arXiv:2602.07761v1 [cs.CE] 8 Feb 2026

# 风险投资组合异常值的概率建模

**Kensei Sakamoto** [1] **Hasan Ugur Koyluoglu** [2] **Fuat Alican** [3] **Yigit Ihlamur** [3]

**摘要**。 本文基于每项投资的个体异常值概率以及投资之间的依赖关系，为风险投资组合绩效定义了概率度量。本研究的灵感源自银行业中用于评估违约风险的贷款组合建模。从数学角度而言，我们计算了由N个非齐次布尔结果（投资成为异常值）之和构成的概率分布，这些结果通过整体市场状况和行业效应等共同因子相互关联。具体来说，我们实施了一种潜在因子模型，其中每项投资的成功取决于一个高斯潜变量是否超过阈值；该变量由特质收益以及来自可解释共享因子的收益（股票市场、行业部门指数、地理区域和创始人类型）构成。该模型采用模拟方法，以保留异质的交易层面成功概率，并使用经验估算的相关矩阵。将模型应用于合成投资组合后，我们发现，仅凭预期异常值数量不足以评估风险投资组合。即使预期结果相同，在纳入不同程度和形式的相关性后，投资组合的可靠性和风险水平（例如，N项投资中没有异常值的概率）也可能大相径庭。在其他条件相同的情况下，多元化能够降低对同一行业等共同冲击的暴露，从而提高达到最低异常值数量的概率。然而，这种改善是以牺牲上行空间为代价的，凸显了可靠性与异常值成功集中出现的规模之间的根本权衡。通过展示投资组合构成、规模和因子暴露如何与相关性相互作用，进而影响出现零个、一个或多个异常值的概率，该框架在交易层面的异常值概率评估与目标导向的投资组合构建之间搭建了一座实用且可检验的桥梁。结合这些结果和建模假设，本文还讨论了局限性、数据代理变量的选择以及扩展方向（包括时变相关性和重尾潜在因子）。

## 1. 引言

众所周知，风险投资回报主要由少数极端结果所主导（Hall and Woodward，2010）。这些异常成功的交易通常被称为独角兽企业，其贡献占基金总绩效的比例远超其数量占比；与此同时，大多数投资都无法带来足够的回报。因此，风险投资组合的表现并不取决于平均回报，而取决于罕见异常值成功是否出现。这一特性使风险投资从根本上有别于传统资产类别，也意味着仅凭期望值不足以评估投资组合的质量。

构建风险投资组合的一项关键挑战在于，投资结果很少彼此独立。行业重点、地理暴露或创始人特征相同的公司，会受到共同的经济和技术冲击。这些共同风险敞口会导致结果之间存在相关性，从而提高共同成功和共同失败的似然。尽管如此，诸如多元化或集中化等常见的投资组合构建经验法则，往往在没有正式分析交易之间潜在依赖关系的情况下就被视为合理。

本文提出一种灵活且易于解释的框架，用于对相关的创业投资结果进行建模。我们采用一种受信用风险模型启发的潜变量方法（Vasicek，1987年；Koyluoglu 和 Hickman，1998年），将每项投资的结果建模为高斯潜变量超过某一阈值的事件。该变量由特质因素以及与行业、地理区域和创始人特征相关的共享潜在冲击共同构成。这种表述方式使我们能够对投资之间的相关性进行建模，同时保留交易层面成功概率的异质性。

本文其余部分安排如下。第2节介绍模型框架。第3节描述如何使用真实世界数据估计相关结构。第4节考察相关性的影响。第5节比较相关性条件下的各种启发式投资组合选择方案。最后，第6节总结全文，并讨论局限性和未来工作。

[1]牛津大学 [2]Oliver Wyman [3]Vela Research。通讯作者：Kensei Sakamoto<kensei.sakamoto@trinity.ox.ac.uk>, Yigit Ihlamur<yigit@vela.partners>。

预印本。*2026*年*2*月*10*日。

## 2. 模型框架

本文介绍用于对相关风险投资结果进行建模的通用框架。其目标是在保留交易层面成功概率的同时，引入可解释的依赖关系，这种依赖关系源于共同的行业、地理和创始人类型暴露。

### 2.1. 潜变量表示

我们使用潜变量 $A_i$ 对每家公司 $i$ 进行建模，该变量代表公司的潜在收益指数。因此，该指数超过某一阈值即意味着该公司成为独角兽。具体而言，我们定义

$$A_i = \mathbf{w}_i^\top \mathbf{Z} + \sqrt{1 - \mathbf{w}_i^\top \mathbf{\Sigma} \mathbf{w}_i}\epsilon_i \tag{1}$$

