

1. 贵州省新型工业化基金、新动能基金 2021 年-2023 年整体投资情况

本节将对贵州省新型工业化基金（工业直投）2021 年-2023 年、新动能基金 2022 年-2023 年投资的所有项目的整体情况进行初步评价，选用 AHP-TOPSIS（层次分析-熵权-优劣解）对所有项目相关评价指标进行权重赋予并以此对各项目的产出水平进行评价。

不过，本节将不关注具体个体的项目运行情况、产出水平评价，而是将投资项目按照资金下达时间与项目开工时间划分为 2021 年、2022 年、2023 年三类，并对各类所包含的个体进行评价，以此得到各类的整体评价。最终观察到新型工业化基金、新动能基金三年投资的运营情况。

依据上文的讨论与数据实际情况，针对个体项目评价，构建指标表如表1所示。表中项目年产值比总投资 = 某项目平均每年所产生产值/该项

TABLE 1. 基金产出水平指标评价体系

一级指标		二级指标
基金项目产出水平	基金效益	基金年投资回报率 (%)
	被投企业经济效益	项目年产值比总投资 (%)
		项目年净利润率 (%)
		企业年利润总额增长率 (%)
被投企业建设发展	被投企业社会效益	项目稳就业人数比总投资 (%)
		企业年就业人数增长率 (%)
		企业年所缴税费增长率 (%)
企业是否定增、公开募集 企业是否上市上板	项目招商引资（包括银行贷款）比总投资 (%)	
	企业是否定增、公开募集	
	企业是否上市上板	

目总投资；企业年利润总额增长率考察企业收到投资，项目企建后企业的运营发展情况，即若某企业 2021 年 4 月收到投资并启动项目建设，那么本文将考察 2021-2022, 2022-2023 两年的利润总额增长率，并取二者的平均值作为企业年利润总额增长率指标：

$$\text{利润总额增长率} = \frac{1}{2} \times \left(\frac{2022 \text{ 利润总额}}{2021 \text{ 利润总额}} + \frac{2023 \text{ 利润总额}}{2022 \text{ 利润总额}} - 2 \right). \quad (1.1)$$

同时，在评价模型中，企业受到投资后若上市标为 1，否则为 0。

进一步地，对于三年的投资项目情况，采用层次分析法标定一级指标的权重，采用 TOPSIS（熵权法）标定二级指标的权重，结果如下表所示：

TABLE 2. 2021 年基金产出水平指标评价权重

一级指标	二级指标	权重
基金效益 (0.159)	基金年投资回报率	0.159
	项目年产值比总投资	0.243
被投企业经济效益 (0.3792)	项目年净利润率	0.110
	企业年利润总额增长率	0.027
	项目稳就业人数比总投资	0.108
被投企业社会效益 (0.3143)	企业年就业人数增长率	0.112
	企业年所缴税费增长率	0.074
	项目招商引资（包括银行贷款）比总投资	0.021
被投企业建设发展 (0.1476)	企业是否定增、公开募集	0.095
	企业是否上市上板	0.053

TABLE 3. 2022 年基金产出水平指标评价权重

一级指标	二级指标	权重
基金效益 (0.1589)	基金年投资回报率	0.159
	项目年产值比总投资	0.194
被投企业经济效益 (0.3792)	项目年净利润率	0.116
	企业年利润总额增长率	0.069
	项目稳就业人数比总投资	0.100
被投企业社会效益 (0.3143)	企业年就业人数增长率	0.121
	企业年所缴税费增长率	0.050
	项目招商引资（包括银行贷款）比总投资	0.043
被投企业建设发展 (0.1476)	企业是否定增、公开募集	0.074
	企业是否上市上板	0.074

TABLE 4. 2023 年基金产出水平指标评价权重

一级指标	二级指标	权重
基金效益 (0.1589)	基金年投资回报率	0.159
被投企业经济效益 (0.3792)	项目年产值比总投资	0.068
	项目年净利润率	0.120
	企业年利润总额增长率	0.192
被投企业社会效益 (0.3143)	项目稳就业人数比总投资	0.089
	企业年就业人数增长率	0.106
	企业年所缴税费增长率	0.056
被投企业建设发展 (0.1476)	项目招商引资（包括银行贷款）比总投资	0.063
	企业是否定增、公开募集	0.074
	企业是否上市上板	0.074

三年二级指标权重不同原因是除去三年投资项目的样本量不同、投资项目产业类别覆盖面不同所导致的系统影响（这一点难以从数据层面消除），还有整体经济环境变化、经济贸易周期等大环境影响所导致的非系统影响，在经济周期处于整体下行的阶段，企业的经济、社会效益不可避免出现整体下滑，这并不是基金或企业运行水平下降导致的结果，因此三年的各项二级指标权重应当不同。

在权重的基础上利用优劣解对个体项目产出水平做出评价，先得出加权标准化样本矩阵 V 。

$$V = v_{ijm \times n} = \{x_{ij} \times z_j\}_{m \times n} = \begin{pmatrix} x_{11}z_1 & x_{12}z_2 & \cdots & x_{1n}z_n \\ x_{21}z_1 & x_{22}z_2 & \cdots & x_{2n}z_n \\ \cdots & \cdots & \ddots & \cdots \\ x_{m1}z_1 & x_{m2}z_2 & \cdots & x_{mn}z_n \end{pmatrix} \quad (1.2)$$

其中， x_{ij} 是 i 项目 j 指标的权重， z_j 是 j 指标的权重。

(1) 计算正理想解 V^+ :

$$V^+ = (\max v_{i1}, \max v_{i2}, \dots, \max v_{in}) \quad (i = 1, 2, \dots, m) \quad (1.3)$$

负理想解 V^- :

$$V^- = (\min v_{i1}, \min v_{i2}, \dots, \min v_{in}) \quad (i = 1, 2, \dots, m) \quad (1.4)$$

(2) 计算各样本到正理想解 V^+ 的距离 D^+ 和到负理想解 V^- 的距离 D^-

$$D_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - V_j^+)^2} (i = 1, 2, \dots, m) \quad (1.5)$$

$$D_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - V_j^-)^2} (i = 1, 2, \dots, m) \quad (1.6)$$

