

分析师 刘华峰
021-38565743
liuhf@xyzq.com.cn
SAC: S0190511070003

聚类高频择时算法交易策略

2011年9月25日

报告关键点

投资要点

相关报告

- 算法交易是指事先设计好交易策略，然后将其编制成计算机程序，利用计算机程序的算法来决定交易下单的时机、价格和数量等。相比于手动订单执行而言，算法交易在管理市场冲击成本、机会成本和风险等方面具有一系列的优势。而不管手动订单执行还是算法交易，如何选择下单时机，将交易成本控制市场平均水平之下，都是交易环节不得不关注的。
- 我们通过对股票高频交易历史时段进行聚类，找出价、量形态存在相似性的时段，研究其概率分布及后续胜率后发现，根据相似时段的上涨概率进行交易决策，胜率在 60% 左右。
- 从样本测试结果来看，根据概率分布调整交易量的分配，不管是大盘股、中盘股、小盘股，也不管是买入择时还是卖出择时，从择时效果看都是成功的，相对基准可获得-2BP~53BP 的日内择时收益，且择时收益较为稳定，正择时收益比例占 80% 以上。从选取样本的择时收益看，小盘股高于中盘股，中盘股高于大盘股。
- 高频择时算法交易策略可用于以下两方面：1、算法交易：通过算法交易系统直接完成下单操作；2、手动订单执行：通过高频择时随时监测市场动向，获得交易信号，供交易员下单操作时参考。

目 录

一、策略思想	- 3 -
(一) 聚类分析	- 3 -
(二) 概率计算	- 4 -
(三) 获得交易信号	- 4 -
(四) 交易操作	- 4 -
二、策略验证	- 5 -
三、交易模拟	- 5 -
(一) 买入择时	- 5 -
(二) 卖出择时	- 9 -
四、报告结论	- 10 -
五、技术实现	- 10 -
表 1、中国石化测试结果	- 5 -
表 2、择时收益（中国石化，买入择时）	- 5 -
表 3、择时收益（大同煤业，买入择时）	- 7 -
表 4、择时收益（威孚高科，买入择时）	- 8 -
表 5、择时收益（中国石化，卖出择时）	- 9 -
图 1、时间加权分配（2011 年 9 月 23 日，中国石化，买入择时）	- 4 -
图 2、模拟交易量分布（2011 年 9 月 13 日，中国石化，买入择时）	- 6 -
图 3、模拟交易量分布（2011 年 9 月 14 日，中国石化，买入择时）	- 6 -
图 4、模拟交易量分布（2011 年 9 月 15 日，中国石化，买入择时）	- 6 -
图 5、模拟交易量分布（2011 年 9 月 23 日，中国石化，买入择时）	- 7 -
图 6、模拟交易量分布（2011 年 9 月 8 日，大同煤业，买入择时）	- 8 -
图 7、模拟交易量分布（2011 年 9 月 20 日，大同煤业，买入择时）	- 8 -
图 8、模拟交易量分布（2011 年 9 月 6 日，威孚高科，买入择时）	- 9 -
图 9、模拟交易量分布（2011 年 9 月 20 日，中国石化，卖出择时）	- 10 -
图 10、程序化交易结构图	- 11 -

算法交易是指事先设计好交易策略，然后将其编制成计算机程序，利用计算机程序的算法来决定交易下单的时机、价格和数量等。相比于手动订单执行而言，算法交易在管理市场冲击成本、机会成本和风险等方面具有一系列的优势。

而不管手动订单执行还是算法交易，如何选择下单时机，将交易成本控制在市场平均水平之下，都是交易环节不得不关注的。我们通过对股票高频交易历史时段进行聚类，找出价、量形态存在相似性的时段，研究其概率分布及后续胜率发现择时效果显著。

高频择时算法交易策略可用于以下两方面：1、算法交易：通过算法交易系统直接完成下单操作；2、手动订单执行：通过高频择时随时监测市场动向，获得交易信号，供交易员下单操作时参考。

