



对外经济贸易大学

## 行业板块轮动效应下的资产组合优化研究

队名：冯老师再捞我一把

成员：梁桂阁

王皓琛

匡查理

**摘要：**股票市场中不同行业间存在相互关联，一个板块的上涨可能会引起其他行业的上涨，投资者可以通过行业轮动提前布局投资组合从而获得超额收益。本文采用数据挖掘的方法，通过关联规则设置最小支持度和最小置信度阈值筛选出满足条件的关联情况。利用改进的多元 AR(p) 模型，找出满足 Markowitz 均值-方差模型中满足全局最小方差的最优投资组合点构造投资组合。通过使得不同行业的历史收益率增加一个微小的值，考察组合期望收益的变化，并以此计算出各行业对收益的贡献程度。在申万行业指数采用的派式指数编制法的基础上，以样本企业主营业务利润占比作为权重，对样本单日流通总市值进行加权，改进后指数更为直接地反映行业本身的变动情况。

**关键词：**关联规则 行业轮动 多元 AR 模型 Markowitz 均值-方差模型

目录

一、以房地产、建材和钢铁行业为对象的行业板块轮动效应研究 ..... 1

    （一）关联规则 ..... 1

    （二）模型与参数选择 ..... 2

    （三）数据处理 ..... 2

    （四）结论 ..... 4

二、基于关联规则下全行业板块轮动检验 ..... 6

三、基于改进的多元 AR 模型与 Markowitz 均值方差模型构建投资组合 ..... 7

    （一）多元 AR 模型简述与改进 ..... 7

    （二）Markowitz 均值-方差模型简述 ..... 8

    （三）样本数据回测结果分析 ..... 9

四、基于敏感性分析计算各行业板块的贡献程度 ..... 10

    （一）敏感性分析及贡献程度的定义 ..... 10

    （二）由历史数据分析各行业的贡献程度 ..... 11

五、电气设备行业指数编制——基于加权主营业务占比的改进 ..... 12

附录 1 运行代码报告 ..... 14

    （一）参数设定及数据导入 ..... 14

    （二）数据预处理——计算 alpha 收益率 ..... 15

    （三）生成滞后 5 阶的变量标记 ..... 18

    （四）基于关联规则行业间关联的数据挖掘 ..... 19

    （五）生成两两配对组合并运行宏 ..... 20

    （六）通过最小置信度和最小支持度阈值筛选关联情况 ..... 21

    （七）对各行业日内收益率分别求回归方程 ..... 22

    （八）根据得出的各行业预测收益计算出权重 ..... 28

    （九）敏感性分析及贡献程度的宏编程 ..... 30

    （十）计算投资组合中各行业占比、收益率、敏感性及贡献程度 ..... 32

    （十一）电气设备行业指数编制 ..... 34

附录 2 数据集说明 ..... 36

参考文献 ..... 37

表目录

表 1 板块涨跌情况示例 ..... 2

表 2 最小支持度为 0.4，最小置信度为 0.6 的关联情况 ..... 4

表 3 最小支持度为 0.35，最小置信度为 0.6 的关联情况 ..... 4

表 4 排除同日上涨，最小支持度为 0.3，最小置信度为 0.6 的关联情况 ..... 5

表 5 最小支持度为 0.3，最小置信度为 0.6 的轮动年份统计 ..... 5

表 6 最小支持度为 0.5，最小置信度为 0.85 的全行业关联情况 ..... 6

表 7 投资组合权重及回测收益率 ..... 9

图目录

图 1 Alpha 收益示例数据集 ..... 3

图 2 示例数据集 ..... 3

图 3 均值方差模型的前沿组合 ..... 9

图 4 行业贡献程度 ..... 11

图 4 编制指数与申万指数对比 ..... 13

# 行业板块轮动效应下的资产组合优化研究

## 一、以房地产、建材和钢铁行业为对象的行业板块轮动效应研究

由于不同产业存在上下游、互为替代或互为补充的关系，当一个行业的供需关系发生变化时，会引起关联行业也发生变动。这种变动不仅影响实体经济的供给与需求，同样会反映在金融市场上，不同行业在供需关系作用下影响其业绩，进而引发股价的变动。这种不同行业间的变动能给投资者提供一种新的投资策略，在观察到一个行业上涨时，预期其带动关联行业上涨，提前进行布局获得超额收益。

### （一）关联规则

关联规则属于数据挖掘算法，可以从海量数据中提取关系较为紧密的模块。例如在一个美国购物商场的案例中，经理发现在顾客的购物行为中，跟尿布一起购买最多的商品竟是啤酒，两个看似无关联的商品实际上却有紧密联系，产生这一现象的原因是：美国的太太们常叮嘱她们的丈夫下班后为小孩买尿布，而丈夫们在买尿布后又随手带回了他们喜欢的啤酒。关联规则可以从一系列包含项的集合中，通过设定某种关系出现的支持度和置信度，当该关系满足最小支持度阈值和最小置信度阈值时，可以认为项之间存在紧密联系，具体定义如下。

$$I = \{I_1, I_2, \dots, I_n\}$$

$I$ 是项的集合，给定一个交易数据库  $D$ ，其中每个事务是 $I$ 的非空子集，即，每一个交易都与一个唯一的标识符对应。关联规则在  $D$  中的支持度(Support)是  $D$  中事务同时包含  $X$ 、 $Y$  的百分比，即该事务发生的概率。置信度(Confidence)是  $D$  中事务已经包含  $X$  的情况下，包含  $Y$  的百分比，即条件概率。

$$Support(X, Y) = \frac{XY}{D} = P(XY)$$

$$Confidence(X \rightarrow Y) = \frac{P(XY)}{P(X)} = P(Y|X)$$

以行业板块涨跌为例，表 1 是板块在 6 日内的涨跌情况，即包含 6 个事务。项集 $I = \{A \text{ 行业}, B \text{ 行业}, C \text{ 行业}, D \text{ 行业}\}$ ，考虑关联规则  $A$  行业和  $B$  行业，事务 12346 中  $A$  行业上涨，事务 126 中  $A$  行业和  $B$  行业同时上涨， $AB=3$ ， $D=6$ 。

$$Support(X, Y) = \frac{3}{6} = 0.5 \quad Confidence(X \rightarrow Y) = \frac{\frac{3}{6}}{\frac{5}{6}} = 0.6$$

表 1 板块涨跌情况示例

日期	A 行业	B 行业	C 行业	D 行业
1	上涨↗	上涨↗	上涨↗	下跌↘
2	上涨↗	上涨↗	下跌↘	下跌↘
3	上涨↗	下跌↘	下跌↘	下跌↘
4	上涨↗	下跌↘	上涨↗	下跌↘
5	下跌↘	上涨↗	下跌↘	上涨↗
6	上涨↗	上涨↗	下跌↘	下跌↘

若设定最小支持度阈值为 0.5，最小置信度阈值为 0.6，则可以认为 A 行业和 B 行业存在同涨的关系。下文将采用上述关联规则对行业板块之间的联动效应进行检验。

## (二) 模型与参数选择

根据 CAPM（资本资产定价模型）模型，股票的收益率可以分为 Alpha（ $\alpha$ ）收益和 Beta（ $\beta$ ）收益，Beta 收益来源于个股相对于市场波动的不同幅度，受经济周期的影响，股票的涨跌很大程度上会由市场的涨跌决定，影响关联规则的使用；而 Alpha 收益能够排除个股受到市场的影响，只考虑其相对于市场的超额收益，更适用于判断不同行业之间相互影响的作用，因此本文将只考虑不同板块之间的 Alpha 收益。

$$E(r_i) = r_f + \beta_i(r_m - r_f)$$

$$\beta_i = \frac{Cov(r_i, r_m)}{\sigma_m^2}$$

$$\alpha_i = r_i - E(r_i)$$

$r_f$ : 日无风险利率

$r_m$ : 市场收益率，取上证指数的对数收益率

$r_i$ : 行业指数的对数收益率

$\beta_i$ : beta 系数，反映个股相对于市场的波动情况，此处用于体现板块相对于市场的波动幅度，本文使用的 beta 系数每年更新一次

$\alpha_i$ : 超额收益，即 alpha 收益

## (三) 数据处理

导入申万二十八个行业自 2000/1/4 至 2019/10/11 的收盘价，计算对数收益率；导入锐思数据中对应的日无风险利率和根据上证指数计算所得的对数收益率，进一步计算可得 alpha 收益如图 1(a)，将其转换为 0-1 变量，若  $\alpha_i > 0$  记为 1，否则记为 0，示例数据集如图 1(b) 所示。

VIEWTABLE: Data.Alpha				VIEWTABLE: Data.Signal			
	alpha2	alpha3	alpha4		alpha2	alpha3	alpha4
1	-0.000814415	0.0198182098	0.0063884259	1	0	1	1
2	0.0032796293	0.0268428794	0.0060811865	2	1	1	1
3	0.007503769	0.0220763028	-0.005212385	3	1	1	0
4	0.005632822	0.0123612657	-0.006715971	4	1	1	0
5	0.0132605314	0.0219120144	-0.005439042	5	1	1	0
6	0.0023635911	-0.006425084	-0.000595082	6	1	0	0
7	-0.004889282	0.0052948484	0.0019939595	7	0	1	1
8	0.0051299402	-0.006084631	-0.000810419	8	1	0	0

(a) Alpha 收益计算结果

(b) Alpha 收益转换为 0-1 变量

图 1 Alpha 收益示例数据集

接下来要利用关联规则两两对比不同行业之间是否存在关联性，考虑金融市场中信息传递及行业轮动可能存在时滞效应，在行业 A 上涨几天后与其关联的 B 行业才会上涨，因此在通过关联规则检验时，将项集扩展为 7 项：

$$I = \{A_{\text{行业}}_{t_0}, B_{\text{行业}}_{t_0}, B_{\text{行业}}_{t_{-1}}, \dots, A_{\text{行业}}_{t_{-5}}\}$$

为了统计事务的支持度和置信度，将项集改写，若 A 行业在  $t_0$  时刻上涨，B 行业  $t_0$  时刻下跌， $t_{-1} - t_{-5}$  时刻均上涨，该条记录则记为“134567”，若“13”满足最小支持度阈值和最小置信度阈值，则可以说 A 行业上涨后，B 行业在下一天有置信度下的概率也上涨。示例数据集如图 2 所示，第一条观测记录为“23456”，说明 A 行业第一天没有上涨，但是 B 行业连涨了五天。