其中 $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ 是因子收益向量，$\epsilon_i \sim \mathcal{N}(0,1)$ 是各公司之间相互独立的特质冲击，而 $\mathbf{w}_i$ 是因子载荷向量。根据定义，$A_i \sim \mathcal{N}(0,1)$。

如果公司 $i$ 满足以下条件，则被归类为独角兽

$$A_i > \Phi^{-1}(1 - p_i) \tag{2}$$

其中，$p_i$ 表示独立成功概率，$\Phi$ 表示标准正态累积分布函数。这一表述在保留个体成功概率的同时，通过共同潜在因子引入相关性。

### 2.2. 因子载荷

在此模型中，我们假设每家公司都属于若干分类组别，包括一个行业、一个地理区域和一种创始人类型。我们使用向量$\mathbf{r}_i$对这些归属关系进行编码，其中非零分量表示公司 $i$所属的组别。

为控制各类组别的相对贡献，我们分别为行业、地理区域和创始人类型赋予权重 $S$、$G$和 $F$，其中 $S + G + F = 1$。本文设定 $S = 0.6$、$G = 0.3$和 $F = 0.1$，体现了这样一种观点：行业风险敞口是共同风险的主要来源，其次是地理区域和创始人类型。这些数值应被视为建模先验，而非经过校准的参数。

对于所属行业为 $s$、地理区域为 $g$、创始人类型为 $f$ 的公司，$\mathbf{r}_i$ 的相应分量设为 $\sqrt{S}$、$\sqrt{G}$ 和 $\sqrt{F}$。接下来，我们定义

$$\mathbf{b}_i = \frac{\mathbf{r}_i}{\sqrt{\mathbf{r}_i^\top \mathbf{\Sigma} \mathbf{r}_i}} \tag{3}$$

从而确保$\mathbf{b}_i^\top \mathbf{\Sigma} \mathbf{b}_i = 1$，使每家公司在缩放之前都具有单位方差。最终的载荷向量为

$$\mathbf{w}_i = w_0 \mathbf{b}_i \tag{4}$$

其中 $w_0$ 控制依赖关系的整体强度。

在此模型中，我们选择 $w_0$，使代表性公司总体中的平均两两相关性达到目标值 0.12。这一做法受到信用风险建模中类似相关性幅度的启发（Gordy，2003），不过在风险投资背景下，其适用性应被视为一种建模先验。

为确保这一点，考虑两家公司的相关性，$i$ 和 $j$

$$\rho_{ij} = \mathbf{w}_i^\top \mathbf{\Sigma} \mathbf{w}_j = w_0^2 \mathbf{b}_i^\top \mathbf{\Sigma} \mathbf{b}_j = w_0^2 \rho'_{ij} \tag{5}$$

其中，我们定义了 $\rho'_{ij} = \mathbf{b}_i^\top \mathbf{\Sigma} \mathbf{b}j$。对所有公司对取平均后，得到 $\bar{\rho}_{ij} = w_0^2 \bar{\rho}'_{ij}$，因此我们选择

$$w_0 = \sqrt{\frac{0.12}{\bar{\rho}'_{ij}}} \tag{6}$$

请注意，为确保模型有效，我们要求 $w_0^2 < 1$，否则潜变量 $A_i$ 的构造将无法成立。

### 2.3. 模拟流程

对于给定的投资组合，其成功概率和群体归属关系均已确定，可以使用蒙特卡洛模拟生成投资组合结果。在每次迭代中，对$\mathbf{Z}$ 和 $\epsilon_i$ 进行抽样，以计算潜变量 $A_i$，再据此通过超过阈值确定独角兽结果。重复此过程足够多次，即可得到独角兽结果的分布。

为了根据$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$高效地抽取相关因子冲击，我们对相关矩阵 $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^\top$ 进行乔列斯基分解，并构造

$$\mathbf{Z} = \mathbf{L}\mathbf{X} \tag{7}$$

其中 $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$。这确保了

$$\mathrm{Var}(\mathbf{Z}) = \mathbf{L}\mathrm{Var}(\mathbf{X})\mathbf{L}^\top = \mathbf{L}\mathbf{L}^\top = \mathbf{\Sigma} \tag{8}$$