(3) 计算相对接进度 C^*

$$C_i^* = \frac{D_i^-}{D_i^+ + D_i^-} \quad (1.7)$$

根据公式 (1.3) 到公式 (1.7)，可以计算出相对接近度。相对接近度越接近 1，代表该项目在同年的投资项目中发展水平越高。**三年相对接近度数值结果详见 APPENDIX A**

但要比较三年的项目投资发展情况无法通过计算该年所有项目的相对接进度 C^* 得到，如前文所述，无法直接排除企业发展受经济周期的影响，因此，本文将通过三年所投项目的各项指标（除去基金年收益率指标）的平均值及中位数来计算一个放大系数，该系数乘以相对接进度 C^* 得到各项目的真实运营情况。2021 年-2023 年基金各项指标如表5所示：

TABLE 5. 三年基金产出水平各指标平均值

年份	2021 所投项目	2022 所投项目	2023 所投项目
项目年产值比总投资	3.481	2.404	1.384
项目年净利润率	0.128	0.113	0.127
企业年利润总额增长率	0.047	0.013	0.323
项目稳就业人数比总投资	191.872	154.853	65.144
企业年就业人数增长率	1.581	4.132	0.169
企业年所缴税费增长率	0.679	0.343	-0.703
项目招商引资（包括银行贷款）比总投资	0.823	0.741	0.741
企业是否增、公开募集	0.081	0.080	0.100
企业是否上市上板	0.324	0.240	0.000

将相对接进度 C^* 映射为标准化的相对接进度 C' ，使得不同年份的项目评价可以排除经济周期影响，以更好地反映基金投资、企业项目运营情况。以下为具体操作步骤：

计算映射的放大系数 $coef$ ：

$$coef = \prod_{j=1}^9 z_j \times \frac{\text{本年度各项目 j 指标平均值}}{\min\{\text{三年 j 指标平均值}\}}. \quad (1.8)$$

计算映射区间的下界 inf ：

$$inf = \max\{0.6 - 0.1 \times (1 - coef), 0.5\}. \quad (1.9)$$

计算映射区间的上界 sup ：

$$sup = \min\{0.7 + 0.25 \times coef, 1\}. \quad (1.10)$$

计算 i 项目映射后接近度 C'_i ：

$$C'_i = \frac{C_i - \min\{C_i\}}{\max\{C_i\} - \min\{C_i\}} \times (sup - inf) + inf. \quad (1.11)$$

最后，根据基金出资额度的大小作为权重，对本年度所有项目的接近度计算加权平均数，以该加权平均数作为本年度基金投资情况的评估指标，得到的最终结果如下：

TABLE 6. 三年基金整体评价指标

年份	2021 所投项目	2022 所投项目	2023 所投项目
放大系数	1.642	1.227	1.000
相对接近度加权平均值	0.829	0.778	0.761

由表5和表6可知，虽然 2023 年所投项目的大多数各项指标平均值相比 2021 年、2022 年都呈现下滑趋势，尤其是企业就业人数增长率呈现大幅下滑。因此 2023 年的放大系数为基准数 1，而 2021 年、2022 年的放大系数都大于 1。但部分排除经济周期（或样本企业财报呈现出的经济周期）影响后，2023 年所投项目、企业的加权相对接近度并未呈现出断崖式下跌。一方面是 2023 年所投的项目较少，样本数量难以覆盖整体情况，

存在幸存者偏差。二是 2021 年呈现出如此高的接近度评价，除了 2021 年经济运行状况较后两年良好之外，样本数据中 2021 年所投资的项目大部分（70%）已完工，建成的生产线或完成的技术升级改造将一定程度上提高企业的各项指标水平，而 2023 年所投项目的最终情况，还需要一定时间的等待与观察。

1.1 基金基于新兴产业类别的整体投资情况

按照上文的讨论，同样可以将基金投资项目按照产业类别而不是时间进行分类来判断基金在新型产业（主要集中于新能源电池及材料行业）上的布局情况与运营情况，由于公式、操作流程上一节已详细介绍，本节不再赘述。最终的结果如下表所示：

TABLE 7. 三年基金新兴产业整体运行评价指标

是否新兴产业	新兴产业	非新兴产业
年产值占比	1.537	2.595
年净利润率	0.114	0.093
利润增长率	0.602	-0.065
稳就业人数占比	69.201	136.179
就业人数增长率	1.024	0.822
税收增长率	-0.006	0.546
招商引资合计	0.783	0.874
是否定增、公开募集	0.120	0.072
上市上板情况	0.240	0.144
放大系数	1.005	1.000
相对接近度加权平均值	0.729	0.741

由上表可以看出，新兴产业的产出水平整体要略高于非新兴产业（因此新兴产业的放大系数大于 1），而相对接近度加权平均值要略低于非新兴产业。非新兴产业略高的原因一是截至 2023 年年底，样本数据中有 67.44% 的项目已经完工，而新兴产业的完工率为 54.34%。二是新兴产业存在投资产出水平波动较大的因素，其内部 Dagum 系数偏大。即投资项目产出水平分布严重不均，部分项目出现投资后大幅增长的情况而某些项目受经济周期影响出现产出水平下滑。2024 年以来，以新能源电池及材料

为代表的新产业遭遇行业寒冬，尤其是 7 月份国内主要城市新能源车零售渗透率达到 51.1%，增长逐渐放缓。新能源电池及材料行业回暖还需耐心地等待。

2. 投资项目预计收益率模型——基于 LSTM，KAN

本文第一节只对过去基金项目的整体运行情况做了初步的探讨，但基于数据分析的角度，最重要的问题是基于所投企业过去的财务数据，如何更加准确的预测所投项目或企业战略性投资的期望收益率。

基于新型工业化基金、新动能基金 2023 年投资的调研报告、可行性调研报告、各被调研企业 2021 年-2023 年的企业财报，本文希望基于数据模型的角度给出期望收益率预测模型。

Ledoit (2014) 认为在大规模样本下，多因子模型很难完整描述影响期望收益的各项因素，且多元回归在大规模样本上表现较差。而多元线性回归的数值实验亦揭示了线性因子回归模型在上市企业组成的样本集上表现极差，在此基础上，本文进一步提出如下问题：