VIEWTABLE: Work.Com2_3_1		
	period	name
1	1	23456
2	1	123457
3	1	12346
4	1	12357
5	1	12467
6	1	13567
7	1	24567
8	1	134567

图 2 示例数据集

此外，各行业的结构在较长时间维度内，相互之间的关系可能发生变动，因此将时间序列按年度划分为 19 个周期，经检验可得，样本的支持度和置信度显著提高，这也说明了样本的轮动效应在短周期内更为显著。

遍历各个行业在所有交易日的数据，找出满足阈值的关联，即为本文所研究的行业板块轮动的关系。

#### （四） 结论

首先选取房地产、建筑材料和钢铁三个行业数据进行分析。最小置信度和最小支持度阈值为经验值，设置最小支持度为 0.4，最小置信度为 0.6 时，获得 2 条观测。

表 2 最小支持度为 0.4，最小置信度为 0.6 的关联情况

持续时间	周期	领涨行业	跟涨行业	支持度	置信度
同日	2016	房地产	建筑材料	0.412955	0.766917
同日	2016	建筑材料	房地产	0.412955	0.693878

调整最小支持度为 0.35，最小置信度为 0.6 时，获得 6 条新观测。

表 3 最小支持度为 0.35，最小置信度为 0.6 的关联情况

持续时间	周期	领涨行业	跟涨行业	支持度	置信度
同日	2015	房地产	建筑材料	0.392713	0.723881
同日	2015	建筑材料	房地产	0.392713	0.702899
同日	2007	房地产	建筑材料	0.356275	0.666667
同日	2015	钢铁	建筑材料	0.360324	0.659259
同日	2015	建筑材料	钢铁	0.360324	0.644928
同日	2007	建筑材料	房地产	0.356275	0.615385

##### 1. 房地产与建筑材料行业关系紧密

由表 1 显示的关联情况，可以看到在 2016 年，房地产和建筑行业十分活跃，年内有 41.3%的交易日内两个行业均获得超越大盘的收益，且置信度较高。在房地产行业上涨时，建筑材料行业有 76.7%的概率上涨；而在建筑材料行业上涨时，房地产行业则有 69.4%的概率上涨。房地产和建筑材料行业同涨现象十分显著，且房地产的领涨效应略高于建筑材料行业。

##### 2. 特定年份的轮动效应显著

结合表 1 表 2，观测到的关联出现的周期较为集中，包括 2007、2015、2016 三个年份。市场运行情况对板块涨跌有显著影响，在这三个年份，我国股市处于牛市，且房地产行业发展势头迅猛，这些因素都能增强投资者对行业板块的信心，这也使得板块之间出现联动的现象次数增多。

##### 3. 近年同日内板块联动显著，早期行业板块轮动存在时滞

支持度反映了特定关联在不同时间长度的出现频率，据表 1 表 2 可知，板块联动在同日内现象支持度较高，这也体现了有效市场内对消息的反应十分迅速。对于可能存在 A 行业上涨后引起 B 行业多日上涨的现象，由于其支持度较同日内上涨的底，可能会被排除在外，因此排除同日内上涨的情况并调整最小支持度阈值为 0.3，最小置信度为 0.6，得到 12 条观测，如表 3 所示。排除同日内上涨的情况后，新观测包含 1 至 5 日不等的滞后，且发生滞后上涨



现象大多时间距今较久，最早追溯到 2005 年，当时的信息流通不如现在及时、全面，故跟涨行业会有数日的时滞，因此可以认为市场的有效性对行业轮动持续时间有影响。

表 4 排除同日上涨，最小支持度为 0.3，最小置信度为 0.6 的关联情况

持续时间	周期	领涨行业	跟涨行业	支持度	置信度
滞后 3 日	2007	钢铁	建筑材料	0.299595	0.654867
滞后 3 日	2005	房地产	钢铁	0.307692	0.633333
滞后 1 日	2007	房地产	建筑材料	0.331984	0.621212
滞后 1 日	2007	钢铁	建筑材料	0.283401	0.619469
滞后 5 日	2011	钢铁	建筑材料	0.279352	0.616071
滞后 4 日	2005	建筑材料	钢铁	0.279352	0.616071
滞后 3 日	2015	钢铁	建筑材料	0.336032	0.614815
滞后 3 日	2008	钢铁	建筑材料	0.295547	0.613445
滞后 3 日	2016	房地产	建筑材料	0.327935	0.609023
滞后 3 日	2008	房地产	建筑材料	0.307692	0.608
滞后 4 日	2016	钢铁	建筑材料	0.271255	0.603604
滞后 4 日	2015	建筑材料	房地产	0.336032	0.601449

#### 4. 行业轮动结构随时间周期变动

根据表 1 表 2，钢铁行业在近几年内与房地产行业与建筑材料行业的关联不大，但是根据表 3 的数据显示，回溯至十年前，钢铁与上述两个行业板块的关系十分密切。这说明了行业轮动结构不是一成不变的，不同行业的发展状况及国家不同时期的政策都会对行业产生重大影响，这就导致了行业之间关联性变动。例如我国在 2013 年后实行去产能，钢铁行业大幅缩减，生产经营受到重大偏离原有内在联系消失或发生变更。

#### 5. 发生板块轮动持续周期存在差异

设置最小支持度为 0.3，最小置信度为 0.6，寻找满足阈值的关联情况，排列出显著出现轮动现象的年份，如表 4 所示，房地产和建筑材料出现轮动频率最高，而钢铁行业相对不活跃。

表 5 最小支持度为 0.3，最小置信度为 0.6 的轮动年份统计

领涨行业	跟涨行业	轮动现象显著年份
建筑材料	房地产	2004、2006、2007、2009
		2013、2015、2016、2018
房地产	建筑材料	2001、2004、2007、2008、2009
		2013、2015、2016、2017
建筑材料	钢铁	2015、2017
房地产	钢铁	2015
钢铁	建筑材料	2001、2008
		2015、2016、2017
钢铁	房地产	2015

## 二、 基于关联规则下全行业板块轮动检验

将检验行业扩大至全行业，对所有行业两两组合，设置最小支持度阈值为 0.5，最小置信度阈值为 0.85，得到满足条件的 37 条观测。

表 6 最小支持度为 0.5，最小置信度为 0.85 的全行业关联情况

持续时间	周期	领涨行业	跟涨行业	支持度	置信度
同日	2015	通信	电气设备	0.522267	0.934783
同日	2015	化工	电气设备	0.54251	0.917808
同日	2015	计算机	电气设备	0.522267	0.908451
同日	2015	休闲服务	电气设备	0.534413	0.90411
同日	2015	商业贸易	化工	0.502024	0.898551
同日	2015	商业贸易	纺织服装	0.502024	0.898551
同日	2015	通信	轻工制造	0.502024	0.898551
同日	2015	电子	电气设备	0.526316	0.896552
同日	2015	轻工制造	电气设备	0.546559	0.89404
同日	2015	化工	轻工制造	0.526316	0.890411
同日	2015	计算机	电子	0.510121	0.887324
同日	2015	农林牧渔	电气设备	0.502024	0.885714
同日	2015	电子	轻工制造	0.518219	0.882759
同日	2014	电气设备	轻工制造	0.518219	0.882759
同日	2015	机械设备	电气设备	0.546559	0.882353
同日	2014	电子	电气设备	0.506073	0.880282
同日	2015	医药生物	电气设备	0.502024	0.879433
同日	2015	综合	轻工制造	0.526316	0.878378
同日	2014	纺织服装	轻工制造	0.522267	0.877551
同日	2015	化工	综合	0.518219	0.876712
同日	2015	纺织服装	电气设备	0.530364	0.873333
同日	2015	计算机	机械设备	0.502024	0.873239
同日	2015	综合	纺织服装	0.522267	0.871622
同日	2015	电子	计算机	0.510121	0.868966
同日	2015	综合	化工	0.518219	0.864865
同日	2015	综合	电气设备	0.518219	0.864865
同日	2014	机械设备	轻工制造	0.518219	0.864865
同日	2015	化工	纺织服装	0.510121	0.863014
同日	2014	电气设备	电子	0.506073	0.862069
同日	2015	轻工制造	化工	0.526316	0.860927
同日	2015	轻工制造	综合	0.526316	0.860927
同日	2015	纺织服装	综合	0.522267	0.86
同日	2015	休闲服务	纺织服装	0.506073	0.856164
同日	2015	电子	化工	0.502024	0.855172
同日	2015	电子	机械设备	0.502024	0.855172
同日	2015	电气设备	轻工制造	0.546559	0.85443
同日	2015	电气设备	机械设备	0.546559	0.85443

### 1. 出现强联动现象的年份集中

由表 5，设置了更高的支持度和置信度后，对于全行业的检验得到较为显著的关联情况。尤其注意到，上表中年份仅为 2014、2015 年，与第一问交叉验证，说明了特定年份对于行业轮动现象确实有显著影响。这两年我国股市处于快速上升期，投资者信心充足，伴随大量资金入市，消息刺激某一行业能够迅速带领其他相同行业一齐上涨。牛市相较于熊市出现板块联动的频率更高，在熊市中，即使有行业上涨，但在大盘的压力下关联行业仍可能下跌，此时的联动效应可能会消失。

### 2. 电气设备行业受多板块推动上涨

此外，表 5 中明显观察到电气设备行业频繁成为跟涨行业，且支持度、置信度很高。多个行业作为电气设备的领涨行业，例如通信、化工、计算机等，从供应链角度考虑，电气设备行业是这些行业的设备提供者，当通信等行业出现扩张时，对设备需求进一步扩大，便能推动电气设备行业的发展。因此，可以认为电气设备行业会受多个行业的推动，其轮动效应较其他行业而言更为显著。

### 3. 工业相关行业间发生轮动更为频繁

进一步对表 5 所列行业进行梳理，发生关联的行业大多属于工业制造行业，包括通信、化工、计算机、电气设备、轻工制造、机械设备等。在工业制造业中，上下游关系比较紧密，往往以产业链的形式参与在市场运转中，因此行业发生上涨时更易产生连锁反应带动上下游一齐上涨。

### 4. 有效市场内板块联动主要体现为同日上涨

在筛选过程中，表 5 中没有任何关联行业出现滞后上涨的现象，均表现为同日上涨，这也说明了我国股票市场是有效市场，信息能很快反映在股价上，不会出现滞后。若在数据集中剔除同日上涨的情况后，关联情况的置信度和支持度明显下降，行业之间的涨跌一般都在日内完成。

