深度报告

金融工程

投资性产品

投资性指标与策略系列之二

2010 年 9 月 7 日

专题报告

本系列报告独到之处

- **提出运用 CART 决策树模型处理行业内股票分类与选股问题。对传统决策树模型进行修正，引入事前修剪、事后修剪以及过滤算法，得到修正动态 CART 模型，并应用在国内各个行业中，得到较好的检验效果。**

CART 决策树在制造业中的选股效果

1. 传统六因子模型在制造业中的选股效果

行业数据来源：证监会行业分类中的制造业，其中剔除成份股数量过少的子行业，最终包括食品饮料，纺织服装，石油化工，电子，金属非金属，机械设备和医药生物等七个子行业。

因子选择：将经典文献的六个因子稍加改造，确定为资产负债率，EPS 一致预期变化率，流通市值，总资产净利率变化率，市盈率与股票收益率。

选股效果：从选股的累积收益来看，纺织服装是决策树样本外检验唯一失败的行业，此外金属非金属效果也一般，其余行业效果较好。模型在样本外出现了几个失效的阶段，我们认为市场持续下跌，市场突然反弹，股指期货上市等因素可以做出部分解释。

	占比情况		月度收益率效果										
行业	多头	空头	1期	2期	3期	4期	5期	6期	7期	8期	9期	10期	累积收益效果
食品饮料	高	低	好	中	好	好	中	好	好	好	好	差	**好**
纺织服装	低	低	好	中	好	差	中	差	差	差	差	差	差
石油化工	中	低	好	好	好	差	好	好	中	差	差	好	**好**
电子仪器	低	低	好	好	差	差	好	好	好	差	差	差	**好**
金属非金属	中	低	好	好	差	差	好	好	差	差	差	好	中
机械设备	中	低	好	好	中	差	好	好	差	差	差	好	**好**
医药生物	高	低	好	好	好	好	中	好	中	中	差	差	**好**

2. 因子加工后的决策树模型选股效果

因子选择：使用逐步回归的方法，分行业对以上六因子进行优选，形成每个行业的显著因子，将显著因子作为决策树模型的输入。

选股效果：在因子优化后，从选股的累积收益来看，纺织服装，电子和金属非金属中模型均失效。我们认为失效的主要原因是因子减少后，可以挖掘的数据过少，导致我们被迫不去对决策树进行任何加工，从而降低模型准确度。在接下来的研究中，我们会大幅增加可选因子数量，将因子选择作为新的研究重点。

	占比情况		月度收益率效果										
行业	多头	空头	1期	2期	3期	4期	5期	6期	7期	8期	9期	10期	累积收益效果
食品饮料	低	低	好	中	好	好	中	好	中	差	中	好	**好**
纺织服装	低	低	中	中	中	差	好	差	好	中	差	好	差
石油化工	低	低	好	好	好	差	中	好	好	差	差	好	**好**
电子仪器	低	低	好	好	差	差	好	好	好	差	差	差	差
金属非金属	低	低	中	中	差	差	差	差	差	差	差	中	差
机械设备	低	低	好	好	中	差	中	差	差	中	差	好	**好**
医药生物	低	低	中	好	好	好	中	差	好	好	中	中	**好**

投资性指标与策略系列相关报告：

《基于 CART 决策树的选股方法》

联系人：赵学昂
电话：010-82254206
E-mail：zhaoxang@guosen.com.cn
分析师：焦健
电话：0755-82130833-6220
E-mail：Jiaojian1@guosen.com.cn
SAC 执业证书编号：S0980210040012

内容目录

图表目录

传统六因子模型在制造业中的选股效果

本报告中，我们将 CART 决策树从国内的科技股板块转移到正规的行业分类上来。建于制造业是我国股市中子行业最多，体系最复杂的行业大类，因此我们检验了决策树模型在该类股票中的选股效果。在此，我们选择了证监会行业分类体系。

行业数据来源及处理

按照证监会分类体系，制造业包含了10个子行业，由下表给出：

表 1：证监会行业分类－制造业，2010-08-22

制造业	成份股个数	成交金额（亿元）	总市值（亿元）
	1184	**1310.87**	**84363.33**
食品、饮料	74	86.06	9166.79
纺织、服装、皮毛	78	67.62	3095.28
木材、家具	8	6.74	265.47
造纸、印刷	39	23.06	1407.6
石油、化学、塑胶、塑料	210	198.68	10732.86
电子	102	115.93	5664.51
金属、非金属	171	210.81	16769.63
机械、设备、仪表	346	418.78	26674.98
医药、生物制品	125	155.85	9296.87
其他制造业	31	27.33	1289.34

数据来源： Wind，国信证券经济研究所

由于决策树模型对样本数量有一定的要求，所以我们没有考虑样本数量过少的子行业，包括木材家具，造纸印刷与其他制造业。这样，我们在其余的7个子行业中应用决策树模型选股。由于我们的因子中包括一致预期数据，因次样本数据的采样期被限制在2005年2月至2010年7月，共66个月。我们将前56个月数据做为样本内数据建立静态树，然后接下来的10个月数据为样本外数据，用其交替建立动态树并同时做选股预测。

因子选择

在首篇报告中，我们应用了 Eric H, Keith L 和 Chee K 在 “The Decision Tree Approach to Stock Selection - An evolving tree model performs the best”（2000）一文选取的 6 个因子来进行决策树在 A 股科技股板块的实证，但在数据的提取上做了细微的修改。这 6 个因子来源于估值、基本面营利性、一致预期与价格动量这四个方面，且主要为美国投资经理的经验选择。

表 2：决策树指标选取（1）

	论文指标	首篇报告指标
Sales-Price	市销率倒数	最近12个月市销率倒数
CashFlow-Price	市现率倒数	最近12个月市现率倒数
EPS-Price	未来12月一致预期EPS比股价	最近12个月市盈率倒数
ROA	ROA变化率	ROA年同比变化率
EPS-MOM	EPS一致预期12周变化	净利润一致预期12周变化
Price-MOM	前一月股票收益率	前一月股票收益率

数据来源： Eric H, Keith L 和 Chee K（2000），国信证券经济研究所

在本篇报告中，我们在这些因子的基础上做了调整，引入总资产负债率（debt to asset）与市值（market value）因子，并去掉了 PC 与 PS 倒数这两个因子。一方面，之前选取了三个估值因子，它们的共线性太强；另一方面，加入债务结构和市场规模，使因子的来源更具多样性。之所以我们在此没有将因子数量增多，主要原因还是基于决策树模型的经典理论：过多的因子将让树形结构的稳定性变差，进而危害选股效果。在未来的研究中，我们也会灵活控制因子的数量，对该理论进行深入的验证。

表 3：决策树指标选取（2）

本报告指标		含义
DTA	资产负债率	负债总额／资产总额*100%
FEPS	EPS一致预期变化率	截止到月底对该年EPS的预测/截止到上月底对该年EPS的预测-1
MV	流通市值	月底A股流通市值
ROA	总资产净利率变化率	（本期的ROA/去年同期的ROA-1）*100%
PE	市盈率	(月底收盘价*汇率*总股本)／净利润（归属于母公司股东）
RETURN	股票收益率	（月底的复权价/月初的复权价-1）*100%

数据来源： Wind，国信证券经济研究所

在指标数据的提取上，由于财报公布时间相对滞后，因此不可避免的出现了因子落后于市场的情况。因此，我们在下一步会加入更多的市场因子。本报告指标数据的财报选取规则如下表所示。