- (1) 如何评估一个基金项目的运营效果。
- (2) 如何确定期望收益率和超大规模因子之间的关系
- (3) 如何根据历史因子评估一个项目的可投性
- (4) 如何解决多因子模型在大规模数据集上表现较差的问题。
- (5) 如何避免大规模因子所产生的多重共线性问题。

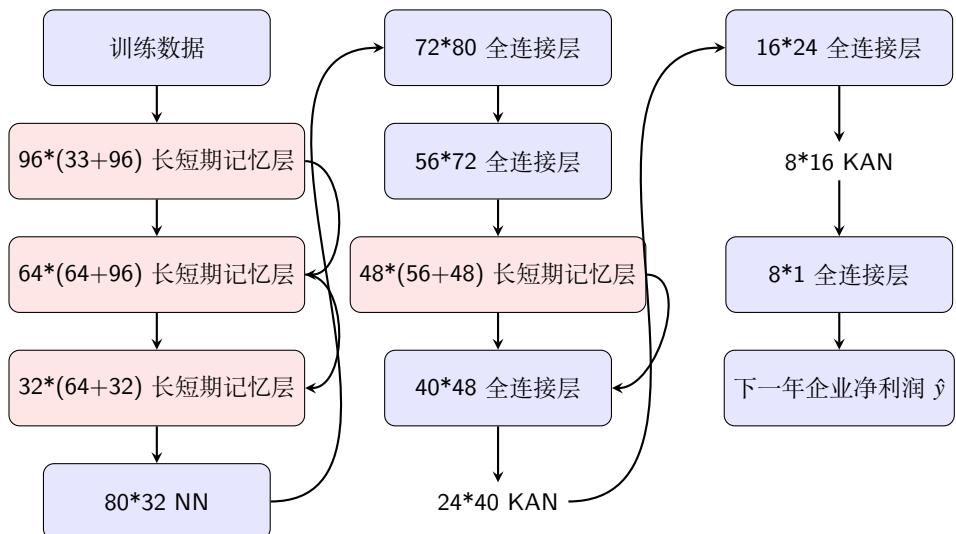
本文将采用深层神经网络来深度学习企业投资预期收益率与最近三年企业财务数据、最近三年市场因子数据之间的关系。由于这是一个高维时间序列数据，为了规避过拟合、多重共线性问题，本文采用长短期记忆神经网络 (LSTM, Felix A. etc 2002) 对时间序列数据进行处理，并将上市公司 $(t-1, t-2, \dots, t-k) i.e. k < t$ 期的收益率作为时间序列数据加入训练，由于 LSTM 本身的特性，它会自动选择不重要的信息予以遗忘，因此能有效规避大规模因子的干扰并筛选掉其中不重要的信息（时间序列指标）。并在浅层使用 Kolmogorov–Arnold Networks(KAN, Ziming Liu. etc 2024)，即使用多维样条插值对神经网络的边进行插值，以取代传统的神经网络对点激活函数来加强学习效果。

由于贵州省新型工业化基金、新动能基金样本数量不足以支持深层神经网络训练，因此选取 2000 年-2023 年 24 年间中国 A 股市场所有上市企业历年财报数据作为样本集（数据来源：YFINANCE 库，wind 数据库），按照 7:3 划分训练集与验证集，选取采样时间点最近三年（如 2000 年的训练数据由 1998, 1999, 2000 年三年的财务数据构成）的财务数据，构成 [125071,3,33] 维度的时间序列数据。并进一步计算如下指标：

TABLE 8. 训练集数据结构

指标名称	计算公式
现金流动性	货币资金/资产总计
流动资产构成率	流动资产合计/资产总计
流动比率	流动负债合计/资产总计
B/M	账面价值/资产总计
净利润率	净利润/营业收入
营业利润率	营业利润/营业收入
资产利润率	净利润/资产总计
资产增长率	(期末资产总计-期初资产总计) /期初资产总计
Market-Weight	企业净资产/市场净资产总额
利税能力	税金及附加/营业收入
未分配利润	未分配利润
回报率	每股收益/股票期初开盘价

基于上述的论述，本文构建了含 14 层的神经网络（12 个隐藏层）共计 46440 个待训练参数，其具体构成如下图所示：



选择带有正则项的损失函数对模型进行训练：

$$Loss = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 + \lambda \|\beta\| \quad (2.1)$$

其中， Y_t 是当前时间点的回报率，它是已知的， \hat{Y}_t 是当前时间点模型预计的预期收益率， λ 是正则化超参数， β 是训练参数集合。

(1) 在每一个 LSTM 隐藏层，隐藏层输出 $h = (h_{t-1}, h_{t-2}, \dots, h_{t-k})$ 的表达式为：

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t, \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h_t &= o_t * \tanh(C_t). \end{aligned} \quad (2.2)$$

其中， $\sigma(\cdot)$ 是激活函数，本文编程中采用 sigmoid: $\sigma(x) = \frac{1}{1+e^{-x}}$ 激活函数。 x_t 是上一个隐藏层的输出值，本层的输入值。 $W_f, b_f, W_i, b_i, W_C, b_C, W_o, b_o$ 都是待训练参数。

(2) 在每一个全连接层 (NN)，隐藏层输出 h 的表达式为：

$$h = \sigma(W \cdot x + b). \quad (2.3)$$

NN 层的激活函数选取 leaky-relu 函数: $\sigma(x) = \max\{\alpha x, x\}$, $\alpha < 0$ 。

(3) 在每一个 KAN 层，输出 $h(x)$ 的表达式为：

$$\begin{aligned} \text{spline}(x) &= \sum_i c_i B_i(x), \\ b(x) &= \text{silu}(x) = x / (1 + e^{-x}), \\ h(x) &= W_b b(x) + W_s \text{spline}(x). \end{aligned} \quad (2.4)$$