## 三、 基于改进的多元 AR 模型与 Markowitz 均值方差模型构建投资组合

### （一） 多元 AR 模型简述与改进

时间序列分析的最基本的理论基础是由 Norbort, Wiener 和 Aandrei. Kolmogorov 于上世纪 40 年代提出的，后经 G. P. Box 和 G. M. Jenkin 完善，提出一套完整的模型、估计、检验和预测的方法。

一元 AR(p) 模型中包含目标变量的 P 阶自回归

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \varepsilon_t$$

式中 c 为模型的常数项， $\Phi_i$  为模型中待估计的自回归系数， $\varepsilon_t$  为一个白噪声过程。

多元 AR(p) 模型是对一元模型的推广，加入其他自变量对目标变量的影响，同时也考虑因变量的滞后效应的影响

$$\begin{aligned} y_t = & c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} \\ & + \Phi_{11} x_{1,t-1} + \Phi_{12} x_{1,t-2} + \cdots + \Phi_{1p} x_{1,t-p} \\ & + \Phi_{21} x_{2,t-1} + \Phi_{22} x_{2,t-2} + \cdots + \Phi_{2p} x_{2,t-p} + \cdots \\ & + \Phi_{m1} x_{m,t-1} + \Phi_{m2} x_{m,t-2} + \cdots + \Phi_{mp} x_{m,t-p} + \varepsilon_t \end{aligned}$$

式中考虑了除自回归 y 外 m 个自变量的影响，即在行业轮动效应的分析中，我们除了考虑该板块自回归的影响，还考虑其他共二十七个板块的影响。

若对于所有行业均考虑滞后 5 期对因变量的影响，则可能产生过拟合的问题，并且待估参数过多，模型估计起来也有难度。因此我们通过前两问分析得出的结论对行业轮动效应的影响进行筛选，通过选择支持度大于 0.15 的各行业滞后影响，选出置信度最大的 10 个自变量，改进后的模型如下：

$$y_t = \sum_{i=1}^5 \alpha_i y_{t-i} + \sum_{i=1}^{10} \beta_i X_{m,t-n} + \varepsilon_t$$

式中， $\alpha_i$  为自回归的待估系数， $\beta_i$  为置信度最大的 10 个自变量的系数， $X_{m,t-n}$  分别为置信度最大的 10 个自变量，模型中不设截距项是考虑到行业收益率长期来看是 0 均值的关系。

对于模型的估计我们采用 OLS 估计方法，选用预测日期 t 前 15 天的历史数据估计出  $\alpha_i$  与  $\beta_i$  共计 15 个参数，以此预测某行业 t 期收益率。

## (二) Markowitz 均值-方差模型简述

上世纪 50 年代，H. Markowitz 提出用均值与方差来衡量证券的收益与风险，并以此建立了一系列关于投资组合的理论。Markowitz 认为在相同的风险水平下，具有最高的收益的投资组合即为有效的投资组合，并以此得到投资组合的有效边界。

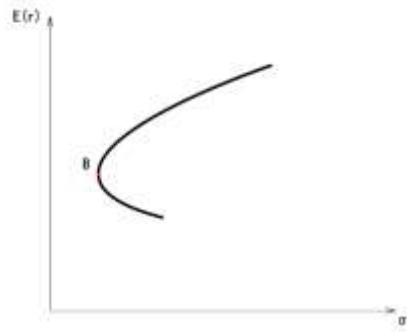


图 3 均值方差模型的前沿组合

如图 3 所示，实线为 Markowitz 均值方差模型下证券组合构成的前沿组合，为相同收益下方差最小的组合，B 点为所有组合里方差最小的组合，B 点以上即为有效投资组合。

在我们的模型中，我们假定投资者均为极度风险厌恶的，因此我们构建全局最小方差组合，即为 B 点。

根据均值方差模型我们可以计算出 B 点的期望收益  $E$

$$E = \frac{A}{C}, \quad A = e^T V^{-1} \mathbf{1}, \quad C = \mathbf{1}^T V^{-1} \mathbf{1}$$

式中  $e$  为 AR 模型估计出的预测各行业的收益率列向量， $V^{-1}$  为该年内方差协方差矩阵的逆， $\mathbf{1}$  为单位列向量。

再使得投资组合方差最小计算出各行业的权重  $w$

$$\begin{aligned} \min_w & w^T V^{-1} w \\ \text{s.t. } & w^T e = E, \quad w^T \mathbf{1} = 1 \end{aligned}$$

### （三） 样本数据回测结果分析

我们使用已有数据对 2019 年 9 月 11 日、12 日、16 日、17 日四个连续交易日进行了预测，首先我们由多元 AR 模型分别对每天各个行业进行回归并由此预测每天各个行业的收益率，然后又由均值-方差模型求出方差最小时各个行业在投资组合中的占比，将某一天各个收益率与其对应占比相乘，得到该日的投资组合收益率，如下表 7 所示

表 7 投资组合权重及回测收益率

日期	2019/09/11	2019/09/12	2019/09/16	2019/09/17
组合收益率	0.6955%	-2.8582%	-2.5510%	4.0812%
市场收益率	-0.4109%	0.7427%	-0.0162%	-1.7518%
残差	1.1065%	-3.6009%	-2.5348%	5.8330%
组合累计收益率	0.6955%	-2.1627%	-4.7137%	-0.6325%
市场累计收益率	-0.4109%	0.3318%	0.3156%	-1.4362%

行业 1 权重	0.118817545	3.79472E-07	0.000273134	0.061553582
行业 2 权重	4.57118E-05	1.66051E-06	2.39595E-05	2.34797E-05
行业 3 权重	2.26777E-05	5.56061E-07	1.4623E-05	1.13286E-05
行业 4 权重	3.76417E-05	1.5008E-06	2.1696E-05	1.49244E-05
行业 5 权重	2.79382E-05	1.98098E-06	3.39653E-05	1.35442E-05
行业 6 权重	1.17126E-05	3.5533E-07	3.58686E-05	3.46125E-06
行业 7 权重	2.65336E-05	9.1564E-07	6.55205E-06	6.7991E-06
行业 8 权重	2.03668E-05	8.29879E-07	0.213717558	8.08987E-06
行业 9 权重	0.048104938	3.15311E-06	4.88679E-05	2.65696E-05
行业 10 权重	2.58511E-05	2.33415E-06	1.08117E-05	1.2712E-05
行业 11 权重	3.61506E-05	7.80461E-07	3.06114E-06	1.90964E-05
行业 12 权重	0.212769786	0.050535563	0.452178145	0.000353094
行业 13 权重	3.92754E-05	1.05111E-06	7.09956E-05	2.06818E-05
行业 14 权重	2.28495E-05	6.61744E-07	1.94052E-05	1.55458E-05
行业 15 权重	3.05924E-05	0.164637755	1.80018E-05	1.31152E-05
行业 16 权重	4.1756E-05	4.08386E-06	1.10631E-05	1.82899E-05
行业 17 权重	1.77891E-05	1.17286E-06	7.04379E-06	3.62511E-06
行业 18 权重	1.90453E-05	6.88684E-07	1.0537E-05	5.11381E-06
行业 19 权重	3.76927E-05	1.30757E-06	1.65796E-05	1.30245E-05
行业 20 权重	2.47709E-05	1.11399E-06	6.81713E-06	0.224590094
行业 21 权重	3.91146E-05	0.151684369	0.00021138	0.080688415
行业 22 权重	1.1302E-05	1.31847E-06	0.002852105	4.95753E-06
行业 23 权重	1.36749E-05	9.90941E-07	2.97435E-06	3.3162E-06
行业 24 权重	1.50357E-05	1.89926E-06	1.5896E-05	3.56375E-06
行业 25 权重	0.619684547	0.633110344	0.330342239	0.632561377
行业 26 权重	8.52913E-06	3.26297E-07	5.90663E-06	2.23102E-06
行业 27 权重	2.46672E-05	2.0396E-06	1.03549E-05	6.0454E-06
行业 28 权重	2.25046E-05	8.68738E-07	3.04582E-05	3.92253E-06

由此可见我们所做投资组合的收益率有时会优于大盘行情，且这四日累积收益率高于大盘情况。

#### 四、 基于敏感性分析计算各行业板块的贡献程度

##### (一) 敏感性分析及贡献程度的定义

敏感性分析是定量分析有关因素发生变化对另一个相关因素的影响，在我们的模型中，不确定性因素主要是对于  $t$  期收益率的预测，我们通过使得不同行业的历史收益率增加一个微小的值，考察模型期望收益的变化，并以此计算出各行业对收益的贡献程度。

$$y_t = \sum_{i=1}^5 \alpha_i y_{t-i} + \sum_{i=1}^{10} \beta_i X_{m,t-n} + \varepsilon_t$$

例如我们考察计算机行业历史收益对于期望收益的影响。我们使计算机行业的前五天历史收益在  $i$  的基础上增加  $\varepsilon$ ，沿用式模型中估计的各项系数，计算出各行业新的预测收益率列向量  $e'_t$ ，再用均值方差模型计算出全

局最小方差点对应的投资组合期望收益率 $E'_t$ 。

$$\{i_m + \varepsilon\}_{m=t-5}^{t-1} \xrightarrow{\text{多元AR模型}} e'_t \xrightarrow{\text{均值方差模型}} E'_t$$

对于每个行业分别变动 $\varepsilon$ ，计算出各行业历史收益变化对应的 $\Delta E = E_t - E'_t$ ，用 $P_i = \frac{\Delta E_i}{\sum \Delta E}$ 表示收益中各行业的贡献程度。

## (二) 由历史数据分析各行业的贡献程度

在历史收益上增加 0.0001 考察各行业对投资组合的贡献程度，因模型中的投资组合仅是方差最小组合，因此轮动效应提升后投资组合的期望收益并不一定会增加，由此算出的贡献程度可能为负数，原因是该行业轮动效应对于我们要求的方差最小并没有好处，因此使得期望收益减小。

我们运用历史数据对各行业的贡献程度进行分析，选取了 2019 年 9 月 12、16、17 共四日的数据进行分析。每天选取贡献最高的 5 个行业与贡献最低的 5 个行业，可见行业对于投资组合的贡献程度并不稳定，12 日贡献最高的化工行业在 16 日贡献最低，而 16 日贡献较低的纺织服装行业在 17 日贡献较高。贡献程度的不同对投资组合的含义即为对于风险控制的贡献，贡献程度越高的行业在投资时更具有抵御风险的能力，而贡献程度低的行业可能存在更高的收益与更高的风险，因此对于方差最小的投资组合只有负的贡献。

