

# 论文《投资者预期协同的资产定价研究——基于多特征机器学习视角》复现代码

本项目是论文《投资者预期协同的资产定价研究——基于多特征机器学习视角》的代码复现包，包含了实现论文中数据预处理、机器学习和实证检验等过程所需的全部代码。

## 目录

- IED 论文复现代码
  - 目录
  - 项目结构
  - 环境与数据
  - 复现流程
  - 图表对应关系
    - 正文图表
      - 图1：样本期内月度股票有效数量
      - 图2：滚动窗口法示例
      - 图3：不同风格投资者预期分散度的时间序列图
      - 图4：股市牛熊市周期划分时序图
      - 表1：描述性统计
      - 表2：投资者预期协同的单变量组合分析
      - 表3：投资者预期协同的FM回归分析
      - 表4：不同参数设定下的FM回归分析
      - 表5：基于错误定价的机制检验
      - 表6：会计信息质量与投资者预期协同的双变量组合分析
      - 表7：不同数据可得性下的单变量组合分析
      - 表8：不同投资者风格下的单变量组合分析
      - 表9：市值与投资者预期协同的双变量组合分析
      - 表10：不同股权性质和市场周期下的单变量组合分析
    - 附录图表
      - 附图1：100个投资者预测值的均方根误差
      - 附图2：投资者预测均值的均方根误差
      - 附图3：100个投资者预测值的对冲组合的FF5- $\alpha$
      - 附图4：投资者预测均值的对冲组合的FF5- $\alpha$
      - 附表1：因子指标名称与文献出处

- 附表2：投资者平均预期与预期协同的双变量组合分析
- 附表3：投资者预期协同的FM回归分析：控制个体定价因子
- 附表4：基于错误定价的机制检验（风险调整后）
- 联系方式

## 项目结构

```
IED_Paper_Replication
├── IED
│   ├── data           # 存放所有数据文件
│   │   ├── cache       # 存放中间数据和分析结果
│   │   └── raw         # 存放原始数据文件
│   └── source         # 存放 Python 源代码
│       ├── machine_learning.py    # 核心机器学习代码
│       ├── run_empirical_tests.py # 主要实证检验代码
│       └── utils             # 存放工具函数、配置文件等
└── run_replication.py      # 主控脚本，用于一键执行不同复现流程
└── README.pdf              # 本文档
```

## 环境与数据

1. **运行环境**: 本项目需要SAS、Python环境及 `tensorflow` 等核心库。此外，因子检验部分使用了作者开发的 `anomalylab` 包。请通过以下命令安装 **0.4.3** 版本：

```
pip install anomalylab==0.4.3
```

2. **原始数据**: 由于版权限制，CSMAR原始数据未包含在内。读者需自行从CSMAR数据库下载，并按**原文件名**存放于 `IED/data/raw/` 对应目录。

3. **机器学习复现性**: 神经网络模型的复现结果受到整个软硬件栈（包括但不限于GPU型号、CUDA/cuDNN库、框架版本等）的影响。因此，即使固定了随机种子，在不同机器上的复现结果也可能出现微小的浮点数差异。

- **我们的测试环境**: GPU为 **NVIDIA A10**，TensorFlow版本为 **2.2.0**。
- **结论稳健性**: 我们已确认，这些微小差异不影响本文核心系数的**符号**、**经济数量级**和**统计显著性**。

4. 中间文件: 机器学习过程的中间数据 (>23GB) 未提供, 可由脚本自动生成。如有需要, 请联系作者获取。
- 

# 复现流程

本项目提供了一个主控脚本 `run_replication.py` 来自动化整个复现流程。该脚本将流程划分为数据预处理、机器学习、实证检验、附录相关分析四个主要阶段, 您可以按需执行。

## 操作步骤:

1. 打开项目根目录下的 `run_replication.py` 文件。
2. 选择要运行的流程: 在文件底部的 `if __name__ == "__main__":` 代码块中, 通过取消注释来选择您想运行的阶段。

