附录 1：基于遗传算法优化的投影寻踪模型计算过程

基于遗传算法优化的投影寻踪模型具体计算过程如下：

第一步，数据标准化处理。为消除测度指标间的量纲差异，保证数据的一致性和可比性，本文采用极差法对数据进行标准化处理，如式（1）所示，式中 $x(i, j)$ 和 $X(i, j)$ 分别为第 i 个样本的第 j 项指标的原始值和标准化值：

$$\text{正向指标: } X(i, j) = \frac{x(i, j) - x_{\min(j)}}{x_{\max(j)} - x_{\min(j)}}; \text{ 负向指标: } X(i, j) = \frac{x_{\max(j)} - x(i, j)}{x_{\max(j)} - x_{\min(j)}} \quad (1)$$

第二步，构造投影目标函数。将 m 维评价指标 $\{X(i, j) | j = 1, 2, 3, \dots, m\}$ 综合为以 $a = (a_1, a_2, \dots, a_m)$ 为投影方向的一维投影值 $z(i)$ 。接着，根据 $\{z(i) | i = 1, 2, 3, \dots, n\}$ 的一维散布图进行分类，构造投影目标函数：

$$z(i) = \sum_{j=1}^m a_j X(i, j) \quad (i = 1, 2, 3, \dots, n) \quad (2)$$

$$Q(a) = S_z \cdot D_Z \quad (3)$$

$$\text{s. t. } \begin{cases} S_z = \sqrt{\frac{\sum_{i=1}^n (z(i) - E(z))^2}{n-1}} \\ D_Z = \sum_{i=1}^n \sum_{j=1}^m [(R - r(i, j)) \cdot u(R - r(i, j))] \end{cases} \quad (4)$$

式中， $Q(a)$ 为投影目标函数； D_Z 为序列 $z(i)$ 的标准差； S_z 为序列 $z(i)$ 的局部密度； $E(z)$ 为序列 $z(i)$ 的平均值； R 为局部密度的窗口半径； $r(i, j) = |z(i) - z(j)|$ 为样本之间的距离； $u(\cdot)$ 为单位阶跃函数，当 $R \geq r(i, j)$ 时取值为1，否则为0。

第三步，优化投影目标函数。投影目标函数随投影方向变化而变化，根据最大化投影目标函数的方式估计最佳投影方向：

$$\max Q(a) = S_z \cdot D_Z \quad (5)$$

$$\text{s. t. } \sum_{j=1}^m a^2(j) = 1 \quad (6)$$

式中， $a(j)$ 为优化变量，本文运用基于实数编码的加速遗传算法（RAGA）进行优化求解。

第四步，确定综合测度值。把求得的最佳投影向量 a 代入式（2），计算得到各样本的投影寻踪值 $z(i)$ ，即各省份新质生产力水平综合测度值。

附录 2：部分指标构造过程

1. 机器人安装原始密度。

随着互联网、大数据、人工智能与实体经济深度融合，以工业机器人为代表的智能化设备安装量快速增长。本文参考 Acemoglu 和 Restrepo（2020）的做法，采用机器人安装原始密度衡量地区智能化生产工具应用水平。具体计算公式如下：

$$Rob_{it} = \sum_{j=1}^n \left(\frac{L_{ijt}}{L_{it}} \times \frac{Rob_{jt}}{L_{jt}} \right) \quad (1)$$

式中， Rob_{it} 为 t 年 i 省份的机器人安装原始密度， Rob_{jt} 为 t 年 j 行业的工业机器人存量， L_{jt} 为 t 年 j 行业的全国总就业人数， L_{it} 为 i 省份 t 年的总就业人数， L_{ijt} 为 i 省份在 t 年 j 行业的就业人数。

2. 专精特新“小巨人”企业专利申请数量。

专精特新“小巨人”企业具有专业化、精细化、特色化等特征，是我国突破“卡脖子”难题和关键核心技术的重要支撑。为获取专精特新“小巨人”企业专利申请数量，首先根据工业和信息化部公布的专精特新“小巨人”企业第一至五批名单获取企业名称及其对应省份；其次，通过国家知识产权局专利数据库检索出专精特新“小巨人”企业专利申请地址、申请日、分类号等信息；最后，整理出样本考察期内专精特新“小巨人”企业专利申请数量，并进一步将企业数据汇总为省级层面数据。

3. 综合技术自主率。

综合技术自主率反映国家产业技术自给能力，本文借鉴中国科学技术发展战略研究院《国家创新能力评价指标体系》中的有关做法进行测算，具体公式如下：

$$\text{综合技术自主率} = \frac{100 \times \text{R\&D 经费}}{2 \times (\text{R\&D 经费} + \text{技术引进费用})} + \frac{100 \times \text{国内发明专利授权数}}{2 \times (\text{国内发明专利授权数} + \text{国外发明专利授权数})} \quad (2)$$

4. 数字经济关注度。

数字经济关注度代表了公共主体的数字化发展预期，是数字经济发展所依赖的外部环境基础。本文基于百度指数（<https://index.baidu.com/v2/index.html#/>），利用 R 语言爬取各省份百度指数中与数字化相关词汇出现的数量来衡量数字经济关注度。

5. 绿色全要素生产率。

绿色全要素生产率在衡量经济增长效率的同时，将资源消耗和环境污染纳入考量，是体现绿色生产力的核心指标。本文运用考虑非期望产出的全局超效率 SBM 模型测算我国各省份绿色全要素生产率，相关数据如下：