贡献程度的不同与我们投资组合的设定有一定的关系，最小方差组合只是风险厌恶的极端情况，大部分投资者的风险偏好并不在这一点上。而通过敏感性分析得出的各行业贡献程度有利于我们选择每日风险更小的行业，从这一点上来说也是有意义的。

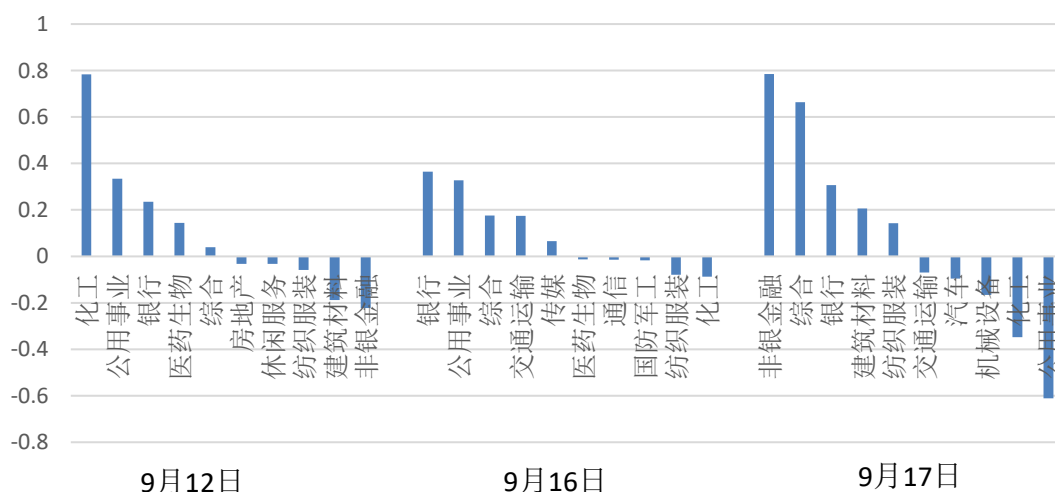


图 4 行业贡献程度

## 五、 电气设备行业指数编制——基于加权主营业务占比的改进

通过查阅申万一级行业指数的编制方法，其采用派式指数编制法指数，以在沪深证券交易所上市的 A 股自由流通市值为权重进行计算。本文采用电气设备行业，选择与申万相同的样本股，在派式编制法的基础上改进。

传统的计算公式为

$$I_t = \frac{LV_t}{LV_{t-1}^{Adj}} \times I_{t-1}$$

$I_t$  为 t 交易日指数

$LV_t$  为 t 交易日样本股的自由流通市值

$LV_{t-1}^{Adj}$  为发生事件例如新股发行、配股等事件时重新调整样本股的自由流通市值

$LV_t$  采用全样本加权

$$LV_t = \sum_{i=1}^n V_{i,t} = \sum_{i=1}^n P_{i,t} \times Q_{i,t}$$

$P_{i,t}$  为样本股 i 在 t 交易日的价格

$Q_{i,t}$  为样本股 i 在 t 交易日的流通股数量

申万指数的计算方法考虑了流通股直接影响样本股的价格变化，因此并没有使用全股本来计算企业市值，剔除了非流通股后能更直接地反映行业总市值的变化率。

但是在选择行业的样本股时，涉及相关行业的企业都有可能作为样本，每个企业主营业务收入占其总收入比例不等，对其利润贡献程度不一。而公司披露的利润往往对其股价有重要影响，但要区分主营业务利润及其他业务产生的利润。如电气设备行业的样本公司年末获得了正的利润，但其电气设备行业产生亏损，而投资于房地产行业带来收益，这一消息能够推动公司的市值上升，但是将其全部归入电气设备行业指数的上涨是不够严谨的。

因此，为了调整不同企业在主营行业内所获收益的贡献程度，通过计算其主营业务利润占比，作为权重调整其流通市值。

重新构造的行业内加权自由流通市值为

$$LV_t^p = \sum_{i=1}^n \omega_i V_{i,t} = \sum_{i=1}^n \omega_i \times P_{i,t} \times Q_{i,t}$$
$$\omega_i = \frac{E_{all} - E_{other}}{E_{all}}$$



$E_{all}$ 为企业总利润

$E_{other}$ 为企业主营业务外产生的利润

复制申万行业电气设备行业的 192 支样本股，从 CSMAR 数据平台获取其自 2019 年 7 月 1 日至 2019 年 10 月 11 日的日流通市值及其 2019 年 6 月 30 日之前发布的半年报损益表，获取总利润及主营业务外利润指标计算得到权重。以 2019 年 7 月 1 日的申万行业指数为新编制指数的基准点，绘出指数如图所示。

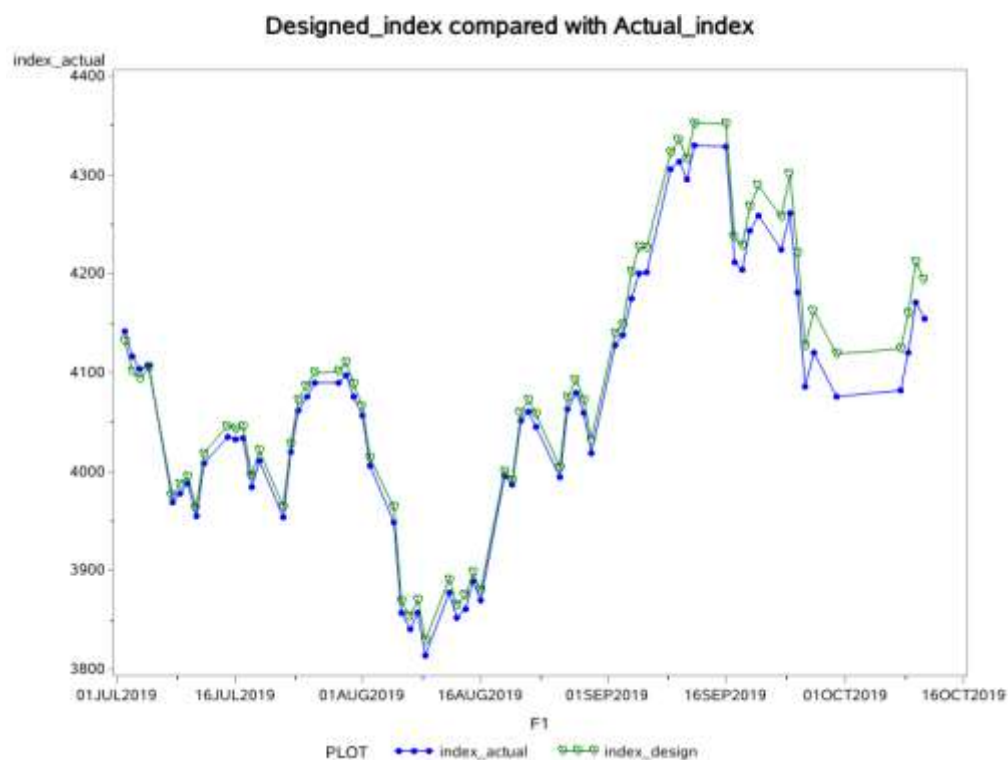


图 5 编制指数与申万指数对比

图 5 展示了加权后的新指数与申万行业指数的对比，从趋势上看两个指数保持一致，都能反映行业整体的变动情况。值得注意的是，2019 年 9 月 7 日以后，新编制的指数出现突增，此后均大于实际指数，这说明电气设备行业在这个区间内的表现优于实际申万指数所反映的增幅，即电气设备业务产生的利润增幅大于样本股总利润的增幅。

加权后编制的新指数相较于申万行业指数能够更好地反映行业本身的变动情况，剔除样本股在其他业务产生的收益对行业指数产生的影响。

但是，这种加权方法也有其局限性，用于计算企业主营业务利润的指标只能随季报更新，不能实时体现样本股的收益构成情况，如果企业在短期内业务结构发生重大改变，无法及时更新其主营业务的权重，造成指数偏离行业本身的变动情况。

# 附录 1 运行代码报告

## （一）参数设定及数据导入

```
/*为保证程序运行速度，将日志输出至外部文件
*/
proc printto log="&path.\log.log";
run;
/*设置文件所在路径！！！！，直接更改宏变量
path即可，其余无需变动*/
%let path = D:\SAS ;
libname data "&path." ;
option noxwait;
x md "&path.\var";
libname var "&path.\var";

proc import out = data.data /*原始股价
数据集*/
datafile = "&path.\申万二十八个行业收盘
价_NoHeader.xlsx"
dbms = excel replace;
getnames=no;
run;
proc import out = data.SZidx /*上证指
数*/
datafile = "&path.\上证指数.xlsx"
dbms = excel replace;
getnames=yes;
run;
proc import out = data.RF_return /*无
风险收益*/
datafile = "&path.\无风险收益.xlsx"
dbms = excel replace;
getnames=yes;
run;
```

data.data 数据集存放初始行业收盘价

	F1	F2	F3	F4
1	04JAN2000	1027.86	1032.97	1025.0
2	05JAN2000	1028.67	1095.08	1034.0
3	06JAN2000	1069.67	1122.06	1081.0
4	07JAN2000	1106.19	1184.13	1114.0
5	10JAN2000	1130.06	1219.1	1129.3
6	11JAN2000	1104.16	1198.66	1074.0
7	12JAN2000	1080.4	1180.74	1042.0
8	13JAN2000	1066.57	1157	1034.7
9	14JAN2000	1062.16	1138.67	1022.4
10	17JAN2000	1085.92	1169.02	1038.6

data.SZidx 数据集存放对应日期内上  
证指数

	TrdDt	CLPr
1	04JAN2000	1406.3
2	05JAN2000	1409.6
3	06JAN2000	1463.9
4	07JAN2000	1516.1
5	10JAN2000	1545.1
6	11JAN2000	1479.7

data.RF\_return 数据集存放对应日期  
内无风险收益率（按月更新）

	Date	DRfRet
1	04JAN2000	0.0000
2	05JAN2000	0.0000
3	06JAN2000	0.0000
4	07JAN2000	0.0000
5	10JAN2000	0.0000