## 3. 相关矩阵估计

为了对潜在因子组之间的依赖关系建模，我们需要估计相关矩阵 $\mathbf{\Sigma}$。由于全面的风险投资结果私募市场数据并未公开，因此本文使用公开股票收益构建经验相关矩阵。

具体而言，我们根据2020年1月至2025年12月的月度对数收益计算经验相关矩阵。行业因子以行业特定的交易所交易基金作为代理，而地理和创始人类型

因子则以一篮子公开交易公司作为代理，这些公司经过筛选，旨在广泛反映相应群体。

将各代理变量的收益汇总为一个数据矩阵，并据此计算相关矩阵 $\boldsymbol{\Sigma}$。在我们的实现中，该矩阵包含五个行业、四个地理区域和两种创始人类型，共计 11 个组。所得相关矩阵如图 1 所示。

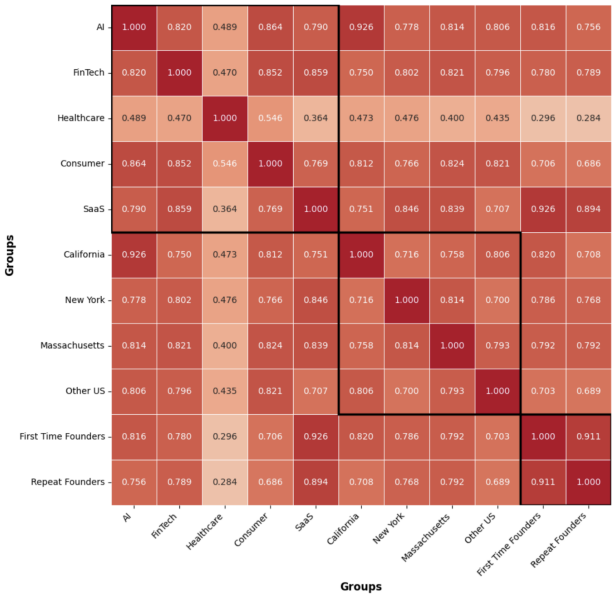


图 *1.* 人工智能、金融科技、医疗保健、消费者、软件即服务、加利福尼亚、纽约州、马萨诸塞州、其他美国、首次创业者和连续创业者的相关矩阵

从估计得到的矩阵中可以观察到几点。特别是，医疗保健行业与其他组的相关性相对较弱，这表明其受到一系列不同经济和技术因素的影响。鉴于所选的抽样时期，COVID-19 疫情也可能加剧了这一效应。相比之下，首次创业者与连续创业者之间的相关性相对较高。这可以解释为，尽管两种创始人类型的基准成功概率有所不同，但它们都受到相似的潜在市场动态影响，因此其结果呈现显著的共同变动。此外，我们还观察到人工智能行业与加利福尼亚州之间存在较强的相关性，这与专注于人工智能的公司在该州的地理集中现象一致。

在继续之前，有必要先说明这种方法存在的一些局限性。主要问题在于，对于一个关注私人市场风险投资结果的模型而言，公开股票收益是否确实是合适的数据来源。尽管公开股票收益并不能完美地代表私人风险投资结果，但它们为可能同时影响两个市场的共同经济敞口提供了具有参考价值的先验信息（Sorensen 和 Jagannathan，2015）。在实际应用中，如果能够获取历史私人市场估值数据，更稳健的实现方式理想情况下应依赖这些数据。尽管如此，所构建的矩阵仍为探索相关性如何影响风险投资组合结果提供了合理且透明的基础。

## 4. 相关性的影响

在本节中，我们考察相关性如何影响风险投资组合结果。首先，在同质概率假设下比较相关模型与非相关模型，然后针对一系列成功概率进行敏感性分析。第5节将探讨更一般的概率假设。

### 4.1. 相关性与独立性

我们首先比较潜在因子模型产生的结果与假设所有结果彼此独立时的结果。为简单起见，我们先考虑一个单因子模型，其中只有行业之间存在相关性。此外，我们假设所有交易的独立成功概率均相同，为4%。

