，结论依然稳健。

附表 3 系统性尾部风险的渠道分析

变量	(1)	(2)	(3)	(4)
	$HousePrice = \text{商品房平均房价}$		$HousePrice = \text{商品房住宅平均房价}$	
	量化 $TailRisk_i$	主观 $TailRisk_i$	量化 $TailRisk_i$	主观 $TailRisk_i$
$Education_i$	0.0526*** (2.6056)	0.0081 (0.7304)	0.0564*** (2.7929)	0.0067 (0.6064)
$HousePrice_i$	0.0423*** (5.1584)	0.0184*** (5.0663)	0.0356*** (3.9581)	0.0191*** (4.9672)
截距项	-0.1207** (-2.2614)	-0.1779*** (-4.9978)	-0.0842 (-1.6050)	-0.1722*** (-4.8454)
控制变量	Yes	Yes	Yes	Yes
样本量	1609	6913	1610	6905
Adj. R^2	0.6244	0.3858	0.6193	0.3813

注：该表的括号内为 t 统计量值，其标准误使用异方差稳健标准误。

附录 6 稳健性检验

我们对本文的三个主要回归分析结果进行了稳健性检验，其中包括系统性尾部风险敞口 β^{TML} （附表 4，对应表 1 中 Panel A）、 α （附表 4，对应表 1 中 Panel A）以及系统性尾部风险与收益率的关系（附表 5，对应表 3 的列（3））。其中，附表 4 中的 Panel A~F 和附表 5 中的第（1）~（6）列分别对应六种稳健性检验的结果。为了报告结果的简明，本文以六因子模型为例进行分析。

（一）系统性尾部风险敞口和 α 的稳健性检验

Panel A、B 表明，调整对重复净值的处理方法和缩尾处理不影响量化私募有正向系统性尾部风险敞口这一结论，说明该结论对于不同数据处理方法较为稳健。Panel C 也表明本文的主要结果依然稳健，但估计值和显著性均较基准回归结果有一定减弱，可能是 36 个月可回溯业绩的要求减少了（近一半的）样本量导致的。Panel D、E 的估计值和显著性均与正文很相近，说明本文主要结论对 $TailRisk$ 的构造方法较为稳健，且只考虑 ES 小于 0 的样本对本文主要结论的影响十分微小。Panel F 结果表明，即使在量化私募发展更为成熟的时期，其也对系统性尾部风险有显著正向暴露。这一方面说明本文结论在更近的时间窗口内也有较强稳健性，另一方面也深化了监管量化风险行为的政策含义。综上，本节关于 β^{TML} 的结果表明，我国量化私募整体对于系统性尾部风险存在正向敞口的主要结论稳健。

另外,附表4中也报告了基金 alpha 的表现。Panel A 表明,调整对重复净值的处理方法也不影响量化私募平均来看无法创造显著为正的 alpha 这一结论。Panel B 表明,对于收益率极端值进行处理后,基金整体 alpha 表现更差,这可能说明未缩尾前收益极高的样本点拉高了基金的整体表现。Panel C 中该结果依然稳健,但显著性、估计值均较基准回归结果有一定减弱。前者可能也是由于 36 个月可回溯业绩的要求减少了样本量,而后者可能表明有较长可靠收益率记录的基金较同业表现稍好。Panel D 和 Panel E 中的结果与基准分析结果十分相近,说明基金 alpha 的估计对不同 Tailrisk 构造方法也较为稳健。Panel F 中的估计值和显著性都比基准分析稍强,这可能反应了量化在发展早期的盈利能力稍好。总之,本节关于 alpha 的结果表明,我国量化私募整体没有创造显著为正 alpha 的能力的结论稳健。

附表 4 系统性尾部风险敞口和 alpha 的稳健性检验

Panel A 重复净值处理						
变量	(1) TML	(2) CAPM 模型	(3) 三因子模型	(4) 四因子模型	(5) 五因子模型	(6) 六因子模型
β^{TML} 均值	0.6767*** (70.2756)	0.5236*** (35.5503)	0.2627*** (17.8871)	0.2601*** (16.8423)	0.2786*** (15.9860)	0.2908*** (14.7585)
Alpha均值	-0.0966*** (-4.9658)	-0.0721*** (-4.0880)	-0.1213*** (-6.0487)	-0.1521*** (-7.5779)	-0.1933*** (-8.3859)	-0.2483*** (-10.2105)
样本量	4674	4674	4674	4674	4674	4674
Panel B 缩尾						
变量	(1) TML	(2) CAPM 模型	(3) 三因子模型	(4) 四因子模型	(5) 五因子模型	(6) 六因子模型
β^{TML} 均值	0.6797*** (73.0712)	0.5153*** (41.9994)	0.2650*** (21.4541)	0.2583*** (20.3757)	0.2696*** (19.3541)	0.2777*** (18.6292)
Alpha均值	-0.1318*** (-8.3099)	-0.1026*** (-6.5414)	-0.1447*** (-8.4988)	-0.1717*** (-10.0321)	-0.2195*** (-11.4843)	-0.2679*** (-12.7681)
样本量	4710	4710	4710	4710	4710	4710
Panel C 36 个月滚动窗口估计						
变量	(1) TML	(2) CAPM 模型	(3) 三因子模型	(4) 四因子模型	(5) 五因子模型	(6) 六因子模型
β^{TML} 均值	0.6535*** (46.5804)	0.3140*** (22.1873)	0.1358*** (9.3136)	0.1237*** (8.1139)	0.1814*** (10.8317)	0.1629*** (8.1195)
Alpha均值	-0.1561*** (-7.0552)	-0.0869*** (-3.9270)	-0.1062*** (-4.4947)	-0.1455*** (-5.7320)	-0.1816*** (-6.9261)	-0.2071*** (-6.2748)
样本量	2389	2389	2389	2389	2389	2389
Panel D TailRisk 构建: OS						
变量	(1) TML	(2) CAPM 模型	(3) 三因子模型	(4) 四因子模型	(5) 五因子模型	(6) 六因子模型
β^{TML} 均值	0.6800***	0.5320***	0.2610***	0.2538***	0.2633***	0.2712***