## (二) 数据预处理——计算 alpha 收益率

```

/*计算对数收益率*/
data data.data_szidx;
merge data.data
data.szidx(rename=(TrdDt=F1
CLPr=F30));
by F1;
run;
data data.log_return;
set data.data_szidx;
array F{ * } _all_;
do i = 2 to 30;
    F(i) = log(F(i)) -
log(lag(F(i)));
end;
drop i;
if F2;
run;

/*计算行业与市场的协方差和相关系数，相除即为
beta值，252个观测为一组计算*/
/*计算各行业之间的协方差矩阵与逆，252个观测
为一组计算，存储在im1中，再次调用使用
load*/
%macro beta;
%do n=2 %to 29;
beta&n=F&n/F30;
%end;
%mend;
/*划分十九年数据为19个周期*/
%macro slice;
%do i=1 %to 18;
data Y&i;
set
data.log_return(firstobs=%eval(1+(&i-
1)*252) obs=%eval(&i*252));
run;
proc corr data=Y&i out=cov&i cov
noprint;
var _all_;
run;
data Y&i;
set cov&i(drop=_TYPE_ _NAME_ F1);
if _N_=30;
%beta;
drop F2-F30;
run;
data Y&i;
set Y&i;
do i =1 to 252;
output;
drop i;
end;
run;
data var.cov&i;
set cov&i(drop=_TYPE_ _NAME_ F1 F30);
if _N_>1;
if _N_<30;
n= n -1;

```

将上证指数、申万行业收盘价均转换为对数收益率

	F1	F2	F3	F4
1	05JAN2000	0.0011767396	0.0219363944	0.006770921
2	06JAN2000	0.0361425947	0.0407917391	0.0444029871
3	07JAN2000	0.0373179692	0.0539420461	0.030644607
4	10JAN2000	0.0213490501	0.0291049539	0.0123804831
5	11JAN2000	-0.023167753	-0.016908615	-0.049275071
6	12JAN2000	-0.021771635	-0.032146533	-0.029641941
7	13JAN2000	-0.012993463	-0.003227285	-0.007634391
8	14JAN2000	-0.004143323	-0.015969534	-0.011977581

计算出各行业与市场收益的相关系数及协方差（一年一更新），period 记录该观测所处的周期，252 个交易日为 1 个周期

beta27	beta28	beta29	period
1.0957357717	0.8722461055	0.9767176754	1
1.0957357717	0.8722461055	0.9767176754	1
1.0957357717	0.8722461055	0.9767176754	1
1.0957357717	0.8722461055	0.9767176754	1
1.0957357717	0.8722461055	0.9767176754	1
1.0957357717	0.8722461055	0.9767176754	1
1.0957357717	0.8722461055	0.9767176754	1
1.0957357717	0.8722461055	0.9767176754	1

```

data cov&i;
set var.cov&i(drop=n);
run;
proc iml;
use cov&i;
read all into cov&i;
invcov&i=inv(cov&i);
store invcov&i;
quit;
run;
%end;
%mend;
%slice;
/*单独处理19年数据*/
data Y19;
set data.log_return(firstobs=4537);
run;
proc corr data=Y19 out=cov19 cov
noprint;
var _all_;
run;
data Y19;
set cov19(drop=_TYPE_ _NAME_ F1);
if _N_=30;
%beta;
drop F2-F30;
run;
data Y19;
set Y19;
do i =1 to 251;
output;
drop i;
end;
run;
data var.cov19;
set cov19(drop=_TYPE_ _NAME_ F1 F30);
if _N_>1;
if _N_<30;
n=_n_-1;
run;
data cov19;
set var.cov19(drop=n);
run;
proc iml;
use cov19;
read all into cov19;
invcov19=inv(cov19);
store invcov19;
quit;
%macro period();
%do i=1 %to 19;
data Y&i;
set Y&i;
format period $5.;
informat period $5.;
period="&i";
%end;
%mend;
%period;
data data.beta;
set Y1-Y19;

```

```
run;
proc datasets library=work noprint;
delete Y1-Y19 cov1-cov19;
quit;
```

```
/*计算ALPHA收益*/
data return_beta;
merge data.log_return data.beta;
run;
data data.return_beta_rf;
merge return_beta
data.rf_return(rename=(Date=F1));
by F1;
if F2;
run;
data data.return_beta_rf;
set data.return_beta_rf;
n=_n_;
run;
%macro alpha;
%do i=2 %to 29;
alpha&i=F&i-DRfRet-beta&i*(F30-
DRfRet);
%end;
%mend;
data data.alpha(keep=alpha2-alpha29
period);
set data.return_beta_rf;
%alpha;
run;
```

```
/*alpha收益转化为逻辑值*/
%macro signal;
%do i=2 %to 29;
if alpha&i>0 then alpha&i=1;
else alpha&i=0;
%end;
%mend;
data data.signal;
set data.alpha;
%signal;
run;
```

解出超额收益率 Alpha

	period	alpha2	alpha3	alpha4
1	1	-0.000814415	0.0198182098	0.0083864259
2	1	0.0032796293	0.0268428794	0.0060811965
3	1	0.007503769	0.0230763029	-0.005212365
4	1	0.005632822	0.0123612657	-0.006719971
5	1	0.0132605314	0.0219120144	-0.005439042
6	1	0.0023635911	-0.006425084	-0.000595082
7	1	-0.004889282	0.0062946484	0.0019939595
8	1	0.0051299402	-0.006084631	-0.000810419

将 Alpha 收益率转换为 0-1 变量，收益率大于 0 为 1，否则为 0

	period	alpha2	alpha3	alpha4
1	1	0	1	1
2	1	1	1	1
3	1	1	1	0
4	1	1	1	0
5	1	1	1	0
6	1	1	0	0
7	1	0	1	1
8	1	1	0	0

### (三) 生成滞后 5 阶的变量标记

```

data data.sum_connect;
run;

/*分析不同行业的关联情况，并放置5阶滞后*/
/*生成代表不同滞后情况的变量插入数据集内*/
%macro varname(n); /*n为滞后阶数*/
%let num0 = %eval(2**&n.);
%let num1 = %eval(2**(&n.+1)-1);
%let num2 = %eval(&n.+1);
%put &num0;
%put &num1;
data varname;
array v{&num2} v1-v&num2.;
do i = &num0. to &num1.;
    q = i;
    do j = &num2 to 1 by -1;
        v(j)=mod(q , 2);
        q = floor(q / 2);
    end;
    output;
end;
drop i q j;
run;
data data.varname2;
set varname;
array v{&num2} v1-v&num2.;
array nv{&num2} nv1-nv&num2.;
format sv1-sv&num2. $1.;
informat var $50. var2 $50.;
format var $50. var2 $50.;
array sv{&num2} sv1-sv&num2.;
do i = 1 to &num2.;
    if v{i}=1 then nv{i}=i;
    else nv{i}=.;
end;
var = "v";
var2 = "v";
do i = 1 to &num2.;
    if nv{i} ~= . then
        sv{i} = put(nv{i}, $1.);
    else sv{i} = "";
    var = catx('% ',var,sv{i});
    var2=cats(var2,sv{i});
end;
var=cats(var,'% ');
keep sv1-sv&num2. var var2; /*var即
变量名*/
run;
%mend;
%varname(6);
/*生成用于%include插入滞后的语句*/
data data.varname;
set data.varname2(keep=var var2);
file"&path.\varname.txt";
a='%gen_var(';
c=",";
b=')';
put a $ var $ c $ var2 $ b $;

```

利用宏生成 64 个不同滞后项的组合

1	2	3
1	v1	v1
2	v2	v2
3	v3	v3
4	v4	v4
5	v5	v5
6	v6	v6
7	v7	v7
8	v8	v8
9	v9	v9
10	v10	v10

```
run;
data data.varname;
set data.varname2( keep=var2);
a='sum(';
b=') as ';
d=cats(a,var2,b);
f=cats(" ",d,var2);
run;
proc sql noprint;
select f
into :list_a separated by ','
from data.varname;
quit;
```

生成滞后项的64个变量v1、v12、v123等储存于宏变量list\_a中

#### (四) 基于关联规则行业间关联的数据挖掘

```
/*进行关联情况分析，ComA表示A行业，ComB表示B行业*/
%macro run_all(comA,comB);
data com&comA._&comB.(keep=11 12 13 14 15 16 17 period);
set data.signal;
l1=lag5(alpha&comA.);
l2=lag5(alpha&comB.);
l3=lag4(alpha&comB.);
l4=lag3(alpha&comB.);
l5=lag2(alpha&comB.);
l6=lag(alpha&comB.);
rename alpha&comB.=l7;
if l1^=. ;
run;
data com&comA._&comB._1;
set com&comA._&comB.;
format l11-l17 $1.;
if l1=1 then l11=1;
if l2=1 then l12=2;
if l3=1 then l13=3;
if l4=1 then l14=4;
if l5=1 then l15=5;
if l6=1 then l16=6;
if l7=1 then l17=7;
name=cats(l11,l12,l13,l14,l15,l16,l17);
keep name period;
run;

data com&comA._&comB.;
set com&comA._&comB._1;
no=_n_;
run;
%macro gen_var(name,name1);
data a;
set com&comA._&comB.;
where name like substr("&name.",2);
&name1.=1;
run;
```

将A行业与B行业滞后0-5阶的日交易数据横向合并

将0-1变量转换为数值1-7表示不同日上涨，如观测1中“23456”表明在t时刻，A行业没有上涨，B行业在t时刻及滞后1、2、3、4天后均有上涨  
出现1说明A行业t时刻上涨，出现2说明B行业t时刻上涨，出现3说明B行业t+1时刻上涨，出现4说明B行业t+2时刻上涨，以此类推

	period	name
1	1	23456
2	1	123457
3	1	12346
4	1	12357
5	1	12467
6	1	13567
7	1	24567

```

data com&comA._&comB.;
merge com&comA._&comB. a;
by no;
run;
%mend;
%include "&path.\varname.txt";
data com&comA._&comB._2;
set com&comA._&comB. (drop=no name);
run;
proc sql;
create table com&comA._&comB._3 as
select
&list_a, period
from com&comA._&comB._2
group by period;
quit;
data com&comA._&comB._2;
set com&comA._&comB._3;
lead=&comA.;
result=&comB.;
run;
data data.sum_connect;
set data.sum_connect;
com&comA._&comB._2;
run;
proc datasets library=work noprint;
delete com&comA._&comB._2
com&comA._&comB._3 com&comA._&comB._1
com&comA._&comB.;
quit;
%mend;