表 4：上市公司财报公布时间

财报分类	公布截止期
年报	4月30日
一季度报	5月31日
半年报	8月31日
三季度报	10月30日

数据来源： Wind，国信证券经济研究所

表 5：月度样本数据应用的财报

月度时间	财报分类	月度时间	财报分类
1月	上年三季度报	7月	当年一季度报
2月	上年三季度报	8月	当年半年报
3月	上年三季度报	9月	当年半年报
4月	上年年报	10月	当年三季度报
5月	当年一季度报	11月	当年三季度报
6月	当年一季度报	12月	当年三季度报

数据来源： Wind，国信证券经济研究所

在 2009 年 10 月开始的样本外检验中，我们对每个行业分别给出每月的选股数量，该数量相对行业内成份股数量的占比，多头组合、空头组合与行业平均月度收益率的对比，样本外各组合的累积收益，以及一个样本外月度决策树模型的实例（为展示清楚，实例图形修剪次数根据情况改变，但模型修剪次数及阈值唯一，分别为“25”与“10”，以避免参数敏感性问题。经过测算，在六个因子下，修剪 15－30 次，结果较稳定）。这里需要强调，收益率均为等权重计算。在本报告中，我们继续进行 CART 决策树方法的实证，并注意因子选择的影响，因此在收益率计算和成交费用上，未进行深入考虑。

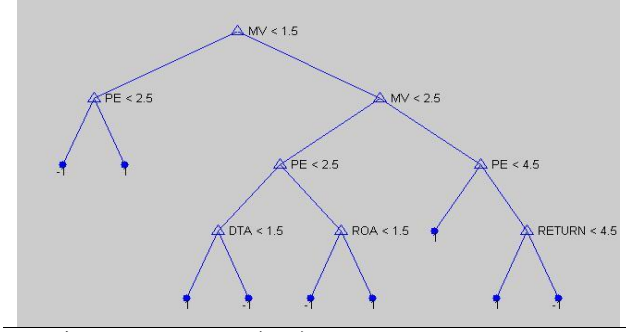
食品饮料行业

表 6：食品饮料样本外选股数量（修剪 25 次，阈值＝10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	25	15	34.72%	20.83%
2009-11	31	11	43.06%	15.28%
2009-12	41	13	56.94%	18.06%
2010-01	43	14	59.72%	19.44%
2010-02	47	10	65.28%	13.89%
2010-03	37	10	51.39%	13.89%
2010-04	26	9	36.11%	12.50%
2010-05	41	7	56.94%	9.72%
2010-06	13	9	18.06%	12.50%
2010-07	25	11	34.72%	15.28%

数据来源： Wind，国信证券经济研究所

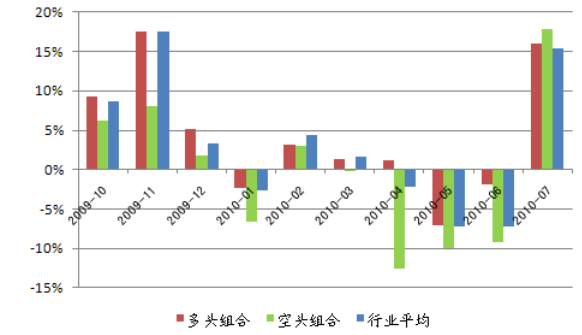
图 1：食品饮料行业最近一期动态 DT 模型（修剪 30 次）



数据来源： Wind，国信证券经济研究所

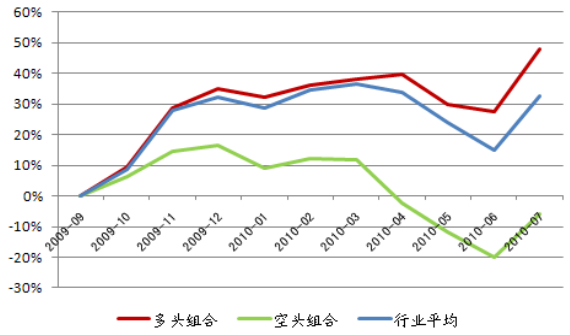
观察样本外每月的多空股票数量，发现多头数量占比行业股票数量较高，最高达到了 65.28%。这在实际的操作中会出现换仓困难，交易费用高昂的问题；空头数量相对很低。

图 2：食品饮料行业样本外组合月度收益率对比



数据来源： Wind，国信证券经济研究所

图 3：食品饮料行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

样本外期初多头与行业平均明显战胜空头，但前两者差距不大。中期多头趋弱，但在 4 月后多头在下跌市场中表现出色。在 7 月市场中，空头突然走强，跑赢了多头和行业平均。从累积收益来看，多头与行业平均差距不大，但远远跑赢空头。

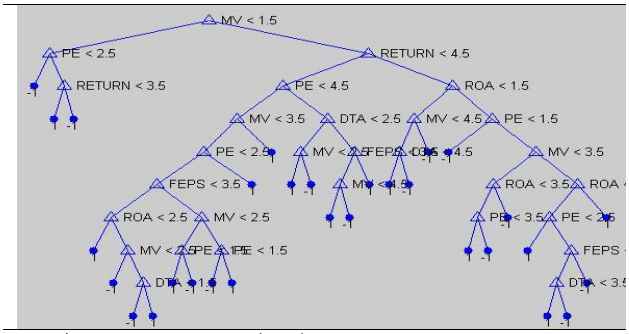
纺织服装行业

表 7：纺织服装样本外选股数量（修剪 25 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	17	15	25.37%	22.39%
2009-11	3	14	4.48%	20.90%
2009-12	5	11	7.46%	16.42%
2010-01	11	12	16.42%	17.91%
2010-02	2	18	2.99%	26.87%
2010-03	7	14	10.45%	20.90%
2010-04	13	13	19.40%	19.40%
2010-05	17	11	25.37%	16.42%
2010-06	19	6	28.36%	8.96%
2010-07	16	8	23.88%	11.94%

数据来源： Wind，国信证券经济研究所

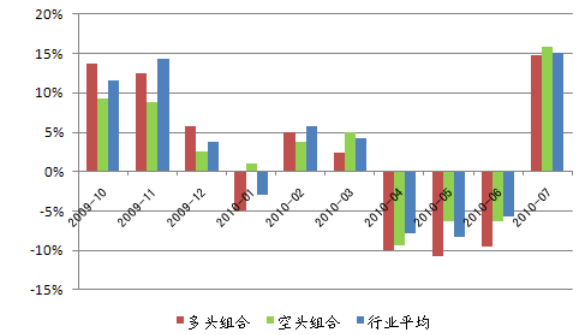
图 4：纺织服装行业最近一期动态 DT 模型（修剪 30 次）



数据来源： Wind，国信证券经济研究所

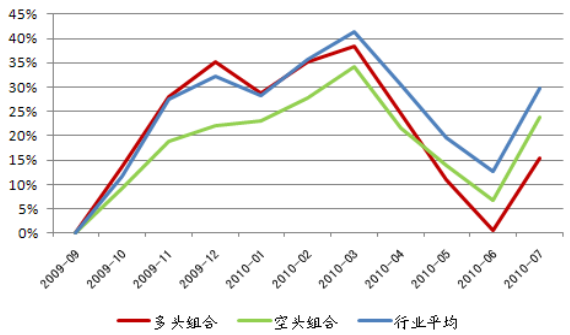
纺织服装行业各组合的月度收益率对比显示，决策树模型出现了严重的问题，多头组合鲜有跑赢行业平均的时候，且多次跑输空头，长期累积收益最低。