其中， $\text{spline}(x)$ 是可学习样条插值函数， c_i, W_b, W_s 都是待训练参数。

模型共计训练 2000 次 epoch，其损失如下图所示：

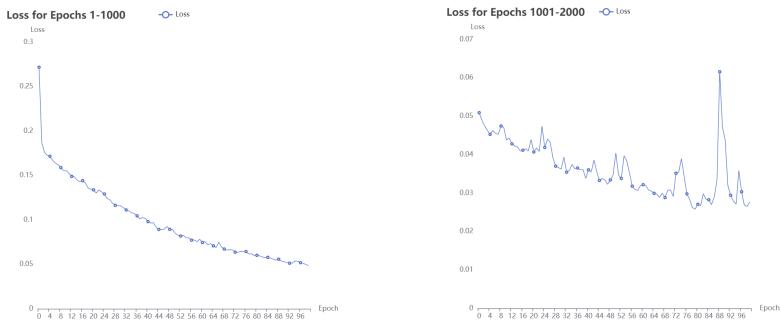


FIGURE 2.1. 模型训练 Loss 示意图，每 10 个 epoch 计算一次损失平均值。

最终 KAN 层的训练可视化结果如下图所示：

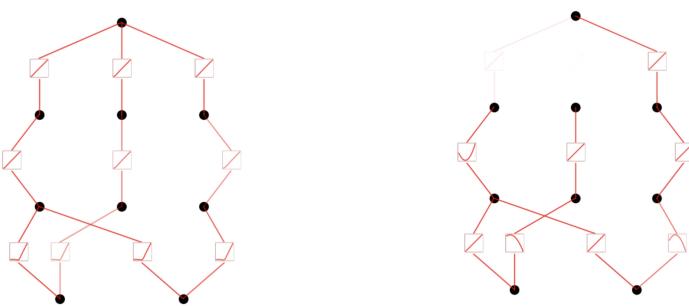


FIGURE 2.2. KAN 层结果示意图

在模型训练过程中的重要参数设置：损失函数：带正则项；初始学习率：**0.005**；学习率衰减函数：指数衰减；全连接层（NN）**dropout** 比率：**0.2**；KAN 层插值方法：高维样条插值；样条插值格点数量：**20**；梯度下降方法：**ADAM**；**batch-size: 64**；超参数最优取值： $\lambda = 0.0976$ ；框架：**TENSORFLOW**。

使用该模型训练结果对 2023 年新型工业化基金、新动能基金所投项目的收益率做出预测，其结果如表9、10、11所示：

TABLE 9. 可使用样本 2023 年指标数据表

企业名称	现金流动性	流动资产构成率	流动比率
兴义市电力有限责任公司	0.026	0.483	0.435
盘州盘誉泰合机械有限公司	0.073	0.621	0.797
盛屯新能源材料（贵州）有限公司	0.644	0.703	0.774
贵安新区大数据科创城产业集群有限公司	0.039	0.429	0.097
贵州中伟兴阳储能科技有限公司	0.082	0.355	0.809
贵州中鼎高精铜箔制造有限公司	0.060	0.620	0.916
贵州六盘水双元铝业有限责任公司	0.138	0.695	0.848
贵州凯迪森新能源材料有限公司	0.084	0.642	1.000
贵州创梦联钙业有限公司	0.001	0.194	0.488
贵州南山婆食品加工有限公司	0.053	0.795	0.409
贵州天润矿业有限公司	0.145	0.148	0.995
贵州天美锂能新材料有限公司	0.274	0.552	0.998
贵州安达科技能源股份有限公司	0.073	0.435	0.986
贵州明晟新材料有限公司	0.609	0.906	0.500
贵州智慧快刻环保科技有限公司	0.011	0.208	0.745
贵州美锦华宇新能源有限公司	0.149	0.332	0.707
贵州航盛锂能科技有限公司	0.050	0.457	0.857
贵州航飞精密制造有限公司	0.112	0.594	0.590
贵州邦盛新能源材料有限责任公司	0.114	0.198	0.562
贵州锂想时代新能源材料有限公司	0.001	0.681	1.000
贵州锦兴轻合金装备制造有限公司	0.027	0.417	0.798
贵州锦源新能源有限公司	0.377	0.594	0.773
贵州页岩气勘探开发有限责任公司	0.080	0.116	0.163
贵州骐信实业有限公司	0.178	0.706	0.451
贵州黔玻永太新材料有限公司	0.215	0.360	0.961
黔西市蕙黔新材料有限公司	0.028	0.407	1.000

TABLE 10. 可使用样本 2023 年指标数据表 (续)

企业名称	B/M	净利润率	营业利润率	Market-Weight
兴义市电力有限责任公司	0.254	0.008	0.010	0.048
盘州盘誉泰合机械有限公司	0.401	0.053	0.062	-0.021
盛屯新能源材料（贵州）有限公司	0.856	0.059	0.159	-0.008
贵安新区大数据科创城产业集群有限公司	0.594	-2.105	-2.806	0.015
贵州中伟兴阳储能科技有限公司	0.698	0.774	0.941	0.056
贵州中鼎高精铜箔制造有限公司	0.728	-0.106	-0.105	-0.035
贵州六盘水双元铝业有限责任公司	0.509	0.013	0.018	0.071
贵州凯迪森新能源材料有限公司	1.011	-10.000	-10.740	-0.027
贵州创梦联钙业有限公司	0.755	-0.342	-0.379	-0.027
贵州南山婆食品加工有限公司	0.520	0.052	-0.021	-0.014
贵州天润矿业有限公司	-0.050	26.642	26.642	-0.043
贵州天美锂能新材料有限公司	0.588	0.001	0.001	-0.033
贵州安达科技能源股份有限公司	0.484	-0.214	-0.232	0.081
贵州明晟新材料有限公司	0.044	0.000	0.000	-0.038
贵州智慧快刻环保科技有限公司	0.840	-0.183	-0.183	-0.035
贵州美锦华宇新能源有限公司	0.368	-15.260	-2.792	0.061
贵州航盛锂能科技有限公司	0.536	0.025	0.016	-0.024
贵州航飞精密制造有限公司	0.564	0.016	0.011	-0.022
贵州邦盛新能源材料有限责任公司	0.760	-85.380	-83.742	-0.005
贵州锂想时代新能源材料有限公司	0.032	-10.000	-19.545	-0.038
贵州锦兴轻合金装备制造有限公司	0.313	-0.010	-0.010	-0.016
贵州锦源新能源有限公司	0.351	-3.108	-3.092	-0.033
贵州页岩气勘探开发有限责任公司	0.532	0.014	0.014	0.094
贵州骐信实业有限公司	0.708	0.000	0.000	0.005
贵州黔玻永太新材料有限公司	0.577	0.012	0.013	0.018
黔西市蕙黔新材料有限公司	0.848	0.148	0.198	-0.031