```

统计每一条观测中不同关联情况是否出现，如 v1=1 说明该观测中含有 1，不含有该关联则变量为空

1	23456	no	v1	v2	v3
1	123456	1	1	1	1
2	12345	1	1	1	1
3	1234	1	1	1	1
4	123	1	1	1	1
5	12	1	1	1	1
6	1	1	1	1	1
7	0	1	1	1	1
8	0	1	1	1	1
9	0	1	1	1	1
10	0	1	1	1	1

统计不同关联出现的次数保存在各变量中，period 记录出现的周期，lead 表示领涨行业序号，result 表示跟涨行业序号，如第一条观测表明在第 1 个周期内，假设第 2 行业为领涨行业，第 3 行业为跟涨行业，则出现“123456”即第 2 行业第 1 天上涨、第 3 行业在同日及随后 4 日内都上涨的次数为 6 次

v123456	v1234567	period	lead	result
6	2	1	2	2
5	3	10	2	2
6	3	11	2	2
4	2	12	2	2
		13	2	2
6	2	14	2	2
1		15	2	2

删除临时逻辑库的数据集释放空间

### （五）生成两两配对组合并运行宏

```

/*生成两两组合的序列数*/
data b;
do i=2 to 29;
do j=2 to 29;
output;
end;
end;
run;
data b;
set b;
if i<=j then output;
run;
data b2;
set b;
file"&path.\index_num.txt";
a='%run_all(';
c=",";
b=')';
put a $ i $ c $ j $ b $;
run;
%include "&path.\index_num.txt";

```

生成 28\*27 的两两组合配对并输出成宏格式运用%include 运行

VIEWTABLE: Work B2				
	i	j	a	b
1	2	3	%run_all(	)
2	2	4	%run_all(	)
3	2	5	%run_all(	)
4	2	6	%run_all(	)
5	2	7	%run_all(	)
6	2	8	%run_all(	)
7	2	9	%run_all(	)



## （六）通过最小置信度和最小支持度阈值筛选关联情况

```
%let given_sup=0.3; /*设置最小支持度*/
%let given_con=0.6; /*设置最小置信度*/
data data.varname;
set data.varname2(keep=var var2);
file "&path.\select.txt";
d='%select(';
b=')';
e=",&given_sup";
f=",&given_con";
put d $ var2 $ e $ f $ b $;
run;
data data.sum_judge_total;
run;
/*计算每一组关联情况的支持度和置信度，并筛选
满足阈值的关联*/
%macro
select(vname,given_sup,given_con);
data data.sum_judge;
set data.sum_connect (firstobs=2);
sup=&vname./247;
con=&vname./v1;
format connection $10.;
informat connection $10.;
connection=substr("&vname.",2);
run;
data data.sum_judge_total;
format connection $10.;
informat connection $10.;
set data.sum_judge_total
data.sum_judge;
run;
data data.sum_judge_total;
set data.sum_judge_total;
format connection $10.;
informat connection $10.;
where sup>=&given_sup and
con>=&given_con;
keep lead result sup con connection
period;
run;
%mend;
%include "&path.\select.txt";
data data.sum_judge_total;
set data.sum_judge_total;
if connection~=1;
run;

proc sort data=data.sum_judge_total
out=data.sum_judge_total;
by descending con;
quit;
```

运设置阈值后，筛选出关联情况，connection 下的“12”表示两个行业同日内上涨，对 con 即置信度降序排序找出置信度较高的关联情况和领涨、跟涨行业

Table 1: Data Sum Judge Total						
	connection	period	lead	result	sup	con
1	12	14	15		19	0.412954676 0.7909172
2	12	14	9		19	0.331805889 0.7907087
3	12	15	15		19	0.392712006 0.7026008
4	12	4	19		19	0.351748093 0.71962814
5	12	16	19		16	0.382712046 0.7026008
6	12	16	19		15	0.412954676 0.8935177
7	12	17	9		19	0.394917817 0.6770804

### (七) 对各行业日内收益率分别求回归方程

```
%macro base_dataset(sum,lr);
data sum_judge_total_2;
    set &sum;
run;

data Log_return;
    set &lr;
run;
%mend;

%base_dataset(data.sum_judge_total_2,
data.Log_return);

%macro split_Sum_judge_total_2; /* 各
年关联关系情况 */
    %do i = 1 %to 19;
        data t&i.;
            set Sum_judge_total_2;
            if period = &i.;
        run;
    %end;
%mend;

%macro F_lag5_all; /* 各行业0天到-5天的
数据 */
    %do i = 2 %to 29;
        data industry&i.;
            set
Log_return_for_reg(keep = F&i.);
            /* 此处为
&date_year.年对数收益率数据 */
            rename F&i. =
F&i._lag_0;

            %do j = 1 %to 5;
                F&i._lag_&j.
= lag&j.(F&i.);
            %end;
            if F&i._lag_5;
        run;
    %end;
%mend;

%macro merge_industries_lag05;
data merge_industries_lag05;
    set industry2;
run;
    %do i = 3 %to 29;
        data
merge_industries_lag05;
            merge
merge_industries_lag05 industry&i.;
        run;
    %end;
%mend;

%macro
reg_rawdata_nodrop;*(target_date);
%split_Sum_judge_total_2;
%F_lag5_all;
%merge_industries_lag05;
```

规定所使用关联规则和对数收益率的数据集。

```

%mend;

%macro dropF_0(indst);
    %do i = 2 %to %eval(&indst.-1);
        drop F&i._lag_0;
    %end;
    %do i = %eval(&indst.+1) %to
29;
        drop F&i._lag_0;
    %end;
%mend;

%macro sql_selectVar;
    %do i = 2 %to 28;
        %do j = 1 %to 5;

            (a.p_F&i._lag_&j.*b.F&i._lag_&j
.) as new_F&i._lag_&j.,
        %end;
    %end;
    %do j = 1 %to 4;

        (a.p_F29_lag_&j.*b.F29_lag_&j.)
as new_F29_lag_&j.,
    %end;
        a.p_F29_lag_5*b.F29_lag_5 as
new_F29_lag_5
%mend;

%macro get_predictdatax;
    %do i = 2 %to 29;
        %do j = 1 %to 5;
            new_F&i._lag_&j. =
lag&j.(F&i.);
        %end;
    %end;
    %do i = 2 %to 29;
        %do j = 1 %to 5;
            if new_F&i._lag_&j.
~= .;
        %end;
    %end;
    %do i = 2 %to 29;
        drop F&i.;
    %end;
%mend;

%macro reg_Findst(indst);
data reg_F&indst;
    set merge_industries_lag05;
    %dropF_0(&indst.);
run;
%mend;

%macro predict(target_date, indst);
data null;
    set Log_return;
    yr = year(F1);
    Format yr 4.;
    if F1 = "&target_date."d then
do;

```

回归解释变量在各行业 1-5 阶滞后中  
选择。

对 target\_date 目标号为 indst  
的行业进行回归

```

call symput('obs_', _n_);
call
symput('date_year', yr);
stop;
end;
run;

%put &obs_.;
%put &date_year.;

%let fstobs = %eval(&obs_. - 20);
data Log_return_for_reg;
set
log_return(firstobs=&fstobs.
obs=&obs_.);
run;
%reg_rawdata_nodrop;

%reg_Findst(&indst.);

%let datasetnm
= %sysfunc(compress(F&indst._suf01_&d
ate_year.));
%put &datasetnm.;

data &datasetnm.; /*某年某行业受哪些行业
影响*/
set sum_judge_total_2;
if result = "&indst.";
if period =
substr(%sysfunc(compress("&date_year.
")), 3);
/*put
substr(%sysfunc(compress("&date_year.
")), 3);*/
if length(connection)=2;
if suf>0.15;
if connection ~= '12';
if con>0.5;
p = 1;
run;

proc sort data = &datasetnm. out =
&datasetnm._sortbycon_suf;
by descending con descending
suf;
run;

data &datasetnm._maininds;
set
&datasetnm._sortbycon_suf(obs = 10);
run;

proc sort data =
&datasetnm._maininds;
by lead;
run;

data varnm;
do i = 2 to 29;
do j = 1 to 5;

```

对关联支持度大于 0.15、置信度前 10 的解释变量进行标记

```

varnames =
compress('p_F' || i || '_lag_' || j);
m + 1;
output;
end;
end;
keep varnames m;
run;
proc sort data = varnm; by varnames;
run;

data &datasetnm._forTrnsps;
set &datasetnm._maininds;
informat varnames $32.;
format varnames $32.;
varnames =
compress('p_F' || lead || '_lag_' || (subst
r(connection, 2, 2) - 2));
*n = _n_;
run;

proc sort data =
&datasetnm._forTrnsps; by varnames;
run;

data &datasetnm._forTrnsps_new;
merge &datasetnm._forTrnsps
varnm;
by varnames;
run;

proc sort data =
&datasetnm._forTrnsps_new; by m; run;

data &datasetnm._forTrnsps_new;
set &datasetnm._forTrnsps_new;
if p = . then p = 0 ;
if find(varnames, "_F&indst._")
then p = 1;
keep varnames p;
run;

proc transpose
data=&datasetnm._forTrnsps_new

out=&datasetnm._Trnsps;
ID varnames;
run;

data &datasetnm._Trnspsed;
set &datasetnm._Trnsps;
drop _name_;
run;

proc sql;
create table F&indst._reg_pre
as
select %sql_selectVar
from &datasetnm._trnspsed
as a, reg_F&indst. as b;
quit;

```

选择要作为回归中解释变量的变量

```

data F&indst._reg_pre_firstrow;
    set F&indst._reg_pre(obs = 1);
run;

proc transpose data =
F&indst._reg_pre_firstrow
    out =
F&indst._reg_pre_firstrow_trpsd;
run;

data F&indst._reg_pre_firstrow_trpsd;
    set
F&indst._reg_pre_firstrow_trpsd;
    if COL1 ~= 0;
    drop COL1;
    m + 1;
    call symput('xvar_num', m);
run;

%put &xvar_num.;

proc transpose data =
F&indst._reg_pre
    out = F&indst._reg_pre_trpsd;
run;

proc sql;
    create table
F&indst._reg_pre_trpsd_dropped as
    select * from
F&indst._reg_pre_trpsd a right join
F&indst._reg_pre_firstrow_trpsd b
        on a._NAME_ = b._NAME_;
quit;

proc sort
data=F&indst._reg_pre_trpsd_dropped;
by m; run;

data F&indst._reg_pre_trpsd_dropped;
set F&indst._reg_pre_trpsd_dropped;
drop m; run;

proc transpose data =
F&indst._reg_pre_trpsd_dropped out =
F&indst._reg_pre_dropped;
    ID _NAME_;
run;

data F&indst._reg_pre_dropped; set
F&indst._reg_pre_dropped; drop _NAME_;
run;

data F&indst._Y;
    set reg_f&indst.(keep =
F&indst._lag_0);
run;

data F&indst._regdata;
    merge F&indst._Y