图 5：纺织服装行业样本外组合月度收益率对比



数据来源： Wind，国信证券经济研究所

图 6：纺织服装行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

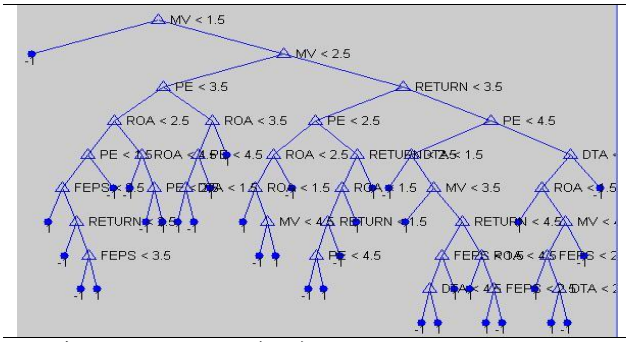
石油化工行业

表 8：石化行业样本外选股数量（修剪 25 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	44	48	24.44%	26.67%
2009-11	49	39	27.22%	21.67%
2009-12	98	40	54.44%	22.22%
2010-01	80	37	44.44%	20.56%
2010-02	72	41	40.00%	22.78%
2010-03	67	34	37.22%	18.89%
2010-04	78	33	43.33%	18.33%
2010-05	61	34	33.89%	18.89%
2010-06	31	37	17.22%	20.56%
2010-07	32	48	17.78%	26.67%

数据来源： Wind，国信证券经济研究所

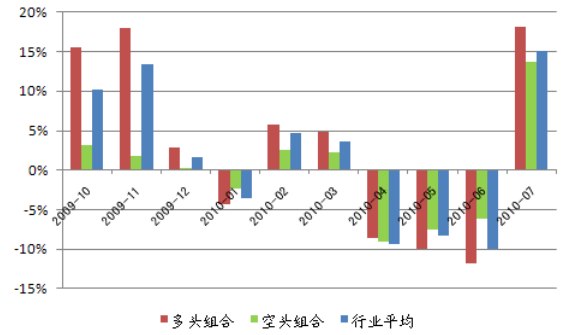
图 7：石油化工行业最近一期动态 DT 模型（修剪 40 次）



数据来源： Wind，国信证券经济研究所

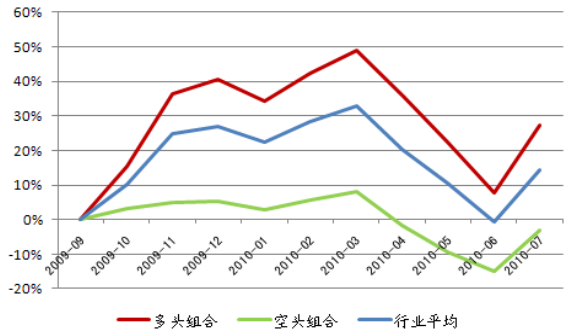
该行业多头占比变动较大，但选股效果不错。除去市场个别下跌月份之外，三个组合的收益率达到了我们的要求。

图 8：石油化工行业样本外组合月度收益率对比



数据来源： Wind，国信证券经济研究所

图 9：石油化工行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

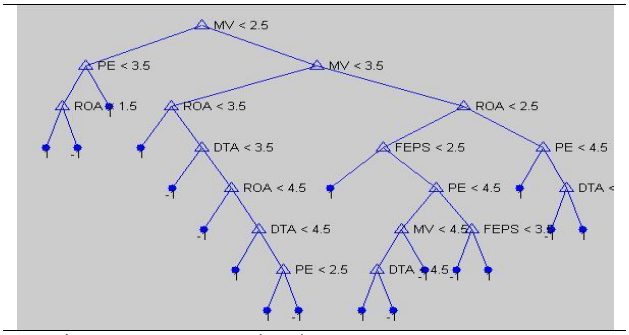
电子仪器行业

表 9：电子行业样本外选股数量（修剪 25 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	10	30	10.87%	32.61%
2009-11	6	24	6.52%	26.09%
2009-12	16	25	17.39%	27.17%
2010-01	11	25	11.96%	27.17%
2010-02	12	23	13.04%	25.00%
2010-03	25	22	27.17%	23.91%
2010-04	27	22	29.35%	23.91%
2010-05	22	21	23.91%	22.83%
2010-06	22	30	23.91%	32.61%
2010-07	20	24	21.74%	26.09%

数据来源： Wind，国信证券经济研究所

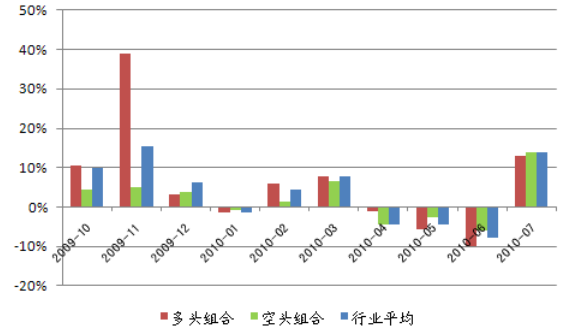
图 10：电子行业最近一期动态 DT 模型（修剪 30 次）



数据来源： Wind，国信证券经济研究所

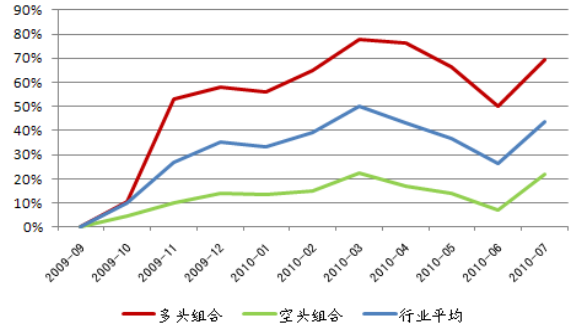
该行业在 2009 年 11 月时，多头有一个非常规的超额收益，且在中期多头表现较好，因此累积收益优秀。

图 11：电子行业样本外组合月度收益率对比



数据来源： Wind，国信证券经济研究所

图 12：电子行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

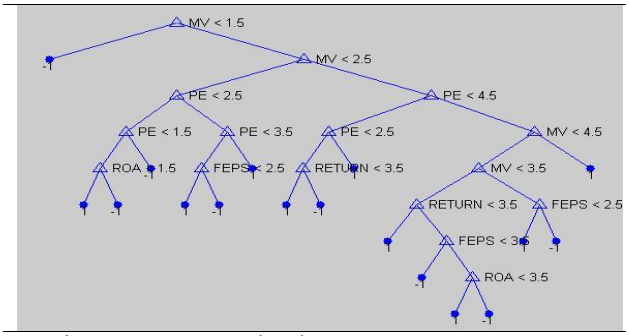
金属非金属行业

表 10：金属非金属样本外选股数量（修剪 25 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	42	37	27.63%	24.34%
2009-11	65	37	42.76%	24.34%
2009-12	63	36	41.45%	23.68%
2010-01	51	32	33.55%	21.05%
2010-02	47	24	30.92%	15.79%
2010-03	67	27	44.08%	17.76%
2010-04	82	28	53.95%	18.42%
2010-05	60	23	39.47%	15.13%
2010-06	36	22	23.68%	14.47%
2010-07	30	35	19.74%	23.03%