由于某些企业并不是 2021 年前就成立，因此无法收集到三年的时间序列数据进行计算；某些企业财务数据有所缺失，故最终选择 14 家完整数据的企业进行测算。

最终的期望收益率数据如下表所示：

TABLE 11. 可使用样本投资预期收益率

企业名称	资产增长率	资产利润率	利税能力	预期收益率
兴义市电力有限责任公司	0.197	0.006	0.001	0.053
盘州盘誉泰合机械有限公司	0.635	0.020	0.004	0.015
盛屯新能源材料（贵州）有限公司	5.798	0.002	0.000	0.002
贵安新区大数据科创城产业集群有限公司	0.426	-0.005	0.003	-0.011
贵州中伟兴阳储能科技有限公司	0.048	0.015	0.024	0.056
贵州中鼎高精铜箔制造有限公司	0.822	-0.033	0.001	0.031
贵州六盘水双元铝业有限责任公司	0.639	0.007	0.003	0.009
贵州凯迪森新能源材料有限公司	3.263	-0.004	0.000	0.001
贵州创梦联钙业有限公司	0.293	-0.023	0.017	0.013
贵州南山婆食品加工有限公司	0.798	0.007	0.012	0.022
贵州天润矿业有限公司	0.111	0.009	0.048	-0.004
贵州天美锂能新材料有限公司	27.308	0.001	0.001	0.024
贵州安达科技能源股份有限公司	-0.164	-0.127	0.004	0.010
贵州明晟新材料有限公司	20.046	0.000	0.000	0.020
贵州智慧快刻环保科技有限公司	-0.019	-0.056	0.007	0.021
贵州美锦华宇新能源有限公司	8.857	-0.004	0.000	0.008
贵州航盛锂能科技有限公司	1.530	0.012	0.000	0.017
贵州航飞精密制造有限公司	1.070	0.003	0.002	0.019
贵州邦盛新能源材料有限责任公司	2.936	-0.002	55.652	0.000
贵州锂想时代新能源材料有限公司	-0.553	-0.029	0.688	0.004
贵州锦兴轻合金装备制造有限公司	0.895	-0.008	0.004	0.006
贵州锦源新能源有限公司	5.598	-0.005	0.593	0.020
贵州页岩气勘探开发有限责任公司	0.223	0.003	0.018	0.022
贵州骐信实业有限公司	4.741	-0.007	0.000	0.003
贵州黔玻永太新材料有限公司	0.765	0.051	0.001	0.008
黔西市蕙黔新材料有限公司	0.805	0.018	0.000	0.021

在本节的基础上本文将更进一步将单体项目的预期收益率模型应用在下游应用中——即投资组合或者说整个基金投资的超额收益率和市场投资组合超额收益率问题。

3. 新型工业化基金、新动能基金超额收益率预计与市场投资超额收益率

为了得到整个基金投资的超额收益率和同时间市场投资超额收益率(投资组合收益率-无风险资产收益率)，以更进一步比较基金在经济效益(即年投资回报率)上的产出水平，本节将立足于上一节对单个项目的预期收益率模型，来比较基金投资的超额收益率和同时间市场投资超额收益率。

在本报告中将采用 Fama-French (2×3) 五因子模型 (FAMA.etc 2015) 来更全面地解释贵州新型工业化基金、新动能基金的超额收益。这需要识别影响贵州基金项目收益的关键因子，通过模型量化市场、规模、账面市值比、盈利能力和投资风格对期望收益的影响，以此来解释期望收益率的变动，以便提升投资组合风险调整后的回报率。

由于新型工业化基金、新动能基金项目样本量不足，因此二者将作为验证集验证模型效果。将 2000-2022 中国所有 A 股上市公司按照贵州省两支基金所投项目行业进行细分，分成不同行业类别的股票类型，获取这些上市公司公司的财务数据、收益率数据（数据来源：wind 数据库），从而训练 Fama-French (2×3) 五因子模型，其表达式如下：

$$R_{it} - R_{ft} = a_i + b_i (R_{Mt} - R_{ft}) + s_i \text{SMB}_t + h_i \text{HML}_t + r_i \text{RMW}_t + c_i \text{CMA}_t + e_{it}, \quad (3.1)$$

其中， R_{it} 表示含有 n 只股票的投资组合 i 在第 t 年的市值平均加权年度收益率； R_{ft} 表示无风险资产收益率，选用中国人民银行公布的人民币年整存整取利率表示； R_{Mt} 表示市场收益率，用所选 A 股上市公司年度收益率表示； SMB 表示市值因子，这里采用净利润来进行分组； HML 表示账面市值比因子，用 B/M 表示账面市值比， B 为公司上一年末的账面价值，由于未上市企业不存在总市值这一数据，因此与 Fama 2015 年的模型不同的是， M 在本文采用上一年末的资产总计； RMW 表示盈利因子，用 OP 表示盈利能力指标，其等于 $t-1$ 年净利润除以 $t-1$ 年的资产总额(t -年的资产利润率)； CMA 表示资产增长因子，用 Inv 表示资产增长指标：

$$\text{INV} = \frac{(t-1) \text{ 年总资产} - (t-2) \text{ 年总资产}}{t-2 \text{ 年总资产}} \quad (3.2)$$