	(70.5504)	(36.4713)	(18.5008)	(16.9997)	(17.2023)	(16.6339)
α 均值	-0.1083***	-0.0787***	-0.1291***	-0.1540***	-0.1949***	-0.2465***
	(-5.7008)	(-4.5419)	(-6.5543)	(-7.8257)	(-8.5156)	(-10.3818)
样本量	4710	4710	4710	4710	4710	4710

Panel E TailRisk 构建：剔除 ES 大于零

变量	(1)	(2)	(3)	(4)	(5)	(6)
	TML	CAPM 模型	三因子模型	四因子模型	五因子模型	六因子模型
β^{TML} 均值	0.6770***	0.5190***	0.2577***	0.2535***	0.2638***	0.2708***
	(70.4539)	(35.4947)	(17.7523)	(16.5411)	(16.7622)	(16.4206)
α 均值	-0.0965***	-0.0727***	-0.1244***	-0.1507***	-0.1934***	-0.2436***
	(-5.0903)	(-4.1833)	(-6.3309)	(-7.6461)	(-8.4953)	(-10.1166)
样本量	4710	4710	4710	4710	4710	4710

Panel F 子样本：2018 年 1 月—2024 年 6 月

变量	(1)	(2)	(3)	(4)	(5)	(6)
	TML	CAPM 模型	三因子模型	四因子模型	五因子模型	六因子模型
β^{TML} 均值	0.6999***	0.5394***	0.2647***	0.2581***	0.2722***	0.2858***
	(70.7262)	(35.7957)	(18.3681)	(16.9130)	(17.8909)	(18.1250)
α 均值	-0.0957***	-0.0697***	-0.1321***	-0.1610***	-0.2072***	-0.2547***
	(-4.8269)	(-3.8413)	(-6.5555)	(-8.0897)	(-8.8063)	(-10.5343)
样本量	4437	4437	4437	4437	4437	4437

注：括号内为零均值假设检验的 t 统计量值。

（二）系统性尾部风险与收益率关系的稳健性检验

附表 5 通过 Fama-MacBeth 回归，检验基金系统性尾部风险与其超额收益率之间的关系。列（1）结果与表 3 相似，说明重复净值处理方法未显著影响该正相关关系。列（2）表明，未对极端值进行缩尾处理可能会低估基金系统性尾部风险与其收益率之间正相关性，这可能是由于缩尾更多减少了高风险基金大幅亏损的样本点所带来的负向影响。列（3）的估计系数也较高，可能是由于 36 个月可回溯业绩的要求筛选了更多在市场尾部幸存下来的基金，忽略了部分因高风险亏损实现而关闭的基金。列（4）、列（5）结果表明上述两种 TailRisk 的构建方法不会对该风险的捕捉和正相关关系产生较大影响。列（6）表明该正相关关系在最近的时间窗口内更加显著，这也与附表 4 中 Panel F 指出该段时间量化私募配置的 β^{TML} 更大一致。综上，本节结果表明，我国量化私募系统性尾部风险与收益率正相关的结论稳健。

附表 5 系统性尾部风险与收益率关系的稳健性检验

	(1)	(2)	(3)	(4)	(5)	(6)
变量	重复净值处理	缩尾	36 个月滚动窗口估计	TailRisk 构建：OS	TailRisk 构建：剔除 ES 大于零	2018 年 1 月—2024 年 6 月
	$ExcessReturn_t$	$ExcessReturn_t$	$ExcessReturn_t$	$ExcessReturn_t$	$ExcessReturn_t$	$ExcessReturn_t$

$TailRisk_{i,t-1}$	0.2731*** (2.7851)	0.3286*** (3.6282)	0.3333** (2.3240)	0.2032** (2.1377)	0.2727*** (2.8632)	0.3211*** (3.3134)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
样本量	132434	133877	70423	133877	132690	124313

注：除列（3）外的括号内为经 Newey-West（24 阶滞后项）调整标准误计算得到的 t 统计量值，列（3）的括号内为经 Newey-West（36 阶滞后项）调整标准误计算得到的 t 统计量值。

参考文献

Agarwal, V., S. Ruenzi and F. Weigert, 2017, “Tail Risk in Hedge Funds: A Unique View from Portfolio Holdings”, *Journal of Financial Economics*, 125(3), pp.610~636.