图2展示了在非相关模型和单因子模型下，由40笔相同交易构成的投资组合所产生的独角兽数量分布，汇总统计数据见表1。

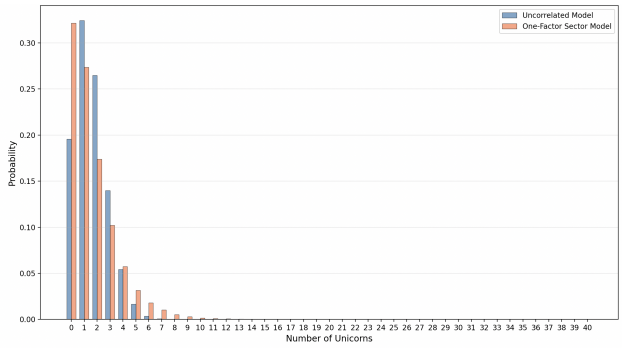


图*2.* 由40笔交易构成的投资组合的独角兽数量分布：在非相关模型和单因子模型下，成功概率均一为4%

我们首先注意到，由于构造方式如此，两种模型的独角兽期望数量相同，这反映了单笔交易层面的成功概率得以保留这一事实。尽管如此，两种分布的形态略有不同，引入相关性提高了

表1. 比较相同交易投资组合的非相关模型与单因子模型的统计数据

| | 非相关 | 单因子 |
|---|---|---|
| $\mathbb{E}[U]$ | 1.6 | 1.6 |
| $\mathbb{P}(U=0)$ | 19.6% | 32.4% |
| $\mathbb{P}(U\leq 1)$ | 52.0% | 59.5% |
| $\mathbb{P}(U\leq 2)$ | 78.5% | 76.9% |

表2. 行业多元化投资组合中非相关模型与单因子模型的统计数据比较

| | 非相关 | 单因子 |
|---|---|---|
| $\mathbb{E}[U]$ | 1.6 | 1.6 |
| $\mathbb{P}(U=0)$ | 19.6% | 29.2% |
| $\mathbb{P}(U\leq 1)$ | 52.0% | 57.8% |
| $\mathbb{P}(U\leq 2)$ | 78.5% | 76.8% |

联合成功与失败的可能性。具体而言，未产生独角兽的概率从独立性假设下的 19.6% 上升至单因子模型下的 32.4%。

另一方面，由于非相关分布的峰值在一家独角兽企业左右，因此在独立性假设下，产生至多两家独角兽企业的概率反而高于相关性假设。这说明，根据所关注的业绩阈值，独立性假设可能会严重低估或高估风险投资结果中的风险。

接下来，我们考虑将投资组合均匀分散到全部五个行业的影响。结果统计数据如表2所示。

我们看到，虽然非相关模型的结果没有变化，但单因子模型的左尾风险有所降低，投资组合未达标的概率也随之下降。这体现了投资组合构建中多元化的意义，因为风险投资的目标往往是实现一定数量的独角兽企业。

最后，我们将分析扩展至完整的多因子模型，纳入行业、地理和创始人类型之间的相关性。例如，考虑前文讨论的同一个完全按行业多元化的投资组合。单因子模型与多因子模型的比较如表3所示。

由于完整模型现在引入了地理区域与创始人类型之间的相关性，即使投资组合看起来已在各行业间实现多元化，仍会面临额外的下行风险。这凸显了对多个依赖渠道进行建模的重要性，因为当其他共同风险敞口依然存在时，仅沿一个维度实现多元化可能并不足够。

表3。 比较行业多元化投资组合单因子模型与多因子模型的统计数据

| | 单因子 | 多因子 |
|---|---|---|
| $\mathbb{E}[U]$ | 1.6 | 1.6 |
| $\mathbb{P}(U=0)$ | 29.2% | 32.5% |
| $\mathbb{P}(U\leq 1)$ | 57.8% | 59.8% |
| $\mathbb{P}(U\leq 2)$ | 76.8% | 76.9% |

表4。非相关模型的概率敏感性分析

| | 2% | 4% | 8% | 16% |
|---|---|---|---|---|
| $\mathbb{E}[U]$ | 0.8 | 1.6 | 3.2 | 6.4 |
| $\mathbb{P}(U=0)$ | 44.5% | 19.6% | 3.7% | 0.1% |
| $\mathbb{P}(U\leq 1)$ | 81.0% | 52.0% | 16.1% | 0.9% |
| $\mathbb{P}(U\leq 2)$ | 95.4% | 78.5% | 36.8% | 3.6% |