,

按上市公司净利润的中位数把全体股票分成小 (S) 和大 (B) 两组；按账面市值比的 30% 分位点和 70% 分位点把全体股票分成高 (H)、中 (N)、

低 (L) 三组, 低于 30% 分位点即为低 (L) 账面市值组, 大于等于 30% 分位点小于 70% 分位点的为中 (N) 账面市值组; 大于 70% 分位点的为高 (H) 账面市值组; 其次, 将市值和账面市值比两个指标进行交叉, 从而把全体股票分成 SH、SN、SL、BH、BN、BL 六个组合; 再次, 分别以资产利润率和资产增长率代替账面市值比重复上述步骤, 可把全体股票分成 SR、SN、SW、BR、BN、BW、SC、SN、SA、BC、BN、BA 这 12 个组合, 其中 R 代表盈利稳健,W 代表盈利较弱,C 代表资产总计低增长,A 代表高增长,N 代表居中的盈利能力或增长速率; 接下来计算上述各组合每一期的市值加权平均收益率; 最后, 利用不同组合收益率之差构造四个因子。具体计算方法如下所示:

$$\begin{aligned}
 SMB_{B/M} &= \frac{(SH + SN + SL)}{3} - \frac{(BH + BN + BL)}{3} \\
 SMB_{OP} &= \frac{(SR + SN + SW)}{3} - \frac{(BR + BN + BW)}{3} \\
 SMB_{Inv} &= \frac{(SC + SN + SA)}{3} - \frac{(BC + BN + BA)}{3} \\
 SMB &= \frac{SMB_{B/M} + SMB_{OP} + SMB_{Inv}}{3} \\
 HML &= \frac{SH + BH}{2} - \frac{SL + BL}{2} \\
 RMW &= \frac{SR + BR}{2} - \frac{SW + BW}{2} \\
 CMA &= \frac{SC + BC}{2} - \frac{SA + BA}{2}
 \end{aligned} \tag{3.3}$$

由于 R_{it} 表示的是投资组合的收益率, 新型工业化基金、新动能基金的各项目资金权重自然是该项目基金出资金额/本年度基金出资总额, 而对于训练集 (**A** 股市场各企业), 其投资组合权重便不好界定, 因此, 采用 Xing, Haipeng, Tze Leung Lai.etc (2010) 的投资组合优化模型 (模型介绍详见 APPENDIX B、C) 得到每次从 **A** 股股票池中随机抽取的 n 只股票的最优投资组合权重。

在五因子模型的回归结果中, HML、RMW、CMA 分别表示账面市值比溢价、盈利溢价和增长溢价。若因子的系数为正数, 则表示因子对于组合收益率的影响贡献大, 存在溢价情况, 反之则不存在。截距项表示该组合的超额收益, 如果回归结果的截距为 0, 表示没有超额收益。

选用 2000-2022 中国 A 股上市企业的历史财务数据进行回归分析，若模型中各因子的回归系数显著不同于零且通过 t 检验，则可以认为这些因子对超额收益率有显著影响。回归结果展示在表12中。

TABLE 12. A 股上市企业五因子结果 Adjusted $R^2 = 0.433$

	coef	std err	t ¹	P > t	[0.025	0.975]
const	-0.0137	0	-46.053	0	-0.014	-0.013
SMB	-0.2631	0.01	-27.365	0	-0.282	-0.244
HML	-0.2189	0.009	-25.149	0	-0.236	-0.202
RMW	0.1783	0.009	19.774	0	0.161	0.196
CMA	0.0054	0.008	0.642	0.521	-0.011	0.022
RM_f	0.154	0.005	28.44	0	0.143	0.165

模型的 t 统计量已经经过异方差性处理。其中，第一行的各个值的含义分别为：coef: 模型各因子系数；std err: 系数标准误差；t 统计量：衡量系数的统计显著性；P 值：显示统计量和响应变量间是否具有显著线性相关性；0.025: 模型因子系数置信区间的上边界值；0.975: 模型因子系数置信区间的下边界值。

在表12中，除 CMA 因子外，其余因子皆通过 t 检验，可以认为除 CMA 因子，投资组合的超额收益率受其余四个因子影响显著。且 RMW 和 RM_f 系数分别为 0.1783 和 0.154，表示盈利因子和市场因子对于所选企业的组合收益率的正面影响大。

下面将比较贵州省工业化基金和新动能基金超额收益率与同时间段市场投资超额收益率，其具体操作步骤如下：

(1) 根据贵州省工业化基金和新动能基金所投项目的行业细分，首先从股票池选取相同行业类型的上市企业总共 645 家进入股票池，再以原贵州省两支基金所投项目在每个行业在整体投资项目中所占的比例以及项目数量为参照，对这 645 家上市企业股票池按照相同比例进行有放回的抽取，(如投资了 4 个基础材料行业项目，那每一次都从股票池中抽取 4 个基础材料行业的企业进入投资组合)。

(2) 对抽取的 A 股企业，与基金投资的项目进行匹配——匹配采用 K 近邻算法，把账面市值比，净利润率，资产增长率，市场权重四个因子作为聚类的中心因子，计算 A 股抽取的企业和基金项目之间的欧氏距离。

(3) 将抽取的 A 股企业作为投资组合，运用 Fama-French 五因子模型计算其超额收益率 $R_i - R_f$ 。

(4) 将基金出资的额度作为投资组合权重，运用 Fama-French 五因子模型计算基金所投项目整体的超额收益率。

(5) 比较市场投资（不断抽取 A 股企业）超额回报率与基金投资超额回报率，得到的结果如下表所示：

TABLE 13. 股票投资组合、基金项目的因子及超额收益率 $R_i - R_f$

group		RM_f	SMB	HML	RMW	CMA	$R_i - R_f$
2023 年基金项目超额收益率		0.0029	-0.0110	0.0026	0.0057	0.0082	-0.0100
上市企业与基金相似度高的组合		0.0384	-0.0303	-0.0060	0.0255	-0.0138	0.0059
所有组合平均超额收益		0.0384	-0.0283	0.0018	0.0151	-0.0078	0.0019

通过三种组合的因子均值分析，可以明显的看出 2023 年所投基金项目的账面市值比溢价（HML）低于 A 股上市企业的对应溢价。而所投项目组合的增长溢价（CMA）和盈利溢价（RMW）却高于其他三项组合。2023 年所投的基金项目组合的超额收益率 (-0.01) 低于所有股票池的随机投资组合的平均超额收益率 (0.0019) 和相似度高的两个投资组合超额收益率 (0.0059)，所投基金项目组合的表现略低于市场平均水平。这表明尽管所投基金项目表现不尽如人意，但与市场整体低迷的情况是一致的。2023 年整体市场水平较低，因此基金项目组合未能显著超越市场表现，虽然基金项目的表现未达到预期，但考虑到市场整体表现不佳，这样的结果仍在可接受范围内。从一定程度上来说，由于 2023 年基金所投的项目较少，且所投企业与 2021 年、2022 年的企业部分重合。因此无法规避非系统误差带来的问题（如样本过少，抽样次数过少，行业局限）。