```

被解释变量数据集，即所求回归行业  
不进行滞后的数据。

```

F&indst._reg_pre_dropped;
run;

proc sql;
    select _NAME_ into : xvar_nm
    separated by " "
    from
    F&indst._reg_pre_firstrow_trpsd;
quit;
%put &xvar_nm.;

proc reg data=F&indst._regdata
outest=F&indst._&target_date._para;
    model F&indst._lag_0 =
    &xvar_nm.
    / noint collinoint;
run;

data
F&indst._&target_date._para_dropped;
    set
F&indst._&target_date._para;
    drop MODEL _TYPE_ _DEPVAR_
    _RMSE_ F&indst._lag_0;
run;

data
F&indst._&target_date._predictdata;
    set Log_return(firstobs
= %eval(&obs_.-4) obs
= %eval(&obs_.+1));
    %get_predictdatax;
    drop F1;
    keep &xvar_nm.;
run;

data
F&indst._&target_date._predict_pre;
    set
F&indst._&target_date._predictdata
F&indst._&target_date._para_dropped;
run;

proc transpose data =
F&indst._&target_date._predict_pre
out =
F&indst._&target_date._predict_trpsd;
run;

data
F&indst._&target_date._predict_trpsd;
    set
F&indst._&target_date._predict_trpsd;
    COL3 = COL1*COL2;
    COL4=COL3;
    call symput("result", COL4);
run;
%put &result.;

data
pred_rslt F&indst._&target_date.;

```

回归方程参数。

例如对 11sep2019 的第十九个行业进行回归，回归结果如下，其中各变量为解释变量对应系数（无截距项）。

	new_F6_lag_2	new_F10_lag_2	new_F10_lag_3
1	-0.367176456	0.3067360919	-0.36922064
new_F10_lag_4	new_F11_lag_1	new_F12_lag_4	new_F17_lag_3
0.4205307805	-0.980476891	-0.758815259	0.174723138
new_F18_lag_2	new_F19_lag_1	new_F19_lag_2	
-1.252860082	0.2114345736	1.323428933	
new_F19_lag_3	new_F19_lag_4	new_F19_lag_5	
0.1083811931	0.2085623752	0.0477184874	
new_F19_lag_5	new_F23_lag_2	new_F29_lag_2	
0.0477184874	1.6915516282	-2.79943172	

```

date = "&target_date."d;
format date yymmdd10.;
pred_var = "F&indst.";
pred_rslt = &result.;

run;
%mend;

*%let target_date = 30AUG2017;
*%predict(&target_date., 5);

%macro all_indst_pred(target_date);
    %do indst = 2 %to 29;
        %predict(&target_date.,
&indst.);
    %end;
%mend;
*%all_indst_pred(30AUG2017);*&target_
date.);

%macro set_all_pred(target_date);
    %do i = 2 %to 29;

        pred_rslt_F&i._&target_date.
    %end;
%mend;
*%put %set_all_pred(&target_date);

%macro get_est_r(target_date);
/*得到某一天各行业的预期收益率 结果储存在
data.est_r_&target_date.数据集中*/
%all_indst_pred(&target_date.);
data pred_&target_date;
    set %set_all_pred(&target_date)
;
run;

data data.est_r_&target_date.;
    set pred_&target_date;
    rename pred_rslt = e_r;
    keep pred_rslt;

run;
/*
proc datasets library=work kill
nolist; *清除work中的数据集;
quit;
run;*/
%mend;

```

某日某行业的预测收益率  
其中行号为行业所对应次序，e\_r 为  
该行业所对应的预测收益率。

	e_r
1	-0.028006398
2	0.001034542
3	-0.011419514
4	-0.001002297
5	-0.01413889
6	0.0647953262
7	-0.042817377
8	0.0977015286

求出某日各行业的预测收益率

#### (八) 根据得出的各行业预测收益计算出权重

```

%macro get_weight(target_date);
data _null_;
    set data.return_beta_rf;
    if F1="&target_date"d;
        call
symput('period',input(period,3.));
run;

```

得到预测日期对应的期数



```

/*p1为optmodel过程调用, p2为iml过程调用
*/
data _null_;
    p1=cats('var.cov', &period);
    p2=cats('invcov', &period);
    call symputx("p1", p1);
    call symputx("p2", p2);
run;

proc iml;
    use data.est_r_&target_date.;
    read all into e;
    store e;
quit;

data rf;
    set data.rf_return;
    if Date="&target_date"d;
    drop Date;
run;

proc iml;
    use rf;
    read all into rf;
    store rf;
quit;

proc iml;
    load rf;
    load e;
    load &p2;
    one=j(28,1,1);
    store one;
    A=t(e)*&p2*one;
    B=t(e)*&p2*e;
    C=t(one)*&p2*one;
    D=B*C-A*A;
    /*E=A/C+D/(A*C-rf*C*C);*/
    E=A/C;
    create data.Er_&target_date.
from E;
    append from E;
quit;

proc optmodel printlevel=0;
    var w{1..28}>=0 <=1;
    set l=1..28;
    set c;
    set est_r;
    set e_er;
    number e{est_r};
    number er{e_er};
    number var{c,l} ;
    read data &p1 into c=[n] {i in
1} <var[n,i]=col("F"||i+1)>;
    read data
data.est_r_&target_date. into
est_r=[_n_] e=e_r;
    read data data.Er_&target_date.
into e_er=[_n_] er=coll;

```

将预测收益转化为矩阵的数据

得到无风险收益

通过模型得到投资组合的期望收益  
Er\_&target\_date.

非线性规划算出投资组合的各项权重  
w\_&target\_date.

```

min sigma2=sum{i in c,j in
l}var[i,j]*w[i]*w[j];
con one:sum{i in 1..28}w[i]=1;
con return:sum{i in
1..28}w[i]*e[i]=er[1];
solve with qp;
create data
data.w_&target_date. from [i] w=w;
quit;

data data.w_&target_date.;
set data.w_&target_date.;
drop i;

run;
%mend;

```

### (九) 敏感性分析及贡献程度的宏编程

```

%macro er(est_r,n_indu);/*date为预测日
的日期使用date7.的格式，est_r为预测各行业
的收益，n为变化的行业取2-29*/
data _null_;
set data.return_beta_rf;
if F1=&date;
call
symput('period',input(period,3.));
run;

/*p1为optmodel过程调用，p2为iml过程调用
*/
data _null_;
p1=cats('var.cov',&period);
p2=cats('invcov',&period);
call symputx("p1",p1);
call symputx("p2",p2);
run;
proc iml;
use &est_r;
read all into e;
store e;
quit;
data rf;
set data.rf_return;
if Date=&date;
drop Date;
run;
proc iml;
use rf;
read all into rf;
store rf;
quit;
proc iml;
load rf;
load e;
load &p2;
one=j(28,1,1);
store one;
A=t(e)*&p2*one;

```

宏 er 计算投资组合最小方差点期望收益，具体数据集内容同上

```

B=t(e)*p2*e;
C=t(one)*p2*one;
D=B*C-A*A;
E=A/C;
create Er&n_indu from E;
append from E;
quit;
data Er&n_indu;
set Er&n_indu;
rename coll=E&n_indu;
run;
%mend;

%macro sensibility;
%do j_indu=2 %to 29;
/*考虑行业n-1的敏感性*/
data data.new_lr;
set data.log_return;
F&j_indu=F&j_indu+0.0001;
run;

/*预测收益率的变化得出数据集est_r_j*/
%base_dataset(data.sum_judge_total_2,
data.new_lr);
%predict(&target_date., &j_indu.);
data data.new_est_r_&target_date.;
set
pred_rslt_F&j_indu._&target_date.(kee
p=pred_rslt rename=(pred_rslt=e_r));
F=%eval(&j_indu.);
run;

data data.est_r_&target_date.;
set data.est_r_&target_date.;
F=_n_+1;
run;

data data.new_est_r_&target_date.;
merge data.est_r_&target_date.
data.new_est_r_&target_date.;
by F;
drop F;
run;
/*算出行业变化后的Er&j*/
%er(data.new_est_r_&target_date.,&j_i
ndu);

%end;
%mend;

%macro delta_e;
%do i=2 %to 29;
dE&i=E&i-coll;
%end;
%mend;
%macro proportion;
%do i=2 %to 29;
pro&i=dE&i./sum;
%end;
%mend;

```

宏 sensibility 通过变动每一个行业计算出投资组合期望收益的变动，进而计算出贡献程度

数据集 new\_lr 为收益增加 0.0001 后的收益

重新估计参数得到新的行业收益 new\_est\_r

<pre> %macro contribution(d); /*设置预测日期*/ %let date="&amp;d"d; %<b>sensitivity</b>;  data data.delta_e_&amp;target_date.; merge data.er_&amp;target_date. er2-er29; %<b>delta_e</b>; sum=sum(of dE2-dE29); %<b>proportion</b>; keep pro2-pro29; run; proc datasets library=work noprint; delete er2-er29; quit; %<b>mend</b>; </pre>	<p>宏 contribution 计算贡献程度</p> <p>数据集 delta_e_&amp;target_date 包含各个行业的贡献程度，pro2-pro29 分别为各个行业的贡献程度</p>
---	--