一、策略思想

当前投资者的投资操作行为或多或少地能为未来一定时期的市场走势提供参考，而最直接反映投资者当前操作的主要是价、量指标，我们假定历史价、量指标与之后一段时期走势之间的对应关系在未来依然成立，因此我们通过分析历史数据来预测未来。

历史市场形态多种多样，依靠人脑无法在短时间内立刻做出准确的判断，从而不错失交易良机，因此我们借助聚类来完成，将历史不同价、量形态的时段通过聚类分析自动分类，判别当前要操作的这一分钟处于历史何种分类中，所在分类的后续走势如何，依此判断当前操作该如何进行，具体步骤如下：

（一）聚类分析

在聚类分析中，动态聚类法是较普遍采用的方法，在数据点分类过程中按照某种准则动态调整数据点类型归属。

通过聚类，将股票高频交易的时间根据指标分成存在差异的几类，每天对历史一段时期内的高频交易时间点按照指标进行快速聚类，对时间点价、量形态的相似性进行分类，找出与当前时间点对应的历史时点组合。

分析过程所使用的相关参数如下：

- 1) 每天的交易时间段：本报告只选取了 9:41-14:30 的交易时间段，前 10 分钟用来做判断，最后的半个小时用以满足特别处理需求；
- 2) 聚类的时间段：以 5 个交易日做历史数据的时间段；
- 3) 聚类指标的时间段：以 10 分钟价、量指标做聚类；
- 4) 聚类指标：涨幅、委比（1 档）、委比（5 档）。

（二）概率计算

计算对应分类时点组合后续一段时期的涨幅及上涨的概率，‘涨幅’和概率的计算公式如下：

- 1) 涨幅=第 10 分钟的平均成交价/当前分钟平均成交价-1
- 2) 概率计算公式：涨幅为正的个数/涨幅总个数*100%

（三）获得交易信号

假定以对应分类时点组合上涨概率超过 65%时作为买入信号（简称买入点），即判断 10 分钟后的成交价将高于当前分钟的成交价；以对应分类时点组合上涨概率低于 35%时作为卖出信号（简称卖出点），即判断 10 分钟后的成交价将低于当前分钟的成交价。

（四）交易操作

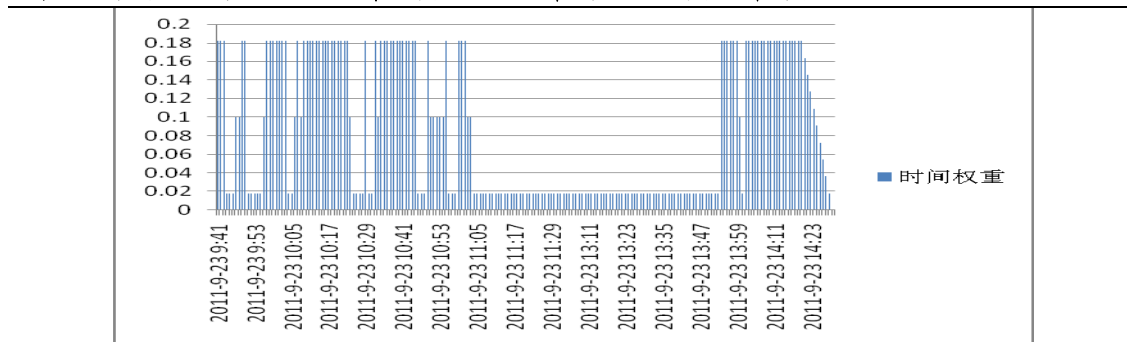
为简单起见，假定基准成交方法为每分钟成交量等分布，且按每分钟平均成交价成交，在此基础上进行高频择时后的交易量重新分配。如果在本基准上的增强有效，那么从理论上说对于在 TWAP（Time Weighted Average Price）、VWAP（Volume Weighted Average Price）上的增强应该也是有效的。