根据 2023 年基金项目中各类行业所占比例，从股票池中随机抽取了 300 个投资组合进行因子计算，表14是随机抽取的几个投资组合的因子值，可以看出每个股票投资组合的因子大小，因为选择的股票数据统一是一个

TABLE 14. 在 A 股市场中反复抽取股票得到的因子值与超额回报率 $R_i - R_f$

group	SMB	HML	RMW	CMA	$R_i - R_f$
1	-0.03174	-0.00567	-0.00103	0.027495	0.00181
7	-0.02021	0.000137	0.003382	-0.01746	-0.00195
22	-0.04004	0.0038	0.023007	0.00295	0.005948
28	-0.03806	0.001948	0.023204	-0.01792	0.005709
32	-0.026	0.016257	0.024475	-0.00981	-0.00013
42	-0.0223	-0.00506	0.006502	0.001437	0.000393
52	-0.02644	0.005079	0.011871	-0.01568	9.91E-05
59	-0.0271	0.007448	0.004471	0.007922	-0.00136
73	-0.03199	-0.01281	0.012413	0.011992	0.00564
105	-0.02694	-0.01106	0.009539	-0.02177	0.00321
144	-0.03358	-0.0125	0.030837	0.005765	0.009177
224	-0.03281	0.002353	0.010019	0.00175	0.002115
273	-0.02995	-0.00036	0.014175	-0.01112	0.002595
289	-0.02316	0.004378	0.0063	-0.00341	-0.00147

年份，所以该年份的所有股票的市场回报率都是一个值。所随机抽选的这几个投资组合的超一半的组合超额回报率高于所有组合的平均超额收益率。

不同因子对各组的回报有显著影响，特别是 CMA 和 HML，表现出高波动性。这些因子的不同值可能对投资组合的超额回报率产生不同的影响。不同的组合在实际市场回报率之外的表现有显著差异，高超额回报率组合可能在因子影响下表现更好。而这些组合中有超一半的组合超额回报率低于整体的超额回报率。在 2023 年经济处于下行周期的过程中，某些企业包括上市企业的净利润甚至跌破负值，而由于样本量只有 26 家，且部分企业与 2022 年、2021 年投资企业存在重合，整体投资的情况存在幸存者偏差，投资可能向着头部企业、优质项目集中，例如：贵州省六盘水双元铝业有限责任公司与贵州中伟兴阳储能科技有限公司，两者均已连续三年获得工业化基金的支持，累计金额分别达到 9 亿元和 15 亿元。而其余项目的超额收益情况普遍低于市场平均水平，因此，2023 年的基金投资情况和产业布局情况还需要进一步地观察和论证。

APPENDIX A: 2021-2023 样本相对接近度计算表

TABLE A.1. 2021 年及之前时间段项目相对接近度表 1

企业名称	相对接近度
贵州中伟兴阳储能科技有限公司	1.000
贵州安达科技能源股份有限公司	0.991
贵州中伟新能源科技有限公司	0.988
国家电投集团贵州金元威宁能源股份有限公司	0.985
贵州磷化新材料科技有限责任公司	0.984
贵州黔玻永太新材料有限公司	0.939
贵州页岩气勘探开发有限责任公司	0.932
贵州酱酒集团有限公司	0.931
贵州航飞精密制造有限公司	0.902
贵州雅光电子科技股份有限公司	0.902
兴义市电力有限责任公司	0.870
贵州誉福隆科技有限公司	0.863
贵州倍易通科技有限公司	0.852
贵州鹏程（集团）纸业有限责任公司	0.849
贵州南山婆食品加工有限公司	0.845
贵州中伟兴阳储能科技有限公司	0.832
贵州华尔盛新材料有限公司	0.824
贵州盘誉泰合机械有限公司	0.819
贵州省六盘水双元铝业有限责任公司	0.818
贵州大龙汇成新材料有限公司	0.817
贵州高精板带箔科技有限公司	0.816
贵州黎阳国际制造有限公司	0.811
贵州威利德制药有限公司	0.810
贵州润生制药有限公司	0.806
贵州瑞和制药有限公司	0.804
贵州木纹石材开发股份有限公司	0.797
贵州长通电气有限公司	0.791
贵州锦麟化工有限责任公司	0.786
仁怀酱酒（集团）有限责任公司	0.785
贵州中航电梯有限责任公司	0.779
贵州羚光新材料有限公司	0.757

TABLE A.2. 2021 年及之前时间段项目相对接近度表 2

企业名称	相对接近度
贵州中伟新能源科技有限公司	0.734
保利新联爆破工程集团有限公司	0.722
贵州航瑞科技有限公司	0.720
贵州航天智能制造产业集群有限公司	0.704
贵州盘誉泰合机械有限公司	0.666
贵州新能源汽车有限公司	0.664

TABLE A.3. 2022 年相对接近度表

企业名称	相对接近度
贵州安达科技能源股份有限公司	1.000
贵州盘江精煤股份有限公司	0.942
贵州盘江精煤股份有限公司	0.893
贵州南山婆食品加工有限公司	0.873
贵州能矿锰业集团有限公司	0.855
贵州凯金新能源科技有限公司	0.844
贵阳块数据城建设有限公司	0.839
贵州华芯半导体技术有限公司	0.830
泰盛（贵州）竹资源发展有限公司	0.826
贵州黑拉嘎科技装备股份有限公司	0.824
贵州金元绿链物流开发有限公司	0.814
贵州航盛锂能科技有限公司	0.779
仁怀酱酒（集团）有限责任公司	0.769
黎平清水江新能源有限公司	0.741
贵州航天天马机电科技有限公司	0.728
贵安新区大数据科创城产业集群有限公司	0.724
贵州筑威新材料科技有限公司	0.722
遵义精进钛业有限责任公司	0.701
贵州省六盘水双元铝业有限责任公司	0.692
贵州水钢新能源有限责任公司	0.675
大唐（印江）新能源有限公司	0.651
大唐（遵义播州）新能源有限公司	0.651
大唐（思南）新能源有限公司	0.651
贵州中航电梯有限责任公司	0.623
云上贵州大数据（集团）有限公司	0.623