### 4.2. 概率敏感性分析

为评估上述影响是否取决于成功概率的选择，我们进行了敏感性分析：在固定投资组合构成的同时，改变单独概率。具体而言，我们考察了由40笔相同交易组成的投资组合，其成功概率分别统一设为2%、4%、8%和16%。非相关模型和完整模型的相应结果分别见表4和表5。

正如预期，提高单独成功概率会使两种模型中的独角兽期望数量成比例增加。然而，相关风险的演变在两种模型之间存在显著差异。

在独立性条件下，成功概率的提高会迅速降低下行风险。例如，将成功概率从2%提高到4%，会使不产生独角兽企业的概率从44.5%降至19.6%。相比之下，在相关模型下，风险下降要平缓得多：在成功概率同样提高的情况下，不产生独角兽企业的概率仅从54.6%降至33.4%。图 3展示了 $\mathbb{P}(U=0)$情形下的这一结果。

可以观察到，随着同质成功概率的提高，非相关模型趋近于零的速度比完整模型更快。这些结果表明，相关性从根本上限制了改善交易质量降低下行风险的幅度。这意味着，提高独立成功概率并不能同样有效地保障投资组合表现的可靠性；管理投资之间的依赖关系也至关重要。

表5. 完整模型的概率敏感性分析

| | 2% | 4% | 8% | 16% |
|---|---|---|---|---|
| $\mathbb{E}[U]$ | 0.8 | 1.6 | 3.2 | 6.4 |
| $\mathbb{P}(U = 0)$ | 54.6% | 33.4% | 14.0% | 3.1% |
| $\mathbb{P}(U \leq 1)$ | 80.7% | 60.1% | 32.2% | 9.3% |
| $\mathbb{P}(U \leq 2)$ | 91.7% | 77.1% | 49.6% | 18.1% |

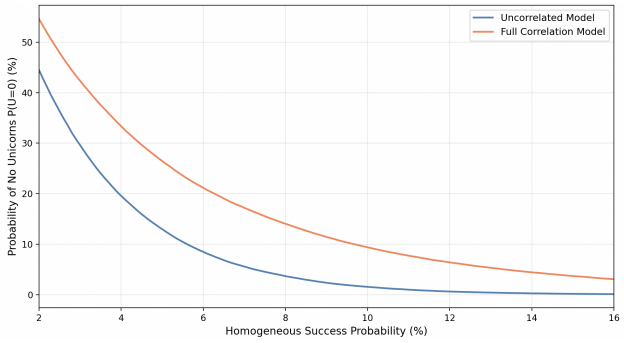


图 3. 在非相关模型和完全相关模型下，未产生独角兽的概率与同质成功概率的关系

## 5. 投资组合比较

在考察了简化概率假设下相关性的影响后，我们现在将完整模型应用于一组典型的投资组合构成。

### 5.1. 赋予合成概率

在实践中，风险投资组合中的交易具有异质的成功概率，通常通过详细分析进行估算。就本文而言，我们根据规则为每笔交易赋予合成成功概率，这些规则大体反映了真实风险投资结果中观察到的经验模式。

具体而言，由首次创业者创立的公司，其成功概率取自定义在 $[0.1\%, 12\%]$ 上的缩放贝塔分布，目标均值为 1.8%；而由连续创业者创立的公司，其成功概率取自定义在 $[1\%, 20\%]$ 上的分布，目标均值为 2.6%。对于首次创业者，我们使用参数为 $\alpha = 1, \beta = 6$ 的贝塔分布；对于连续创业者，则使用参数为 $\alpha = 1, \beta = 11$ 的贝塔分布，从而得到明显右偏的分布，这与风险投资结果的重尾特征相符。

为纳入地理和行业效应，我们对总部位于加利福尼亚或纽约州，或属于通常被视为高增长行业（如人工智能、金融科技和软件即服务）的公司，施加 +1% 的微调。此外，所有概率均限制在各自的上限以内。