以买入股票为例，分配规则如下：将交易时间划分为以分钟为单位的交易段，首先我们对各个交易段的交易量进行等分布，在每个交易段进行交易前阅读交易信号，为买入信号时，判断未来 10 分钟为上涨趋势，对未来 10 分钟的交易量按时间加权分配，依次分配 10/55、9/55、……、1/55，下一个交易段依此类推；当历史上涨概率低于 35%时，做相反分配；否则平均分配。

卖出股票与买入股票的交易操作相反。即当交易信号为买入时，对未来 10 分钟的交易量按时间加权分配，依次分配 1/55、2/55、……、10/55；当交易信号为卖出时，依次分配 10/55、9/55、……、1/55；否则平均分配。

以中国石化买入择时为例，时间加权分配如下：

图 1、时间加权分配（2011 年 9 月 23 日，中国石化，买入择时）



数据来源：兴业证券研究所 天软科技

二、策略验证

为验证策略的有效性，我们以中国石化（600028）为例，对 2011 年 4 月至 8 月共 5 个月的数据进行了测算，以买入后涨幅为正或卖出后涨幅为负为胜，总胜率基本维持在 60% 左右，且择时买入后的平均涨幅全部为正，择时卖出后的平均涨幅全部为负。

表 1、中国石化测试结果

开始日期	结束日期	买入点	卖出点	买入点胜率	卖出点胜率	总胜率	买入后平均涨幅	卖出后平均涨幅
2011/4/1	2011/5/1	579	902	0.56	0.68	0.63	0.08	-0.08
2011/5/1	2011/6/1	845	854	0.54	0.66	0.60	0.02	-0.06
2011/6/1	2011/7/1	562	466	0.58	0.63	0.60	0.05	-0.06
2011/7/1	2011/8/1	525	800	0.58	0.65	0.62	0.01	-0.05
2011/8/1	2011/9/1	285	991	0.52	0.60	0.58	0.05	-0.05

数据来源：兴业证券研究所 天软科技

三、交易模拟

根据前述交易操作方法，我们分别选取大盘股（中国石化，600028）、中盘股（大同煤业，601001）、小盘股（威孚高科，000581）一支，以 9 月份 1-25 日各交易日为例进行交易模拟，交易股票数量分别为 1 万手、5000 手、1000 手，计算择时收益如下（成交额单位：万元）：