TABLE A.4. 2023 年相对接近度表

企业名称	相对接近度
贵州中伟兴阳储能科技有限公司	0.950
贵州中云数据服务有限公司	0.794
贵州航飞精密制造有限公司	0.792
贵州红星发展股份有限公司	0.787
贵州黔玻永太新材料有限公司	0.713
保利新联爆破工程集团有限公司	0.708
贵州锦源新能源有限公司	0.702
贵州锦兴轻合金装备制造有限公司	0.697
贵州省六盘水双元铝业有限责任公司	0.658
兴义市电力有限责任公司	0.600

APPENDIX B: A 股市场股票协方差预计

在优化投资的框架下，需要同时考虑收益和风险。协方差矩阵用于计算组合的系统风险（即量化系统风险，这是由项目组合带来的系统性风险，其一般导致投资期望回报不稳定，并随着时间、利率变化、经济周期等波动较大。针对非系统风险如投资不合规等财务风险，请做好背调。）。是通过选择不同资产的权重，使得在给定风险水平下组合的预期收益最大化，或者在给定收益水平下组合的风险最小化。在投资组合优化中使用贝叶斯和收缩估计量能够改进未知参数的估计，特别是资产收益的均值 (μ) 和协方差矩阵 (σ)。传统方法依赖历史数据估计这些参数，当资产数量 (m) 相对于观察数 (n) 较大时，往往会产生高估计误差。通过贝叶斯和收缩方法，可以提供更可靠的估计，从而提升投资组合的表现。

贝叶斯估计量结合了先验信息和样本数据，通过加权平均方式整合先验协方差矩阵和样本协方差矩阵，并引入修正项来调整均值估计的不确定性，从而得出更稳健的协方差矩阵估计，协方差矩阵的估计如下：

$$\begin{aligned}\hat{\Sigma} = & \frac{n_0 - m - 1}{n + n_0 - m - 1} \frac{\Psi}{n_0 - m - 1} \\ & + \frac{n}{n + n_0 - m - 1} \left\{ \frac{1}{n} \sum_{i=1}^n (\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^T \right. \\ & \left. + \frac{\kappa}{n + \kappa} (\bar{\mathbf{r}} - \mathbf{v})(\bar{\mathbf{r}} - \mathbf{v})^T \right\}. \end{aligned} \quad (B.1)$$

Ledoit 和 Wolf (2003) 将 MLE 缩小到结构化协方差矩阵，而不是直接使用这种贝叶斯估计器，将其与样本协方差矩阵进行组合，优化收缩参数，以减少估计误差。收缩估计器的形式为：

$$\hat{\Sigma} = \delta \hat{F} + (1 - \delta) S \quad (B.2)$$

其中， S 是样本协方差矩阵， \hat{F} 是结构化协方差矩阵， δ 是收缩系数。通过优化收缩参数，使得估计的协方差矩阵在统计意义上更接近真实协方差矩阵，从而提高估计精度。

根据 A 股市场股票收益采用 Co-Variance Matrix Shrinkage 方法得到系统风险，并在此基础上对股票投资组合配置进行优化。

APPENDIX C: A 股市场投资组合最优化方案

本文给出了期望收益率、期望协方差矩阵（系统风险）的估计方式，在此基础上，为了最大化风险收益，最小化系统风险，采用最优化模型如下所示：

$$\mathbf{w}(\eta) = \arg \min_{\mathbf{w}: \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0} \{ \lambda \mathbf{w}^T \mathbf{V}_n \mathbf{w} - \eta \mathbf{w}^T \boldsymbol{\mu}_n \}, \quad (\text{C.1})$$

其中， \mathbf{V}_n 是期望协方差矩阵， $\boldsymbol{\mu}_n$ 是期望收益， \mathbf{w} 是组合权重，即资金分配权重，该权重可设置上下限，以保证重点项目不会出现权重过低，或某些项目出现过高权重情况，即式 (C.1) 变为：

$$\begin{aligned} \mathbf{w}(\eta) &= \arg \min \{ \lambda \mathbf{w}^T \mathbf{V}_n \mathbf{w} - \eta \mathbf{w}^T \boldsymbol{\mu}_n \}, \\ s.t \quad \mathbf{w}^T \mathbf{1} &= 1 \\ \mathbf{w} &\geq \mathbf{c}, \mathbf{w} \leq \mathbf{C} \end{aligned} \quad (\text{C.2})$$

\mathbf{C} 和 \mathbf{c} 分别是权重的上下限。 λ 和 η 是超参数，根据不同的 λ 和 η 值所得出的不同最优权重组称为有效前沿面。有效前沿面上的所有权重都是有效的投资组合， λ 越大，那么风险厌恶越大，更倾向于保守投资； η 越大，高波动，易受经济周期影响的股票（如材料行业股票、金融行业股票）但收益高的就越受青睐。

由式 (C.3) 将得到现有股票池的最优投资组合方案，它将通过 λ 和 η 参数的调节实现保守投资或收益最大化两种方案，并充分考虑主观投资意愿的权重平衡。

若投资组合存在设定好的预期收益目标，那么式 (C.3) 将变为：

$$\begin{aligned} \mathbf{w}(\eta) &= \arg \min \{ \lambda \mathbf{w}^T \mathbf{V}_n \mathbf{w} \}, \\ s.t \quad \mathbf{w}^T \mathbf{1} &= 1 \\ \mathbf{w}^T \boldsymbol{\mu} &= \mu^* \\ \mathbf{w} &\geq \mathbf{c}, \mathbf{w} \leq \mathbf{C} \end{aligned} \quad (\text{C.3})$$

其中， μ^* 是整个投资组合期望达到的收益目标，因此权重组乘以各个项目的期望收益应该达到该目标。并在该过程中，尽量减小系统风险，尤其尽量规避行业周期、经济周期带来的收益不稳定风险。