表6。基准投资组合的构成

| | A | B | C | D |
|---|---|---|---|---|
| 连续创业者 | 30% | 100% | 50% | 100% |
| 首次创业者 | 70% | 0% | 50% | 0% |
| AI | 30% | 100% | 20% | 35% |
| 金融科技 | 15% | 0% | 20% | 32.5% |
| 医疗保健 | 15% | 0% | 20% | 0% |
| 消费者 | 15% | 0% | 20% | 0% |
| SaaS | 25% | 0% | 20% | 32.5% |
| CA | 40% | 100% | 25% | 50% |
| NY | 20% | 0% | 25% | 50% |
| MA | 10% | 0% | 25% | 0% |
| 其他美国 | 30% | 0% | 25% | 0% |

表 7. 基准投资组合的绩效比较

| | A | B | C | D |
|---|---|---|---|---|
| $\mathbb{P}(U = 0)$ | 37.0% | 28.6% | 37.5% | 28.0% |
| $\mathbb{P}(U \leq 1)$ | 65.5% | 54.6% | 65.8% | 54.1% |
| $\mathbb{P}(U \leq 2)$ | 81.8% | 72.4% | 82.2% | 72.3% |
| $\mathbb{E}[U\|U \geq 1]$ | 2.14 | 2.58 | 2.13 | 2.57 |
| $\mathbb{E}[U\|U \geq 2]$ | 3.09 | 3.49 | 3.06 | 3.46 |
| $\mathbb{E}[U\|U \geq 3]$ | 4.06 | 4.46 | 4.05 | 4.42 |

### 5.2. 基准投资组合比较

为总结投资组合结果，我们报告用于衡量下行风险和条件上行空间的统计数据。具体而言，我们考虑以下指标：

- $\mathbb{P}(U = 0)$、$\mathbb{P}(U \leq 1)$ 和 $\mathbb{P}(U \leq 2)$分别表示投资组合中独角兽企业数量最多为 0、1 和 2 家的概率。
- $\mathbb{E}[U|U \geq 1]$、$\mathbb{E}[U|U \geq 2]$和 $\mathbb{E}[U|U \geq 3]$ 分别表示在至少有 1、2 和 3 家独角兽企业的条件下，独角兽期望数量。

我们首先考虑以下交易数量固定为 40 笔的投资组合：

- 投资组合A：行业平均投资组合。
- 投资组合B：完全集中于人工智能、加利福尼亚州和连续创业者的投资组合。
- 投资组合C：涵盖所有群体的完全多样化投资组合。
- 投资组合D：精选但多元化的投资组合。

这些投资组合的构成汇总于表 6，其结果见表 7。

这些结果揭示了一些值得关注的见解。

我们首先观察到，投资组合 A 和 C 的风险较高，条件上行收益低于其他投资组合。这反映出它们对再次创业者的配置较少，从而降低了实现独角兽企业的总体概率。鉴于

表8. 投资组合E、F和G的行业配置

| | E | F | G |
|---|---|---|---|
| AI | 25% | 40% | 20% |
| 金融科技 | 25% | 20% | 20% |
| 医疗保健 | 25% | 20% | 40% |
| SaaS | 25% | 20% | 20% |

表9. 投资组合E、F和G的业绩比较

| | E | F | G |
|---|---|---|---|
| $\mathbb{P}(U = 0)$ | 27.45% | 27.53% | 27.38% |
| $\mathbb{P}(U \leq 1)$ | 53.69% | 53.80% | 53.64% |
| $\mathbb{P}(U \leq 2)$ | 72.12% | 72.20% | 72.09% |

创始人因子的权重较低，且两种创始人类型之间的相关性相对较高，通过多元化配置创始人类型所降低的相关性，无法抵消平均成功概率的损失。

现在比较投资组合 B 和 D，可以发现，投资组合 D 中独角兽企业数量至多为 0、1 和 2 家的概率略低。尽管差异幅度不大，但这与这样一种直觉相符：通过多元化降低相关性，会减少共同失败的似然。另一方面，尽管 $\mathbb{E}[U]$ 相同，多元化程度较高的投资组合的条件期望值仍略低。这一模式与这样的观点一致：较为集中的投资组合在避免共同失败的前提下，往往会实现更大的成功集群。综合来看，这些结果揭示了两个理想指标之间的根本张力。较低的相关性提高了达到最低成功阈值的可靠性，而较高的相关性则会在达到该阈值后扩大成功的规模。