（一）买入择时

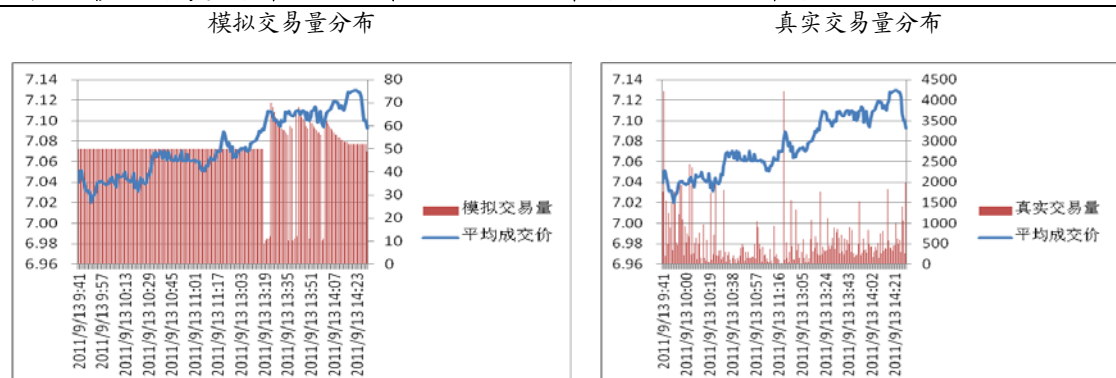
表 2、择时收益（中国石化，买入择时）

时间	买入点	卖出点	总胜率	择时前成交额	择时后成交额	择时买入收益（基点）
20110901	20	21	0.68	737.05	736.93	1.59
20110902	2	12	0.86	722.27	722.21	0.78
20110905	22	0	0.77	709.37	709.34	0.54
20110906	38	74	0.54	710.68	710.35	4.59
20110907	32	8	0.50	711.13	711.13	0.05
20110908	80	0	0.36	715.37	715.39	-0.26
20110909	32	1	0.67	715.08	715.04	0.49
20110913	0	16	0.63	707.58	707.61	-0.34
20110914	46	32	0.44	710.09	710.11	-0.34
20110915	0	0		715.14	715.14	0.00
20110916	0	18	0.72	719.46	719.45	0.12
20110919	2	0	1.00	707.02	707.02	0.01
20110920	59	73	0.57	704.50	704.10	5.69
20110921	0	0		710.94	710.94	0.00
20110922	0	0		709.15	709.15	0.00
20110923	82	103	0.75	694.73	694.21	7.48

数据来源：兴业证券研究所 天软科技

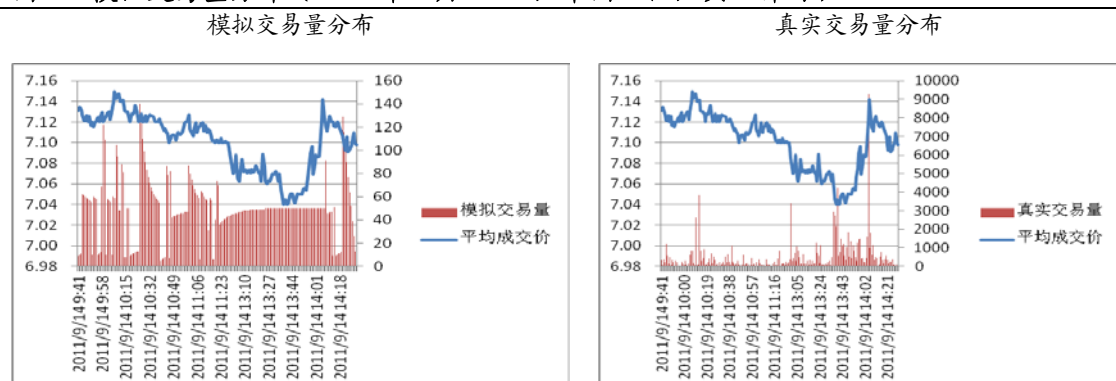
从上表可以看到，在中国石化 9 月 1 日-25 日的 16 个交易日内，只有 3 个交易日的择时收益为负，且负的幅度较小，其中 2011 年 9 月 13 日、14 日负收益最大，也只有-0.34BP；3 个交易日即 2011 年 9 月 15 日、21 日、22 日未监测到交易信号，按基准成交方法平均成交，无择时收益；其他 10 个交易日的择时收益皆为正，其中 2011 年 9 月 23 日有较大的正收益，达到 7.48BP。2011 年 9 月 13 日、2011 年 9 月 14 日、2011 年 9 月 15 日、2011 年 9 月 23 日的模拟交易量分布图如下（交易量单位：手）：