（1）投入要素包括资本、劳动力和能源。第一，在资本方面，借鉴单豪杰（2008）的做法，以永续盘存法估算的资本存量表征资本投入，将固定资本形成总额作为每年新增投资额，固定资产投资价格指数作为投资额价格指数，估算以 2006 年为基期的省份年度实际投资额，同时采用 2007—2011 年实际投资增长率的平均值代替 2007 年资本存量增长率计算基期资本存量，进而估算出我国各省份资本存量。第二，在劳动力方面，采用各省份就业人员数代表劳动力投入。第三，在能源方面，以折合成万吨标准煤的能源消费总量表征各省份能源投入情况。

（2）期望产出为各省份地区生产总值，以 2006 年为基期进行平减处理。

（3）非期望产出包括工业 SO_2 排放量和工业化学需氧量排放。

6.人才发展环境综合指数。

高质量人才发展环境是推进人才高地建设的重要方面，本文从生活压力、交通便利、就业环境等方面出发，选取居民人均支出占可支配收入比重、高速公路覆盖率和居民人均可支配工资性收入三个指标，运用熵值法对各指标进行权重分配得到衡量人才发展环境的综合指数。

7.泰尔熵。

产业结构合理化反映不同产业之间协调程度，本文以干春晖等（2011）提出的泰尔熵度量。计算公式如下：

$$TL = \sum_{i=1}^n \left(\frac{Y_i}{Y}\right) \ln \left(\frac{Y_i/Y}{L_i/L}\right) \quad (3)$$

式中， Y 为产业产值， L 为就业人员数， i 为产业， n 为产业部门数。

参考文献：

- [1] 干春晖,郑若谷,余典范. 中国产业结构变迁对经济增长和波动的影响[J]. 经济研究,2011,46(5):4-16, 31.
- [2] 单豪杰. 中国资本存量K的再估算:1952~2006年[J]. 数量经济技术经济研究,2008,25(10):17-31.
- [3] Acemoglu D, Restrepo P. Robots and Jobs: Evidence from US Labor Markets[J]. Journal of Political Economics, 2020, 128(6): 2188-2244.

附录3：数据来源

附表1

新质生产力水平测度结果

地区	省份	总指数	总排名	新的变革	排名	质的跃升	排名
东部地区	北京	3.2325	2	1.6326	3	1.7383	1
	天津	1.3514	8	0.7540	11	1.0407	6
	河北	1.0486	18	0.5744	17	0.5453	18
	山东	2.1648	6	1.3770	5	1.0065	8
	辽宁	1.1181	17	0.6258	16	0.6503	17
	江苏	2.7084	3	1.9083	2	1.1348	5
	浙江	2.3609	5	1.4677	4	1.2229	4
	上海	2.4206	4	1.2100	6	1.6022	2
	广东	3.6007	1	2.3207	1	1.3837	3
	福建	1.4270	7	0.8478	7	1.0098	7
海南	0.8508	19	0.4751	20	0.5180	19	
中部地区	内蒙古	0.7876	24	0.3118	26	0.4616	22
	吉林	0.7981	20	0.3463	23	0.4906	20
	黑龙江	0.7484	26	0.2751	28	0.4898	21
	山西	0.8080	25	0.3806	21	0.3892	26
	安徽	1.2561	10	0.8081	8	0.8270	12
	江西	0.9996	16	0.5342	19	0.7443	13
	河南	1.1173	15	0.6356	15	0.6607	16
	湖北	1.3748	9	0.7686	9	0.8830	9
湖南	1.1885	12	0.7672	10	0.7159	14	
西部地区	广西	0.8229	23	0.3148	25	0.4587	23
	重庆	1.1490	13	0.6441	14	0.8373	11
	四川	1.4270	11	0.6739	12	0.8408	10
	贵州	0.7656	27	0.3442	24	0.4192	25
	云南	0.7125	21	0.3525	22	0.4314	24
	陕西	1.1475	14	0.6699	13	0.6661	15
	甘肃	0.6784	28	0.2712	29	0.3581	27
	青海	0.6587	29	0.2074	30	0.2425	28
	宁夏	0.8224	22	0.5450	18	0.2285	29
新疆	0.6928	30	0.3110	27	0.1315	30	

附表2 区域高新质生产力水平组态路径

条件组态	高新质生产力水平			
	S1	S2	S3	S4
科技创新	●	●	●	●
人才建设	⊗	⊗	●	●
产业结构	⊗	●	⊗	
资金供给	⊗	●		●
市场环境		⊗	●	●
政府支持	●	●		●
一致性	0.993	1.000	0.986	1.000
原始覆盖度	0.203	0.116	0.522	0.443
唯一覆盖度	0.077	0.048	0.172	0.108
总体一致性	0.989			
总体覆盖度	0.768			

注：●和⊗分别表示核心条件的存在与缺失；●和⊗分别表示边缘条件的存在与缺失；空格表示条件变量的出现与否对于结果的发生可有可无。下同。

附表3 调整校准锚点、提高 PRI 一致性阈值的稳健性检验

条件组态	高新质生产力水平 (调整校准锚点)				高新质生产力水平 (改变 PRI 一致性阈值)			
	S1	S2	S3	S4	S1	S2	S3	S4
科技创新	●	●	●	●	●	●	●	●
人才建设	⊗	⊗	●	●	⊗	⊗	●	●
产业结构	⊗	●	⊗		⊗	●	⊗	
资金供给	⊗	●		●	⊗	●	⊗	●
市场环境		⊗	●	●		⊗	●	●
政府支持	●	●		●	●	●		●
一致性	0.962	0.957	0.984	1.000	0.976	0.988	0.982	1.000
原始覆盖度	0.179	0.101	0.504	0.459	0.204	0.119	0.221	0.458
唯一覆盖度	0.068	0.036	0.169	0.118	0.075	0.049	0.094	0.345
总体一致性	0.979				0.984			
总体覆盖度	0.757				0.704			