### 5.3. 完全多元化的局限

一个自然的问题是，在成功概率相似的群体之间进行完全多元化，是否必然能将左尾风险降至最低。为探究这一细微之处，我们暂时将医疗保健归类为高增长行业，并赋予其与人工智能、金融科技和软件即服务相同的 +1 % 增幅。这一选择是基于医疗保健与其他行业之间的相关性较弱。

现在来看投资组合E、F和G，它们的创始人和地理构成与投资组合D相同，但行业配置各不相同，如表8所示。相应结果列于表9。

我们首先观察到，投资组合 E 和 F 的下行风险几乎相同，而投资组合 F 出现不良结果的概率略高，这与其在人工智能行业中的集中化程度增加相一致。相比之下，尽管投资组合 G 的行业多样性同样有所降低，但其下行风险略低于投资组合 E，这反映出医疗保健与其他行业的相关性较弱。

这一结果突显了一个关键点：即使在成功概率相似的群体中，相关性最低的投资组合也不一定是多元化程度最高的。因此，有效的多元化需要了解其背后的相关结构。

### 5.4. 投资组合规模敏感性分析

接下来，我们考察投资组合规模如何影响相应结果。在投资组合构成保持不变的情况下，我们将投资数量从5项调整到40项，并比较行业平均投资组合（投资组合A）与精选但多元化投资组合（投资组合D）。对于每种规模，我们都报告独角兽期望数量以及未能获得任何独角兽企业的概率。结果如图4和图5所示。

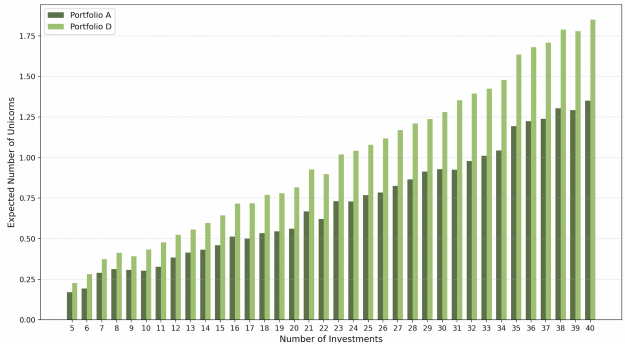


图4. 投资组合A和D按投资数量划分的独角兽期望数量

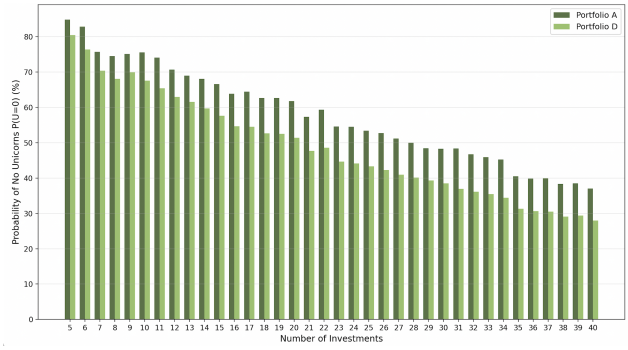


图5. 投资组合 A 和 D 中，独角兽企业数量为零的概率与投资数量的关系

正如预期，图4显示，在两种情况下，独角兽期望数量都随投资组合规模近似线性增长。此外，由于精选组合的独立成功概率平均更高，因此在所有规模下，其独角兽期望数量始终更高。更重要的是，随着投资组合规模扩大，两个投资组合的独角兽期望数量差距也随之扩大，这反映了更高质量的项目筛选所产生的累积效应。

相比之下，图5呈现出略有不同的表现。首先，随着投资组合规模扩大，未产生独角兽的概率必然趋近于零，因此风险降低带来的收益会逐渐递减，而非呈线性改善。此外，尽管精选组合未产生独角兽的概率始终低于行业平均投资组合，但这一优势的幅度仅随投资组合规模逐步扩大。这与预期独角兽企业数量所呈现的情况形成鲜明对比，表明扩大投资组合规模在提升预期结果方面，比在降低风险方面更有效。

## 6. 结论

本文提出了一个用于对相关风险投资结果进行建模的潜在因子框架，并比较了多种启发式组合筛选方法。模型将每项投资的成功表示为潜在高斯变量超过阈值的结果，并纳入共享的行业、地理和创始人类型暴露，从而在保留交易层面成功概率异质性的同时，捕捉现实的依赖结构。