(十) 计算投资组合中各行业占比、收益率、敏感性及贡献程度

<pre> /*获取大盘收益率*/ <b>data</b> data.Market_yield;     <b>set</b> data.Log_return(<b>keep</b> = F1 F30);     <b>informat</b> F1 date9.;     <b>format</b> F1 date9.;     date_F1 = put(F1, date9.); <b>run</b>;  %let target_date = 11SEP2019; %<b>base_dataset</b>(data.sum_judge_total_2, data.Log_return); %<b>get_est_r</b>(&amp;target_date.); %<b>get_weight</b>(&amp;target_date.); %<b>contribution</b>(&amp;target_date.); %let target_date = 12SEP2019; %<b>base_dataset</b>(data.sum_judge_total_2, data.Log_return); %<b>get_est_r</b>(&amp;target_date.); %<b>get_weight</b>(&amp;target_date.); %<b>contribution</b>(&amp;target_date.); %let target_date = 16SEP2019; %<b>base_dataset</b>(data.sum_judge_total_2, data.Log_return); %<b>get_est_r</b>(&amp;target_date.); %<b>get_weight</b>(&amp;target_date.); %<b>contribution</b>(&amp;target_date.); %let target_date = 17SEP2019; %<b>base_dataset</b>(data.sum_judge_total_2, data.Log_return); %<b>get_est_r</b>(&amp;target_date.); %<b>get_weight</b>(&amp;target_date.); %<b>contribution</b>(&amp;target_date.);  <b>data</b> data.pred_er;     <b>set</b>         data.Er_11sep2019         data.Er_12sep2019 </pre>	<p>获取大盘收益率</p> <p>各日预测投资组合中各行业占比、投资组合预期收益率、投资组合中各行业敏感性及其贡献程度</p>
---	--

```

data.Er_16sep2019
data.Er_17sep2019;
rename COL1 = pred_E_r;
run;

data data.real_er;
set data.Market_yield(firstobs
= 4771 obs = 4774);
rename F1 = date F30 =
market_yield;
keep F1 F30;
run;

data data.er;
merge data.pred_er
data.real_er;
epsilon = pred_E_r -
market_yield;
run;

data data.er;
set data.er;
sum_pred + pred_E_r;
sum_mkt + market_yield;
run;

data data.weight;
merge

data.W_11sep2019(rename=(W =
W_11sep2019))

data.W_12sep2019(rename=(W =
W_12sep2019))

data.W_16sep2019(rename=(W =
W_16sep2019))

data.W_17sep2019(rename=(W =
W_17sep2019));
run;

PROC EXPORT DATA= data.er
OUTFILE= "&path.\Er.xlsx"
DBMS=EXCEL REPLACE;
SHEET="Sheet1";
RUN;

PROC EXPORT DATA= data.weight
OUTFILE=
"&path.\weight.xlsx"
DBMS=EXCEL REPLACE;
SHEET="Sheet1";
RUN;

```

投资组合预期收益率与大盘收益率对比，行号为日期对应的次序，列名分别为预期投资组合收益率、日期、市场收益率、市场收益率与投资组合预期收益率之差、累积投资组合预期收益率、累积市场收益率

	pred_E_r	date	market_yield
1	0.0069951852	11SEP2019	-0.004109452
2	-0.029581999	12SEP2019	0.007427125
3	-0.025509764	16SEP2019	-0.000161663
4	0.0408120022	17SEP2019	-0.017517885
epsilon	sum_pred	sum_mkt	
0.0110646369	0.0069951852	-0.004109452	
-0.036009124	-0.021626813	0.0033176733	
-0.025348101	-0.047136578	0.0031560102	
0.0583298869	-0.006324576	-0.014361875	

投资组合各行业占比，行号为行业对应次序，列名表示权重所对应的日期。

	W_11sep2019	W_12sep2019
1	0.1188175446	3.7947157E-7
2	0.0000457118	1.6605077E-6
3	0.0000226777	5.5606143E-7
4	0.0000376417	1.5007972E-6
5	0.0000279382	1.9809805E-6
6	0.0000117126	3.5533003E-7
7	0.0000265336	9.156404E-7
8	0.0000203668	8.2987899E-7
9	0.0481049384	3.153109E-6
10	0.0000258511	2.3341522E-6
11	0.0000361506	7.8046074E-7
12	0.2127697861	0.0505355628
13	0.0000392754	1.0511055E-6
14	0.0000228495	6.617441E-7
15	0.0000305924	0.1646377547
16	0.000041756	4.0838633E-6
17	0.0000177891	1.1728553E-6
18	0.0000190453	6.886841E-7
19	0.0000376927	1.3075665E-6
20	0.0000247709	1.1139937E-6
21	0.0000391146	0.1516843686
22	0.000011302	1.3184669E-6
23	0.0000136749	9.909415E-7
24	0.0000150357	1.8992591E-6
25	0.6196845473	0.6331103444
26	8.5291295E-6	3.2629716E-7
27	0.0000246672	2.0395958E-6
28	0.0000225046	8.6873769E-7

## (十一) 电气设备行业指数编制

```

proc import out = data.dqsb_day /*日
流通数据*/
datafile = "&path.\TRD_Dalysr.xlsx"
dbms = excel replace;
getnames=yes;
proc import out = data.dqsb_profit /*
财务报表*/
datafile = "&path.\FS_Comins.xlsx"
dbms = excel replace;
getnames=yes;
run;
data data.index_design;
run;
data dqsb_day_weight;
run;
data dqsb_profit;
set data.dqsb_profit;
keep stkcd percentage;
run;
data dqsb_day_weight;
merge data.dqsb_day dqsb_profit;
by stkcd;
run;
data dqsb_day_weight;
set dqsb_day_weight;
dsmvosd_weight=dsmvosd*percentage;
where percentage~=.;
run;

proc sql;
create table q5_c as
select trddt,sum(dsmvosd) as sum from
data.dqsb_day
group by trddt;
quit;
data q5_d;
set q5_c;
dt = input(Trddt, yymmdd10.);
format dt date9.;
lag_dsmvosd=lag(sum);
change=sum/lag_dsmvosd-1;
if change~=.;
drop trddt;
run;

data q5_b;
set q5_d(rename=(dt=trddt));
retain index_design 4154.113;
index_design=index_design*(1+change);
run;
data data.index_design;
merge data.data(rename=(F1=trddt
F21=index_actual)) q5_b ;
by trddt;
keep trddt index_actual index_design;
if index_design~=.;
label index_actual=index_actual;
run;

```

加权后计算指数记为 index\_design

	F1	index_actual	index_design
1	02JUL2019	4141.715	4132.3882675
2	03JUL2019	4116.808	4101.9973976
3	04JUL2019	4103.818	4094.4795264
4	05JUL2019	4107.923	4105.0032113
5	06JUL2019	3969.473	3975.2918469
6	09JUL2019	3978.633	3987.7314318
7	10JUL2019	3987.721	3994.2441031
8	11JUL2019	3955.763	3963.4738383
9	12JUL2019	4000.747	4010.267727
10	15JUL2019	4035.652	4046.1203748
11	16JUL2019	4032.609	4042.8162853
12	17JUL2019	4034.464	4045.8982486
13	18JUL2019	3984.529	3995.8132032

```
AXIS1 MINOR=(COLOR=BLUE NUMBER=1);  
AXIS2 MINOR=(COLOR=BLUE HEIGHT=0.25  
NUMBER=1);  
SYMBOL1 INTERPOL=JOIN VALUE=DOT  
C=BLUE;  
SYMBOL2 INTERPOL=JOIN VALUE=#  
C=GREEN;  
PROC GPLOT DATA=data.index_design;  
    TITLE 'Designed_index compared  
with Actual_index';  
    PLOT index_actual*trddt  
index_design*trddt/OVERLAY LEGEND  
HAXIS=AXIS1 VAXIS=AXIS2 ;  
RUN;  
QUIT;
```

## 附录 2 数据集说明

数据集名称 (data 逻辑库内)	说明
Alpha	包含计算得出的各行业 Alpha 收益, Alpha2-Alpha29 分别对应申万行业中第 1-28 行业的 Alpha 收益
Beta	包含计算得出的各行业 Beta 值, Beta2-Beta29 分别对应申万行业中第 1-28 行业的 Beta 值
Data	变量 F2-F29 储存了 28 个申万行业的日收盘价
Data_szidx	申万行业与上证指数收盘价的横向合并表
Dqsb_day	电气设备行业的日流通市值, stkcd 为股票代码, dsmvosd 为日流通市值
Dqsb_profit	电气设备行业 2019 年 6 月 30 日之前发布二季报的利润表, percentage 为主营业务利润比例
Index_design	Index_actual 为电器设备行业的申万指数, index_design 为加权后编制的新指数
Log_return	包含各个行业和上证指数的对数收益率
Return_beta_rf	包含各个行业的 Beta 收益
Rf_return	19 个周期的无风险收益, 一年更新一次
Signal	转换为 0-1 变量的 alpha 收益
Sum_connect	存放 28 个两两组合的行业的关联情况
Sum_judge	储存待筛选关联情况的临时数据集
Sum_judge_total	满足最小阈值的关联情况, connect 反应同涨情况, period 为周期数, lead 为领涨行业, result 为跟涨行业, sup 为支持度, con 为置信度
Szidx	储存上证指数日收盘价
Varname	通过滞后变量合并控制语句生成宏命令行
Varname2	储存 64 个 5 阶滞后变量
Est_r_&target_date	储存模型估计的各行业收益率
Er_&target_date	储存均值方差模型下投资组合的期望收益
W_&target_date	储存投资组合中各行业的权重
Er	2019/9/11、2019/9/12、2019/9/16、2019/9/17 各日投资组合预期收益率以及大盘收益率对比
Weight	2019/9/11、2019/9/12、2019/9/16、2019/9/17 各日投资组合中各行业所占权重



## 参考文献

- [1]孙守坤. 基于沪深 300 的量化选股模型实证分析[D]. 复旦大学, 2013.
- [2]武巍巍, 龚玉晶. 中国股票市场板块轮动的机理研究[J]. 现代商贸工业, 2010, 22(11):197-198.
- [3]陈庆伟, 程刚. 我国股市板块效应研究[J]. 金融经济, 2010(10):110-111.
- [4]张福芬. 中国股票市场板块轮动的机理研究[J]. 科协论坛(下半月), 2010(04):139-141.
- [5]叶咸尚. 中国股市板块轮动: 特征、影响因素与趋势[J]. 浙江金融, 2007(08):42+41.
- [6]燕吉吉. 关联规则在中国股市行业轮动中的应用[D]. 青岛大学, 2018.
- [7]彭惠, 刘欣雨. 基于关联规则的中国股票市场行业轮动现象研究[J]. 北京邮电大学学报(社会科学版), 2016, 18(01):66-71.
- [8]刘欣雨. 基于关联规则的中国 A 股市场行业轮动现象研究[D]. 北京邮电大学, 2016.
- [9]叶银龙. 基于关联规则挖掘的股票板块指数联动分析[J]. 统计教育, 2008(09):56-58.
- [10] Harry Markowitz. Portfolio Selection [J]. Journal of Finance, 1952: Vol 7, pp 77-91