例如, 若要运行数据预处理流程, 请确保代码如下:

```
if __name__ == "__main__":
    # --- 按需取消注释来运行特定流程 ---

    # 选项 1: 仅运行数据预处理
    execute_workflow(DATA_PREPROCESSING_SCRIPTS, workflow_name="数据预处理")

    # 选项 2: 仅运行机器学习模型 (假定数据已处理好)
    # execute_workflow(MACHINE_LEARNING_SCRIPTS, workflow_name="机器学习")

    # 选项 3: 仅运行实证检验 (假定模型已跑完)
    # execute_workflow(EMPIRICAL_TEST_SCRIPTS, workflow_name="实证检验")

    # 选项 4: 运行附录分析
    # execute_workflow(APPENDIX_SCRIPTS, workflow_name="附录相关分析")
```

若要转而运行其他流程, 只需将上一行注释掉, 并取消对应行的注释即可。

3. 从终端运行脚本:

```
python run_replication.py
```

**特别注意：**阶段二：机器学习 由于涉及多个模型超参数、子样本等的调整，我们建议您根据 `run_replication.py` 文件中的注释指引，分别手动执行该阶段的脚本，以便更好地观察和管理结果。

# 图表对应关系

本部分详细列出了论文中各主要图表与本项目代码生成的结果文件之间的对应关系，方便您快速定位和检验研究结果。

## 正文图表

**图1：样本期内月度股票有效数量**

- 对应结果文件：
  - [IED/data/cache/stock\\_number/stock\\_number.png](#)

**图2：滚动窗口法示例**

- 对应结果文件：
  - [IED/data/cache/rolling\\_window/滚动窗口法图例.png](#)

**图3：不同风格投资者预期分散度的时间序列图**

- 对应结果文件：
  - [IED/data/cache/different\\_investors/ft\\_results/investors\\_ied \(average = False\).png](#)

**图4：股市牛熊市周期划分时序图**

- 对应结果文件：
  - [IED/data/cache/different\\_market\\_status/results/index\\_price.png](#)

**表1：描述性统计**

- 对应结果文件：
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)

- [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_3\\_0\\_100\\_16.xlsx](#)
- [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_2\\_0\\_100\\_16.xlsx](#)
- **工作表:** : summary

**表2：投资者预期协同的单变量组合分析**

- **对应结果文件:**
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
- **工作表:** : uni\_vw

**表3：投资者预期协同的FM回归分析**

- **对应结果文件:**
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_0\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_3\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_2\\_0\\_100\\_16.xlsx](#)
- **工作表:** : fm\_results

**表4：不同参数设定下的FM回归分析**

- **对应结果文件:**
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_3\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_6\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_24\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_36\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_50\\_12\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_200\\_12\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_500\\_12\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_1000\\_12\\_0\\_1\\_1\\_0\\_100\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_1\\_0\\_25\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_1\\_0\\_50\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_1\\_0\\_200\\_16.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_1\\_1\\_100\\_16\\_L1.xlsx](#)
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_1\\_0\\_1\\_100\\_16\\_L1.xls](#)  
X
  - [IED/data/cache/empirical\\_tests/NN\\_R\\_100\\_12\\_0\\_1\\_1\\_0\\_01\\_100\\_16\\_L1.xls](#)  
SX

- IED/data/cache/empirical\_tests/NN\_R\_100\_12\_0\_1\_1\_0.001\_100\_16\_L1.xlsx
  - IED/data/cache/empirical\_tests/NN\_R\_100\_12\_0\_1\_1\_1\_100\_16\_L2.xlsx
  - IED/data/cache/empirical\_tests/NN\_R\_100\_12\_0\_1\_1\_0.1\_100\_16\_L2.xls
  - IED/data/cache/empirical\_tests/NN\_R\_100\_12\_0\_1\_1\_0.01\_100\_16\_L2.xlsx
  - IED/data/cache/empirical\_tests/NN\_R\_100\_12\_0\_1\_1\_0.001\_100\_16\_L2.xlsx
- **工作表:** fm\_results

**表5：基于错误定价的机制检验**

- **对应结果文件:**
  - IED/data/cache/mispricing/short\_sale\_constraints/short\_available.xlsx
  - IED/data/cache/mispricing/institutional\_ownership/IOR.xlsx
  - IED/data/cache/mispricing/limits\_to\_arbitrage/illiq.xlsx
- **工作表:** bi\_vw

**表6：会计信息质量与投资者预期协同的双变量组合分析**

- **对应结果文件:**
  - IED/data/cache/char\_quality/results/EvaluationResult\_IED.xlsx
- **工作表:** bi\_vw