### 6.1. 关键要点

我们的分析揭示了若干重要见解。

首先，仅凭预期独角兽企业数量不足以评估风险投资组合。存在相关性时，预期结果相同的投资组合，其可靠性和风险水平可能大相径庭。

其次，多元化可以通过降低对共同冲击的暴露，提高达到最低成功门槛的概率。然而，达到这些门槛后，条件上行收益会因此降低，这凸显了可靠性与集群式成功规模之间的根本权衡。重要的是，相关性最低的投资组合并不一定是简单意义上多元化程度最高的投资组合；即使弱相关群组的独立成功概率较低，将资金配置给这些群组也可能有益。

第三，尽管增加投资数量会放大预期结果，但在降低相关性下的左尾风险方面，收益会逐渐递减，即使是经过筛选且多元化的投资组合也不例外。因此，即使规模相对较大的投资组合，仍容易受到共同经济冲击的影响，从而导致交易共同失败。

第四，相关性从根本上限制了提高单笔交易成功概率所能带来的下行风险降幅。当结果存在相关性时，即使独立成功概率大幅提高，左尾风险也只会逐步降低，这意味着仅靠提升交易质量不足以确保投资组合表现可靠。

最后，最优风险投资组合构建在很大程度上取决于投资者目标。将未能达到最低成功门槛的概率降至最低的投资组合，与在成功发生后最大化条件上行收益的投资组合并不相同，这凸显了在存在依赖关系时采用目标导向型组合设计的必要性。

更广泛地说，所提出的框架为在现实的依赖结构下对风险投资组合进行压力测试提供了一种易于处理的方法，搭建了启发式投资组合构建与正式风险建模之间的桥梁。该框架奠定了灵活的基础，可用于未来研究中确定最优投资组合结构。

### 6.2. 局限性与未来工作

我们分析的一项关键局限在于假设每家公司只属于一个行业，而现实中往往并非如此。将模型扩展以允许多行业归属，可以更真实地呈现风险投资组合，但会大幅增加模型复杂性，并可能降低可解释性。

出于类似原因，我们的分析在很大程度上抽象掉了单笔交易层面行业、地理区域和创始人类型之间的相互作用。尽管我们发现，在所假设的相关结构和概率分配下，提高连续创业者的敞口通常有益，但如果连续创业者交易集中于特定行业或地理区域，这一结论可能并不成立。考虑此类相互作用对于开展更严谨的分析十分重要，但超出了本文的范围。

此外，尽管我们仅将行业、地理区域和创始人类型视为共同风险的来源，但风险投资结果也可能受到其他潜在因子的影响，这些因子可在未来工作中加以探讨。

转向模型本身，如前所述，一个主要问题是使用公开股票代理指标来估算各组之间的相关性。虽然公开市场可能反映了共同的深层经济因素，但它们很可能无法准确替代私募市场动态。未来工作可以纳入私募市场

估值数据或融资轮次结果，从而更可靠地估算风险投资之间的依赖关系。

另一个局限在于，假设不同群组之间的相关性随时间保持不变。实际上，风险投资市场具有周期性，系统性风险上升的时期之后，往往会出现更具特异性的时期。在更稳健的模型中，我们希望能够让相关矩阵随时间变化，从而使模型涵盖风险投资周期中的不同阶段。

最后，按照潜在因子信用风险模型的常见做法，该框架假设潜在因子服从高斯分布。然而，鉴于风险投资结果具有重尾特性，这一假设可能会低估联合极端聚集的似然。若将重尾潜在因子（如 t 分布）纳入框架，模型便能更好地捕捉风险投资结果中的极端共同变动。

## 参考文献

Gordy, M. B. (2003). 基于评级的银行资本规则的风险因子模型基础。《金融中介期刊》。

Hall, R. E. 和 Woodward, S. E.（2010）。创业中不可分散风险的负担。《美国经济评论》。

Koyluoglu, H. U. 和 Hickman, A.（1998）。可调和的差异。风险，11(10)：56–62。信用投资组合风险建模。Sorensen, M. 和 Jagannathan, R.（2015）。公开市场等值法与私募股权表现。《金融分析师期刊》。Vasicek, O.（1987年）。贷款组合损失概率。*KMV* 公司。