数据来源： Wind，国信证券经济研究所

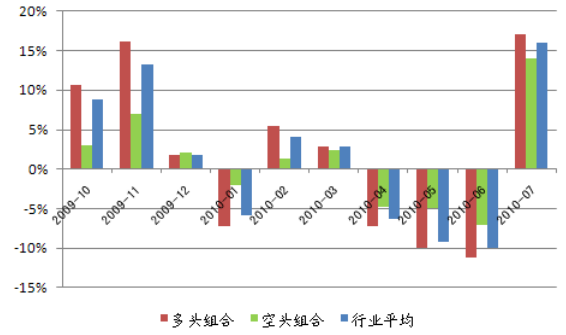
图 13：金属非金属行业最近一期动态 DT 模型（修剪 40 次）



数据来源： Wind，国信证券经济研究所

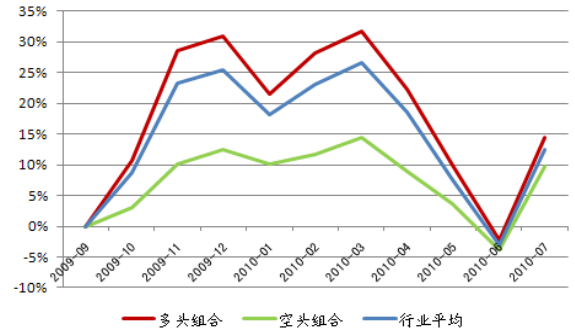
多头组合在前面累积的超额收益，在 2010 年 4 月至 6 月的下跌中基本消失殆尽，空头在这段时间内一直表现最强，这也是我们所有行业样本外检验的共同问题。该行业的多头占比较高。

图 14：金属非金属行业样本外组合月度收益率对比



数据来源： Wind，国信证券经济研究所

图 15：金属非金属行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

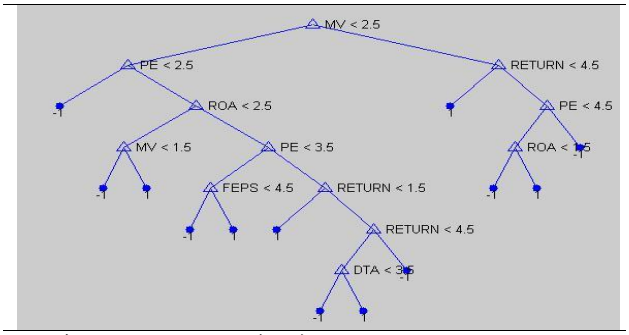
机械设备行业

表 11：机械设备样本外选股数量（修剪 25 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	101	78	32.17%	24.84%
2009-11	94	71	29.94%	22.61%
2009-12	97	69	30.89%	21.97%
2010-01	92	61	29.30%	19.43%
2010-02	83	51	26.43%	16.24%
2010-03	105	58	33.44%	18.47%
2010-04	99	43	31.53%	13.69%
2010-05	84	41	26.75%	13.06%
2010-06	61	38	19.43%	12.10%
2010-07	60	51	19.11%	16.24%

数据来源： Wind，国信证券经济研究所

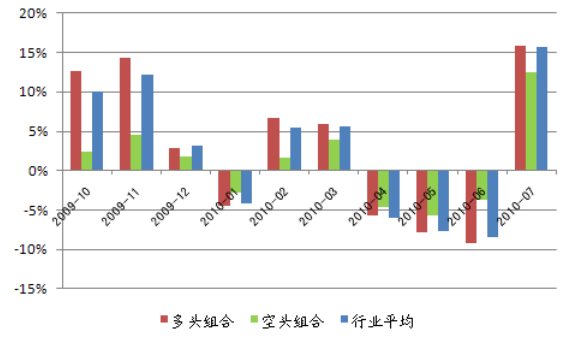
图 16：机械设备行业最近一期动态 DT 模型（修剪 60 次）



数据来源： Wind，国信证券经济研究所

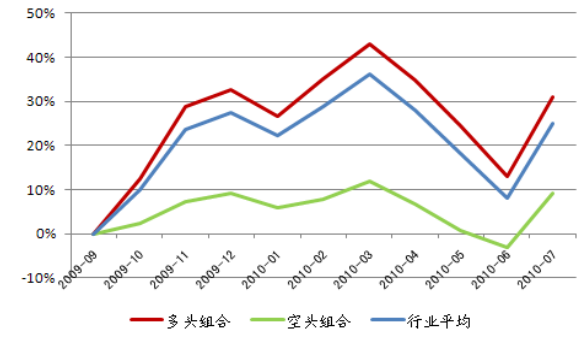
该行业选股效果相对较好，但多头没有与行业平均明显拉开差距。多空头的占比相对稳定，但多头数量偶尔会出现极值。由于成份股数量众多，该行业的决策树模型生长的极为复杂，我们认为修剪次数限制在 25 次，降低了模型选股的准确性。

图 17：机械设备行业样本外组合月度收益率对比



数据来源： Wind，国信证券经济研究所

图 18：机械设备行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

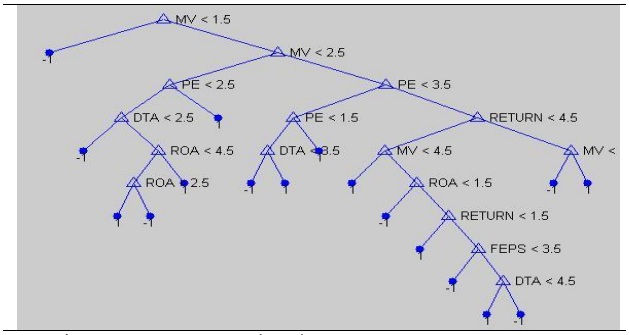
医药生物行业

表 12：医药生物样本外选股数量（修剪 25 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	33	26	28.21%	22.22%
2009-11	36	18	30.77%	15.38%
2009-12	37	19	31.62%	16.24%
2010-01	42	16	35.90%	13.68%
2010-02	52	14	44.44%	11.97%
2010-03	53	14	45.30%	11.97%
2010-04	53	15	45.30%	12.82%
2010-05	58	10	49.57%	8.55%
2010-06	50	9	42.74%	7.69%
2010-07	32	3	27.35%	2.56%

数据来源： Wind，国信证券经济研究所

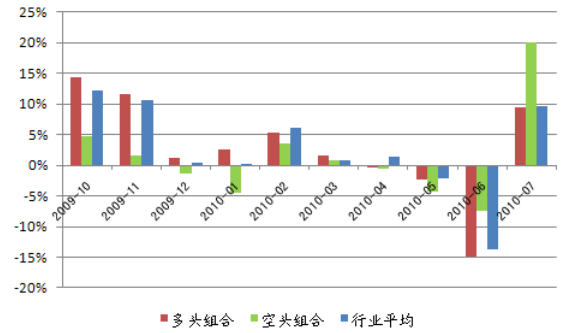
图 19：医药生物行业最近一期动态 DT 模型（修剪 40 次）



数据来源： Wind，国信证券经济研究所

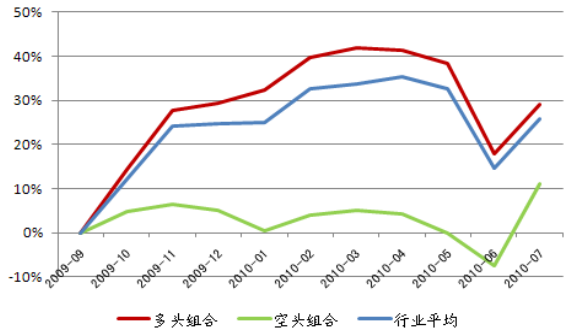
该行业多头数量明显超过空头，甚至个别月度近 6 倍。样本外最后一个月度的空头跑赢多头与行业平均极多，这严重的抵消了前面模型对空头组合精准的判断。此外，该行业多头与行业平均的区分也不明显。