图 2、模拟交易量分布（2011 年 9 月 13 日，中国石化，买入择时）



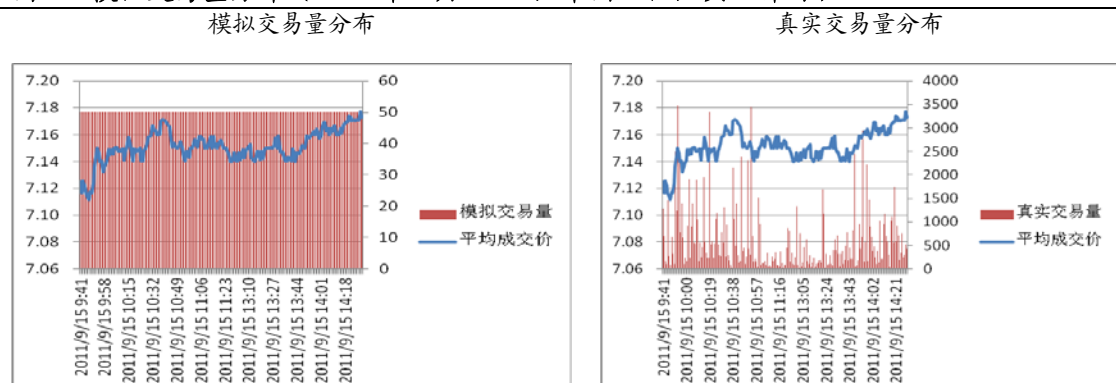
数据来源：兴业证券研究所 天软科技

图 3、模拟交易量分布（2011 年 9 月 14 日，中国石化，买入择时）



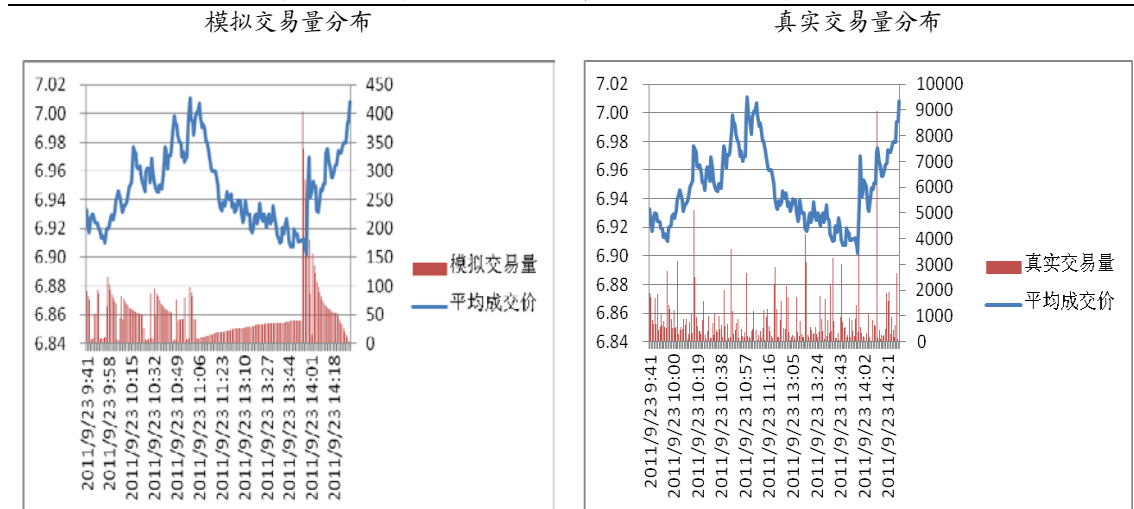
数据来源：兴业证券研究所 天软科技

图 4、模拟交易量分布（2011 年 9 月 15 日，中国石化，买入择时）



数据来源：兴业证券研究所 天软科技

图 5、模拟交易量分布（2011 年 9 月 23 日，中国石化，买入择时）



数据来源：兴业证券研究所 天软科技

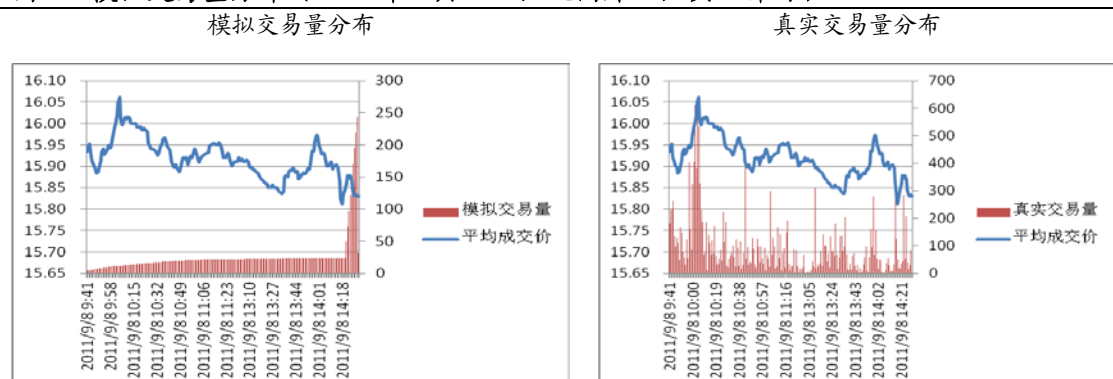
表 3、择时收益（大同煤业，买入择时）

时间	择时 买入点	择时 卖出点	总胜率	择时前 成交额	择时后 成交额	择时买入 收益(基点)
20110901	2	42	0.89	806.98	806.60	4.66
20110902	21	6	0.78	798.41	798.33	1.05
20110905	0	38	0.68	778.50	777.62	11.29
20110906	17	9	0.92	775.22	775.09	1.76
20110907	153	18	0.59	786.89	786.60	3.71
20110908	1	199	0.59	795.90	794.84	13.26