**表7：不同数据可得性下的单变量组合分析**

- **对应结果文件:**
  - IED/data/cache/char\_quantity/ft\_results/NN\_R\_RC21\_100\_12\_0\_1\_1\_0\_100\_16.xlsx
  - IED/data/cache/char\_quantity/ft\_results/NN\_R\_RC42\_100\_12\_0\_1\_1\_0\_100\_16.xlsx
  - IED/data/cache/char\_quantity/ft\_results/NN\_R\_RC63\_100\_12\_0\_1\_1\_0\_100\_16.xlsx
  - IED/data/cache/char\_quantity/ft\_results/NN\_R\_RC84\_100\_12\_0\_1\_1\_0\_100\_16.xlsx
- **工作表:** uni\_vw

**表8：不同投资者风格下的单变量组合分析**

- 对应结果文件:
  - IED/data/cache/different\_investors/ft\_results/0% value investors (average = False).xlsx
  - IED/data/cache/different\_investors/ft\_results/100% value investors (average = False).xlsx
  - IED/data/cache/different\_investors/ft\_results/0% value investors (average = True).xlsx
  - IED/data/cache/different\_investors/ft\_results/100% value investors (average = True).xlsx
- 工作表: : uni\_vw

**表9：市值与投资者预期协同的双变量组合分析**

- 对应结果文件:
  - IED/data/cache/empirical\_tests/NN\_R\_100\_12\_0\_1\_1\_0\_100\_16.xlsx
- 工作表: : bi\_vw

**表10：不同股权性质和市场周期下的单变量组合分析**

- 对应结果文件:
  - IED/data/cache/empirical\_tests/NN\_R\_100\_12\_0\_1\_3\_0\_100\_16.xlsx
  - IED/data/cache/empirical\_tests/NN\_R\_100\_12\_0\_1\_2\_0\_100\_16.xlsx
  - IED/data/cache/different\_market\_status/results/vw\_bear.xlsx
  - IED/data/cache/different\_market\_status/results/vw\_bull.xlsx
- 工作表: : uni\_vw

## 附录图表

**附图1：100个投资者预测值的均方根误差**

- 对应结果文件:
  - IED/data/cache/appendix/performance\_evaluation/original\_investors\_absolute\_predictive\_power.png

**附图2：投资者预测均值的均方根误差**

- 对应结果文件:

- [IED/data/cache/appendix/performance\\_evaluation/cumulative\\_mean\\_absolute\\_predictive\\_power.png](#)

附图3：100个投资者预测值的对冲组合的FF5- $\alpha$

- 对应结果文件：
  - [IED/data/cache/appendix/performance\\_evaluation/original\\_investors\\_relative\\_predictive\\_power.png](#)

附图4：投资者预测均值的对冲组合的FF5- $\alpha$

- 对应结果文件：
  - [IED/data/cache/appendix/performance\\_evaluation/cumulative\\_mean\\_relative\\_predictive\\_power.png](#)

附表1：因子指标名称与文献出处

- 对应结果文件：
  - [IED/data/raw/data\\_for\\_ml/char\\_description.pdf](#)

附表2：投资者平均预期与预期协同的双变量组合分析

- 对应结果文件：
  - [IED/data/cache/appendix/mean\\_synergy\\_analysis/output/investor\\_expectation\\_metrics\\_factor\\_test.xlsx](#)
- 工作表：：bi\_vw\_return, bi\_vw\_mean

附表3：投资者预期协同的FM回归分析：控制个体定价因子

- 对应结果文件：
  - [IED/data/cache/appendix/advanced\\_fm/advanced\\_fm\\_results.xlsx](#)
- 工作表：：fm\_results

附表4：基于错误定价的机制检验（风险调整后）

- 对应结果文件：
  - [IED/data/cache/mispricing/short\\_sale\\_constraints/short\\_available.xlsx](#)
  - [IED/data/cache/mispricing/institutional\\_ownership/IOR.xlsx](#)
  - [IED/data/cache/mispricing/limits\\_to\\_arbitrage/illiq.xlsx](#)
- 工作表：：bi\_vw

# 联系方式

---

- [唐国豪] - [ghtang@hnu.edu.cn]
- [陈海玮] - [chenhaiwei@stu.sufe.edu.cn]
- [朱 琳] - [zhulin@gdufe.edu.cn]
- [姜富伟] - [fwjiang@xmu.edu.cn]