图 20：医药生物行业样本外组合月度收益率对比



数据来源： Wind，国信证券经济研究所

图 21：医药生物行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

六因子模型选股效果小结

从选股的累积收益来看，纺织服装是决策树样本外检验唯一失败的行业，此外金属非金属效果也一般，其余行业效果较好。从日期来看，市场的因素对模型的效果影响很大，1，2 期表现都很好，但在 4 期几个行业选股效果集体减弱，在 8，9 期则出现了集体失效的情况。我们相信 2010 年 4 月开始的市场持续下跌，股指期货上市对市场结构的改造等因素可以部分解释模型失效的原因。此外，依照月度收益率的图形，我们发现 2010 年 7 月市场的突然反弹，也造成了一些问题，尤其是原本在食品饮料行业长期有效的模型突然给出了收益率较差的多头组合。由于决策树模型根本上可以理解为是一种因子动量模型，我们期待在过去某些产生超额收益的因子，在未来依然会产生超额收益。因此，在因子动量失效的情况下，模型结果有可能较差。

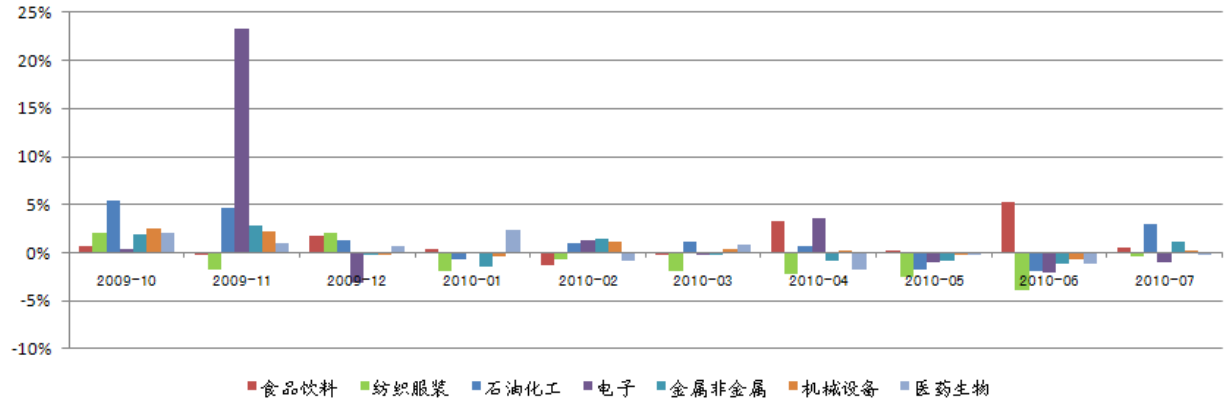
表 13：各行业样本外决策树选股效果对比

	占比情况		月度收益率效果										
行业	多头	空头	1期	2期	3期	4期	5期	6期	7期	8期	9期	10期	累积收益效果
食品饮料	高	低	好	中	好	好	中	好	好	好	好	差	好
纺织服装	低	低	好	中	好	差	中	差	差	差	差	差	差
石油化工	中	低	好	好	好	差	好	好	中	差	差	好	好
电子仪器	低	低	好	好	差	差	好	好	好	差	差	差	好
金属非金属	中	低	好	好	差	差	好	好	差	差	差	好	中
机械设备	中	低	好	好	中	差	好	好	差	差	差	好	好
医药生物	高	低	好	好	好	好	中	好	中	中	差	差	好

数据来源： Wind，国信证券经济研究所

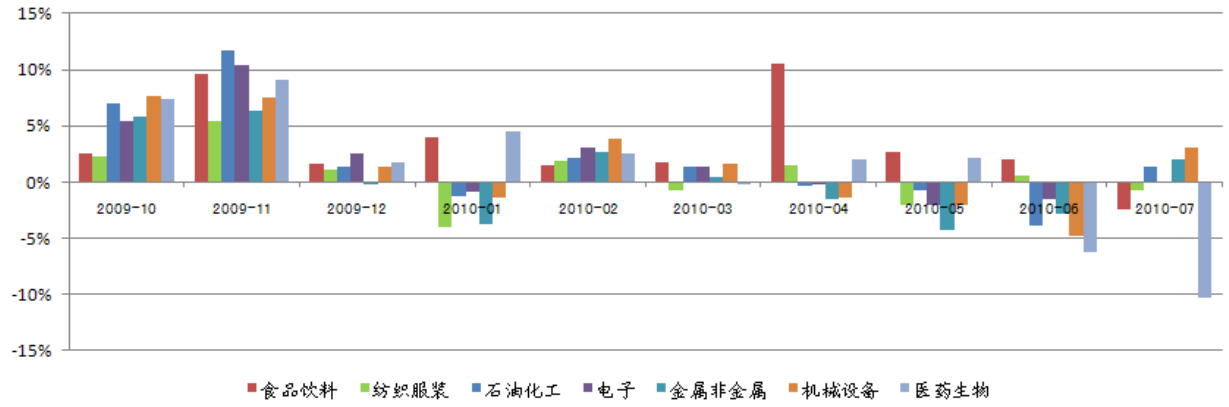
虽然决策树模型在纺织服装行业效果较差，但我们不能认为 DT 模型对该行业无效，这很有可能是样本数据过少，或这些因子并不适合应用在该行业中等因素造成的。我们现在能证明的是，在上文的六个因子下，DT 模型对食品饮料，石油化工，电子，机械设备与医药生物这五个行业有较好的选股效果。值得注意的是，随着我们对决策树修剪次数的更改，阈值设置的变换，或者是因子选择的更替，这些操作都有可能对模型实现进一步的优化，增强选股效果。**但在本节中，我们强调的是相同因子下行业间同期选股效果的对比。**

图 22：各行业样本外多头组合超额收益对比



数据来源： Wind，国信证券经济研究所

图 23：各行业样本外空头组合超额收益对比



数据来源： Wind，国信证券经济研究所

因子加工后的决策树模型选股效果

以上我们对制造业的子行业均采用了传统的六因子决策树模型，得到了不同的选股效果。我们在第一篇报告中谈到，不同的行业应该有不同的显著因子，因此在这里我们对传统的六因子分行业进行一个优选，来检验差异化的因子是否会令决策树的效果更好。这里需要注意的是，因为优选后因子数量减少，修剪次数我们统一定为 0，即不对决策树进行修剪，否则模型会直接失效。经过测算，在较少因子的情况下，修剪 0-5 次，模型效果较为稳定。因为存在参数设置的不同，我们不建议用本节的结果与上一节的结果进行同行业同期的比较，我们给出的累积收益比较图仅供参考。**在本节中，我们关注的是各行业优选的因子是否会让模型在行业间的相对有效性发生变化，以及因子数量减少，修剪次数不同等因素对模型的影响。**在下一篇报告中，我们会引入更多的因子并进行大范围的因子优选。

逐步回归的因子优选法

逐步回归的定义如下：在建立多元回归方程的过程中，按偏相关系数的大小次序将自变量逐个引入方程，对引入方程中的每个自变量偏相关系数进行统计检验，

效应显著的自变量留在回归方程内，循此继续遴选下一个自变量。如果效应不显著，停止引入新自变量。由于新自变量的引入，原已引入方程中的自变量由于变量之间的相互作用其效应有可能变得不显著者，经统计检验确证后要随时从方程中剔除，只保留效应显著的自变量。直至不再引入和剔除自变量为止，从而得到最优的回归方程。

