#正太性JB检验

import pandas as pd

from scipy.stats import jarque\_bera

short = pd.read\_csv("financial\_data.csv")

# 删去 'TICKER\_SYMBOL' 和 'END\_DATE' 'FLAG'列

short = short.drop(['TICKER\_SYMBOL', 'END\_DATE','FLAG'], axis=1)

alpha = 0.05

for column in short.columns:

    data\_to\_test = short[column]

    stat, p\_value = jarque\_bera(data\_to\_test)

    print(f'Column: {column}')

    print(f'Jarque-Bera Statistic: {stat}')

    print(f'P-value: {p\_value}')

    # 根据 p-value 判断是否拒绝正态性假设

    if p\_value > alpha:

        print("样本数据可能来自正态分布")

    else:

        print("样本数据不是来自正态分布")

    print('\n')  # 为了输出结果之间有一些间隔

#相关系数

# 引入必要的库

import matplotlib.pyplot as plt

import seaborn as sns

# 创建一个新的Figure对象，指定图形的大小

fig = plt.figure(figsize=(13, 10))

# 计算DataFrame（short）中列之间的Spearman秩相关性

short\_corr = short.corr(method='spearman')

# 使用seaborn库的heatmap函数创建热力图

# cmap='Blues': 设置热力图的颜色映射为蓝色调。

# linewidth=0: 指定热力图中的格子之间的边框宽度为0，以获得更紧凑的显示。

# annot=True: 在热力图的每个格子中显示相关性系数的数值。

# fmt='.2f': 数值的格式，表示显示小数点后两位。

sns.heatmap(short\_corr, cmap='Blues', linewidth=0, annot=True, fmt='.2f')

# 旋转x轴标签，以便更好地适应较长的列名

plt.xticks(rotation=45)

# 保存图形为PNG文件，指定dpi（每英寸点数）为600

plt.savefig('result2\_1.png', dpi=600)

plt.show()

print(short)

short.to\_csv(r'..\result\short.csv',encoding='utf-8',index=False)