20110909	31	88	0.64	792.69	792.22	6.01
20110913	3	0	1.00	784.39	784.37	0.21
20110914	67	0	0.52	792.03	792.03	-0.05
20110915	0	21	0.86	800.88	800.79	1.06
20110916	0	17	0.53	799.29	799.25	0.54
20110919	59	102	0.54	783.25	782.89	4.60
20110920	0	103	0.49	784.11	784.26	-1.92
20110921	8	1	0.78	801.28	801.21	0.81
20110922	7	90	0.60	802.58	802.35	2.90
20110923	83	15	0.83	789.92	789.32	7.57

数据来源：兴业证券研究所 天软科技

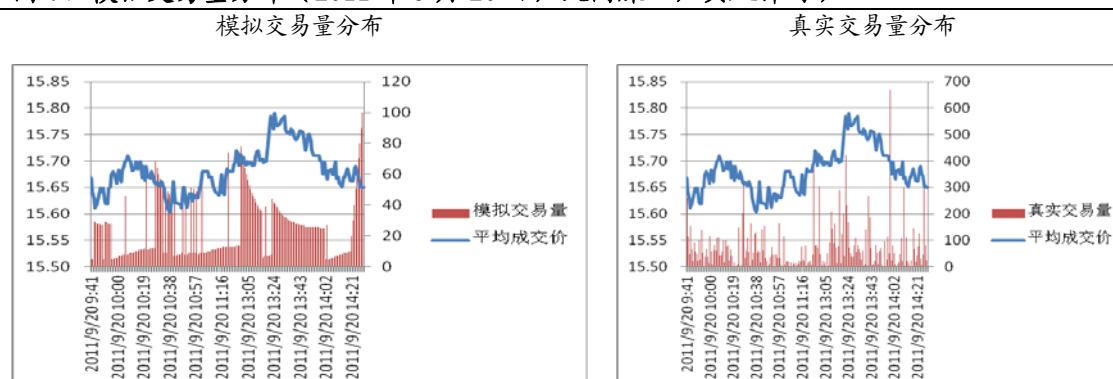
从上表可以看到，在大同煤业 9 月 1 日-25 日的 16 个交易日，只有 2 个交易日的择时收益为负，其中 2011 年 9 月 20 日负收益较大，为-1.92BP；其他 14 个交易日的择时收益皆为正，其中 2011 年 9 月 8 日有较大的正收益，达到 13.26BP。2011 年 9 月 8 日、2011 年 9 月 20 日的模拟交易量分布图如下（交易量单位：手）：