我们的行业内六因子优选按照如下的流程进行：选定某行业内某只股票，将某时期样本内六因子数据作为自变量，将下一期该股票的收益率作为因变量进行逐步回归，得到对于该股票的显著因子。同理，将该行业内所有的股票逐个计算，得出所有股票各自的显著因子。根据行业内每个因子作为显著因子出现的总的次数，我们将出现次数大于出现次数中位数的因子保留，作为该行业的显著因子，并应用于决策树选股之中。

表 14：各行业因子出现次数及显著因子列表

	DTA	FEPS	MV	ROA	PE	RETURN	中位数
食品饮料	7	8	13	6	5	5	6.5
纺织服装	8	7	6	7	2	2	6.5
石油化工	12	17	16	20	15	16	16
电子仪器	4	9	11	11	7	2	8
金属非金属	11	17	8	16	6	8	9.5
机械设备	32	27	37	44	26	27	29.5
医药生物	9	14	15	9	9	4	9

	DTA	FEPS	MV	ROA	PE	RETURN
食品饮料	×	×	×			
纺织服装	×	×		×		
石油化工		×	×	×		×
电子仪器		×	×	×		
金属非金属	×	×		×		
机械设备	×		×	×		
医药生物	×	×	×	×	×	

数据来源： Wind，国信证券经济研究所

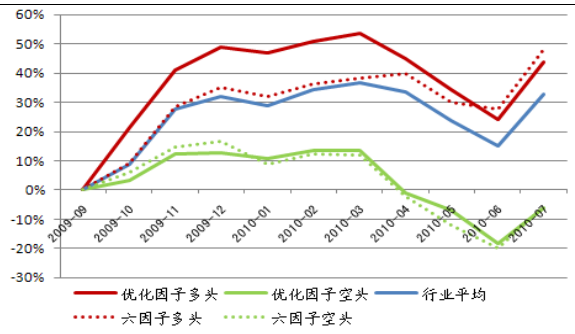
食品饮料行业

表 6：食品饮料样本外选股数量（修剪 0 次，阈值=10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	4	4	5.56%	5.56%
2009-11	11	5	15.28%	6.94%
2009-12	17	3	23.61%	4.17%
2010-01	15	3	20.83%	4.17%
2010-02	17	4	23.61%	5.56%
2010-03	19	4	26.39%	5.56%
2010-04	11	4	15.28%	5.56%
2010-05	20	5	27.78%	6.94%
2010-06	10	6	13.89%	8.33%
2010-07	7	8	9.72%	11.11%

数据来源： Wind，国信证券经济研究所

图 24：食品饮料行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

在压缩了因子数量后，各个行业的多空头股票数量有显著的下降，这主要是决策树模型本身的原因，而不能理解为多空头组合的优化。在食品饮料行业中，因子优化后，选股效果有很大的改善。

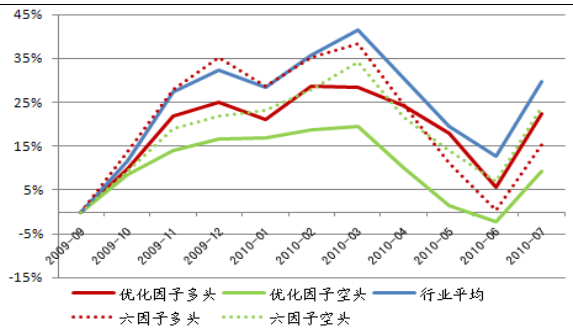
纺织服装行业

表 15：纺织服装样本外选股数量（修剪 0 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	5	12	7.46%	17.91%
2009-11	6	8	8.96%	11.94%
2009-12	10	10	14.93%	14.93%
2010-01	10	8	14.93%	11.94%
2010-02	5	11	7.46%	16.42%
2010-03	6	11	8.96%	16.42%
2010-04	13	8	19.40%	11.94%
2010-05	11	6	16.42%	8.96%
2010-06	9	14	13.43%	20.90%
2010-07	4	12	5.97%	17.91%

数据来源： Wind，国信证券经济研究所

图 25：纺织服装行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

纺织服装行业空头组合效果得到强化，但多头效果变化不大，依然远低于行业平均水平，可见模型在该行业中依然失效。

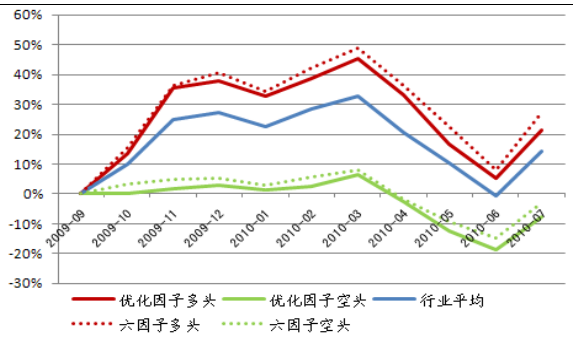
石油化工行业

表 7：石化行业样本外选股数量（修剪 0 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	21	37	11.67%	20.56%
2009-11	39	40	21.67%	22.22%
2009-12	26	40	14.44%	22.22%
2010-01	33	38	18.33%	21.11%
2010-02	50	34	27.78%	18.89%
2010-03	37	41	20.56%	22.78%
2010-04	49	36	27.22%	20.00%
2010-05	15	18	8.33%	10.00%
2010-06	23	22	12.78%	12.22%
2010-07	21	34	11.67%	18.89%

数据来源： Wind，国信证券经济研究所

图 26：石油化工行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

因子变动前后，石化行业选股的效果变化不大，但组合中股票数量大幅下降。再次注意，因为模型的修剪次数前后不同，累积收益率对比不能说明因子优化后的效果弱于未优化，只能说明模型在新的设置后的效果弱于设置前。

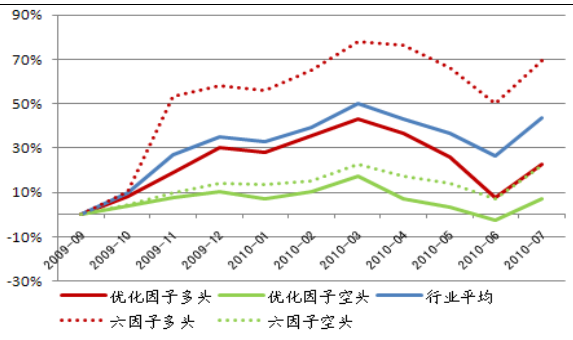
电子仪器行业

表 16：电子行业样本外选股数量（修剪 0 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	6	26	6.52%	28.26%
2009-11	7	26	7.61%	28.26%
2009-12	18	26	19.57%	28.26%
2010-01	17	24	18.48%	26.09%
2010-02	18	26	19.57%	28.26%
2010-03	15	24	16.30%	26.09%
2010-04	17	23	18.48%	25.00%
2010-05	16	23	17.39%	25.00%
2010-06	11	30	11.96%	32.61%
2010-07	6	20	6.52%	21.74%

数据来源： Wind，国信证券经济研究所

图 27：电子行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

该行业在新的模型下消除了之前的一次非常规超额收益，结果跑输行业平均，不过空头效果有所加强。

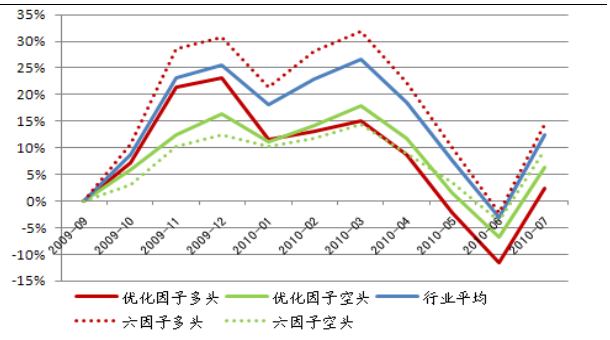
金属非金属行业

表 17：金属非金属样本外选股数量（修剪 0 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	21	41	13.82%	26.97%
2009-11	26	33	17.11%	21.71%
2009-12	32	36	21.05%	23.68%
2010-01	18	33	11.84%	21.71%
2010-02	19	32	12.50%	21.05%
2010-03	29	36	19.08%	23.68%
2010-04	23	34	15.13%	22.37%
2010-05	14	23	9.21%	15.13%
2010-06	11	36	7.24%	23.68%
2010-07	28	43	18.42%	28.29%

数据来源： Wind，国信证券经济研究所

图 28：金属非金属行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

因子修改后的模型在该行业中完全失效，多头跑输行业平均，甚至跑输于空头，此外空头的效果也弱于修改前。

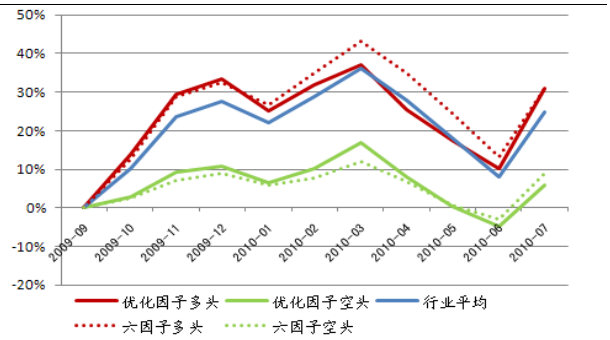
机械设备行业

表 18：机械设备样本外选股数量（修剪 0 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	53	87	16.88%	27.71%
2009-11	59	79	18.79%	25.16%
2009-12	70	74	22.29%	23.57%
2010-01	68	72	21.66%	22.93%
2010-02	45	74	14.33%	23.57%
2010-03	61	69	19.43%	21.97%
2010-04	68	70	21.66%	22.29%
2010-05	50	61	15.92%	19.43%
2010-06	24	72	7.64%	22.93%
2010-07	23	70	7.32%	22.29%

数据来源： Wind，国信证券经济研究所

图 29：机械设备行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

该行业选股效果相变化不大。我们认为虽然优化后，因子减少为 3 个，但是由于样本股有 314 只，模型没有受到数据变少的过多影响。这表明，决策树模型对数据量的需求比较大。

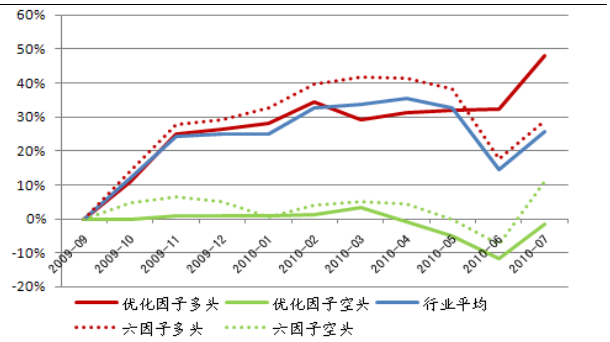
医药生物行业

表 19：医药生物样本外选股数量（修剪 0 次，阈值 = 10）

时间	多头组合	空头组合	多头占比	空头占比
2009-10	3	19	2.56%	16.24%
2009-11	5	16	4.27%	13.68%
2009-12	9	17	7.69%	14.53%
2010-01	7	10	5.98%	8.55%
2010-02	10	11	8.55%	9.40%
2010-03	12	10	10.26%	8.55%
2010-04	7	11	5.98%	9.40%
2010-05	4	10	3.42%	8.55%
2010-06	2	8	1.71%	6.84%
2010-07	1	1	0.85%	0.85%

数据来源： Wind，国信证券经济研究所

图 30：医药生物行业样本外组合累积收益率对比



数据来源： Wind，国信证券经济研究所

该行业内多头组合开始比较疲弱，但末期有明显加强的态势，空头效果则一直更好。对比前后，多头的数量有极大幅度的减小。

优化因子后的模型选股效果小结

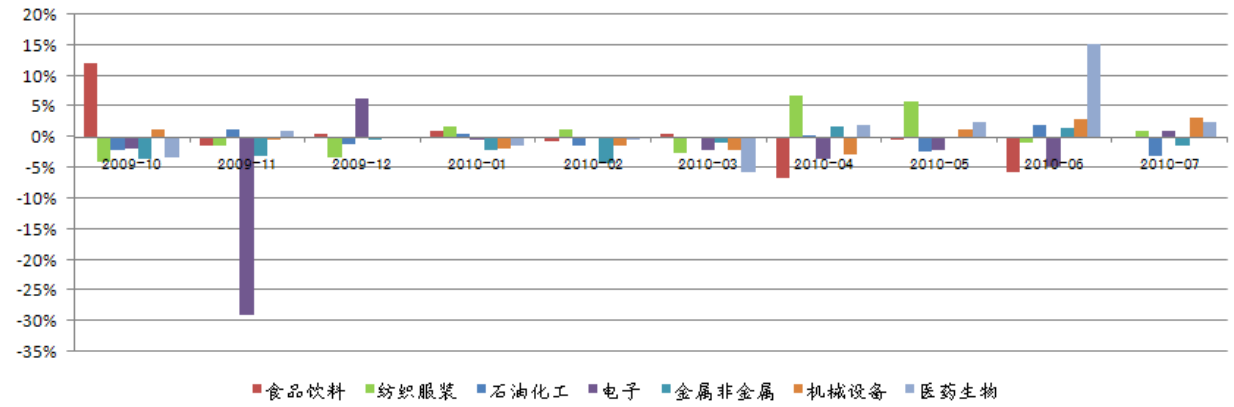
在因子优化后，从选股的累积收益来看，纺织服装，电子和金属非金属中模型均失效。我们认为失效的主要原因是因子减少后，可以挖掘的数据过少，导致我们被迫不去对决策树进行任何加工。在接下来的研究中，我们会大幅增加可选因子数量，依照行业分别去进行因子优化。

表 20：各行业样本外决策树选股效果对比

行业	占比情况		月度收益率效果										
	多头	空头	1期	2期	3期	4期	5期	6期	7期	8期	9期	10期	累积收益效果
食品饮料	低	低	好	中	好	好	中	好	中	差	中	好	好
纺织服装	低	低	中	中	中	差	好	差	好	中	差	好	差
石油化工	低	低	好	好	好	差	中	好	好	差	差	好	好
电子仪器	低	低	好	好	差	差	好	好	好	差	差	差	差
金属非金属	低	低	中	中	差	差	差	差	差	差	差	中	差
机械设备	低	低	好	好	中	差	中	差	差	中	差	好	好
医药生物	低	低	中	好	好	好	中	差	好	好	中	中	好

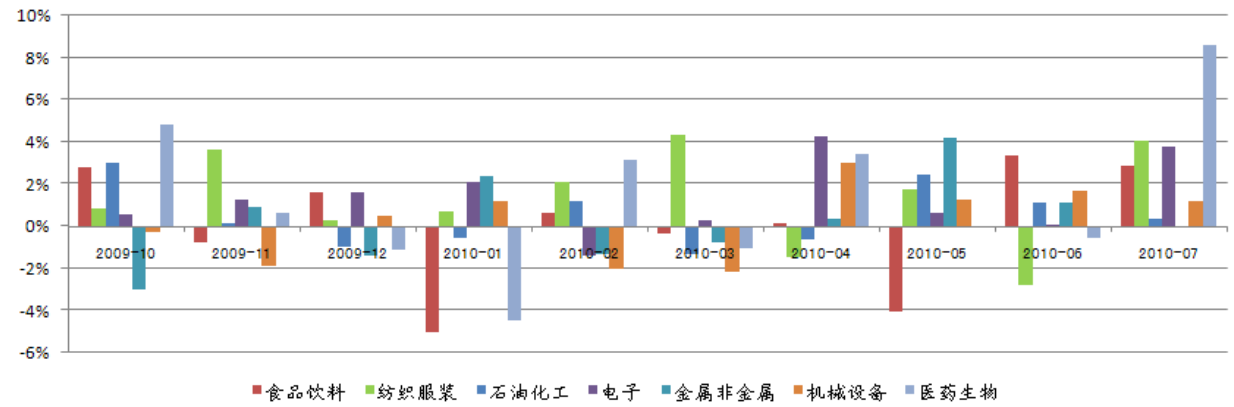
数据来源： Wind，国信证券经济研究所

图 31：各行业样本外多头组合超额收益对比



数据来源： Wind，国信证券经济研究所

图 32：各行业样本外空头组合超额收益对比



数据来源： Wind，国信证券经济研究所

国信证券投资评级

类别	级别	定义
股票投资评级	推荐	预计 6 个月内，股价表现优于市场指数 20%以上
	谨慎推荐	预计 6 个月内，股价表现优于市场指数 10%-20%之间
	中性	预计 6 个月内，股价表现介于市场指数±10%之间
	回避	预计 6 个月内，股价表现弱于市场指数 10%以上
行业投资评级	推荐	预计 6 个月内，行业指数表现优于市场指数 10%以上
	谨慎推荐	预计 6 个月内，行业指数表现优于市场指数 5%-10％之间
	中性	预计 6 个月内，行业指数表现介于市场指数±5%之间
	回避	预计 6 个月内，行业指数表现弱于市场指数 5%以上

免责条款

国信证券经济研究所团队成员

宏观
周炳林 0755-82130638
林松立 010-66026312

策略
黄学军 021-60933142
崔 嵘 021-60933159
廖 喆 021-60933162

交通运输
郑 武 0755- 82130422
陈建生 0755- 82133766
岳 鑫 0755- 82130432
高 健 0755-82130678

银行
邱志承 021- 60875167
黄 飙 0755-82133476
谈 煊 010- 66025229

房地产
方 焱 0755-82130648
区瑞明 0755-82130678
黄道立 0755- 82133397

机械
余爱斌 0755-82133400
黄海培 021-60933150
陈 玲 0755-82130646
杨 森 0755-82133343
李筱筠 010-66026326

汽车及零配件
李 君 021-60933156
左 涛 021-60933164

钢铁
郑 东 010- 66026308
秦 波 010-66026317

商业贸易
孙菲菲 0755-82130722
吴美玉 010-66026319
祝 彬 0755-82131528

基础化工
张栋梁 0755-82130532
陈爱华 0755-82133397
邱 斌 0755-82130532

医药
贺平鸽 0755-82133396
丁 丹 0755- 82139908
陈 栋 021-60933147

石油与石化
李 晨 021-60875160
严蓓娜 021-60933165

电力设备与新能源
皮家银 021-60933160

传媒
陈财茂 021-60933163

有色金属
彭 波 0755-82133909
谢鸿鹤 0755-82130646

电力与公用事业
徐颖真 021-60875162
谢达成 021-60933161

非银行金融
邵子钦 0755- 82130468
田 良 0755-82130513
童成敦 0755-82130513

通信
严 平 021-60875165
程 峰 021-60933167

造纸
李世新 0755-82130565
邵 达 0755-82130706

家电
王念春 0755-82130407

计算机
段迎晟 0755- 82130761

电子元器件
段迎晟 0755- 82130761
高耀华 0755-82130771

纺织服装
方军平 021-60933158

农业
张 如 021-60933151

旅游
廖绪发 021-60875168
刘智景 021-60933148

食品饮料
黄 茂 0755-82138922

建材
杨 昕 021-60933168

煤炭
李 然 010-66026322
陈 健 010-66215566
苏绍许 021-60933144

建筑
邱 波 0755-82133390
李遵庆 0755-82133055

固定收益
李怀定 021-60933152
高 宇 0755- 82133538
侯慧娣 021-60875161
张 旭 010-66026340
蔺晓熠 021-60933146
刘子宁 021-60933145

指数与产品设计
焦 健 0755-82133928
王军清 0755-82133297
彭甘霖 0755-82133259
阳 璀 0755-82133538
周 琦 0755-82133568
赵学昂 0755-66025232

投资基金
杨 涛 0755-82133339
彭怡萍 0755-82133528
刘舒宇 0755-82133568
康 亢 010-66026337
刘 洋

数量化投资
葛新元 0755-82133332
董艺婷 021-60933155
林晓明 0755-25472656
赵斯尘 021-60875174
程景佳 021-60933166
郑 云 021-60875163
毛 甜 021-60933154

交易策略
戴 军 0755-82133129
秦国文 0755-82133528
徐左乾 0755-82133090
黄志文 0755-82133928

国信证券机构销售团队

华北区（机构销售一部）		华东区（机构销售二部）		华南区（机构销售三部）	
王立法	010-66026352 13910524551 wanglf@guosen.com.cn	盛建平	021-60875169 15821778133 shengjp@guosen.com.cn	万成水	0755-82133147 13923406013 wancs@guosen.com.cn
王晓建	010-66026342 13701099132 wangxj@guosen.com.cn	马小丹	021-60875172 13801832154 maxd@guosen.com.cn	魏　宁	0755-82133492 13823515980 weining@guosen.com.cn
焦　戬	010-66026343 13601094018 jiaojian@guosen.com.cn	郑　毅	021-60875171 13795229060 zhengyi@guosen.com.cn	邵燕芳	0755-82133148 13480668226 shaoyf@guosen.com.cn
李　锐	010-66025249 13691229417 lirui2@guosen.com.cn	黄胜蓝	021-60875166 13761873797 huangsl@guosen.com.cn	林　莉	0755-82133197 13824397011 linli2@guosen.com.cn
徐文琪	010-66026341 13811271758 xuwq@guosen.com.cn	刘　塑	021-60875177 13817906789 liusu@guosen.com.cn	王昊文	0755-82130818 18925287888 wanghaow@guosen.com.cn
		叶琳菲	021-60875178 13817758288 yelf@guosen.com.cn	甘　墨	0755-82133456 15013851021 ganmo@ guosen.com.cn
		孔华强	021-60875170 13681669123 konghq@guosen.com.cn	段莉娟	0755-82130509 18675575010 duanlj@guosen.com.cn
				黎　敏	0755-82130681 13902482885 limin1@guosen.com.cn
				徐　冉	13632580795 xuran1@guosen.com.cn
				颜小燕	13590436977 yanxy@guosen.com.cn