图 6、模拟交易量分布（2011 年 9 月 8 日，大同煤业，买入择时）



数据来源：兴业证券研究所 天软科技

图 7、模拟交易量分布（2011 年 9 月 20 日，大同煤业，买入择时）



数据来源：兴业证券研究所 天软科技

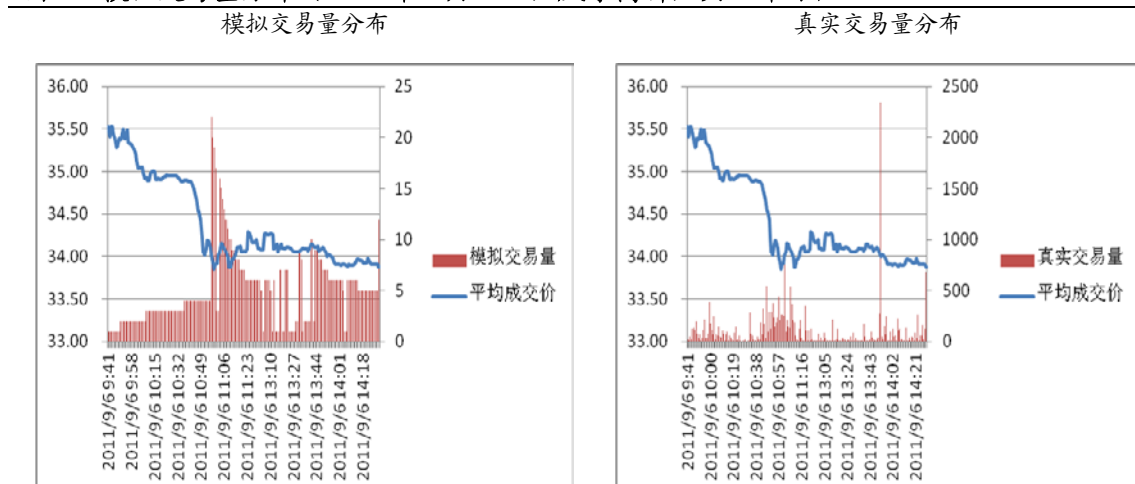
表 4、择时收益（威孚高科，买入择时）

时间	择时 买入点	择时 卖出点	总胜率	择时前 成交额	择时后 成交额	择时买入 收益(基点)
20110901	0	69	0.68	360.14	359.93	5.64
20110902	0	97	0.60	356.80	356.46	9.48
20110905	57	35	0.61	353.99	353.82	5.03
20110906	0	101	0.75	343.92	342.08	53.37
20110907	0	0		338.82	338.82	0.00
20110908	0	85	0.53	337.93	337.82	3.20
20110909	0	60	0.78	337.10	336.51	17.56
20110913	0	53	0.74	329.96	329.88	2.31
20110914	70	0	0.50	330.64	330.61	1.00
20110915	12	51	0.68	333.27	333.12	4.54
20110916	38	2	0.55	332.69	332.69	0.07
20110919	0	173	0.59	329.79	329.55	7.24
20110920	33	31	0.48	330.15	330.14	0.35
20110921	0	33	0.58	339.29	339.23	1.98
20110922	0	2	1.00	336.16	336.16	0.04
20110923	9	161	0.54	326.97	326.60	11.57

数据来源：兴业证券研究所 天软科技

从上表可以看到，在威孚高科 9 月 1 日-25 日的 16 个交易日内，只有 1 个交易日即 2011 年 9 月 7 日未监测到交易信号，按基准成交方法平均成交，无择时收益；其他 15 个交易日的择时收益皆为正，其中 2011 年 9 月 6 日的正收益最大，高达 53.37BP。2011 年 9 月 6 日的模拟交易量分布图如下（交易量单位：手）：

图 8、模拟交易量分布（2011 年 9 月 6 日，威孚高科，买入择时）



数据来源：兴业证券研究所 天软科技

从择时点和模拟交易量分布看，作买入交易时，模拟交易量在低平均成交价区域有明显的放大。

（二）卖出择时

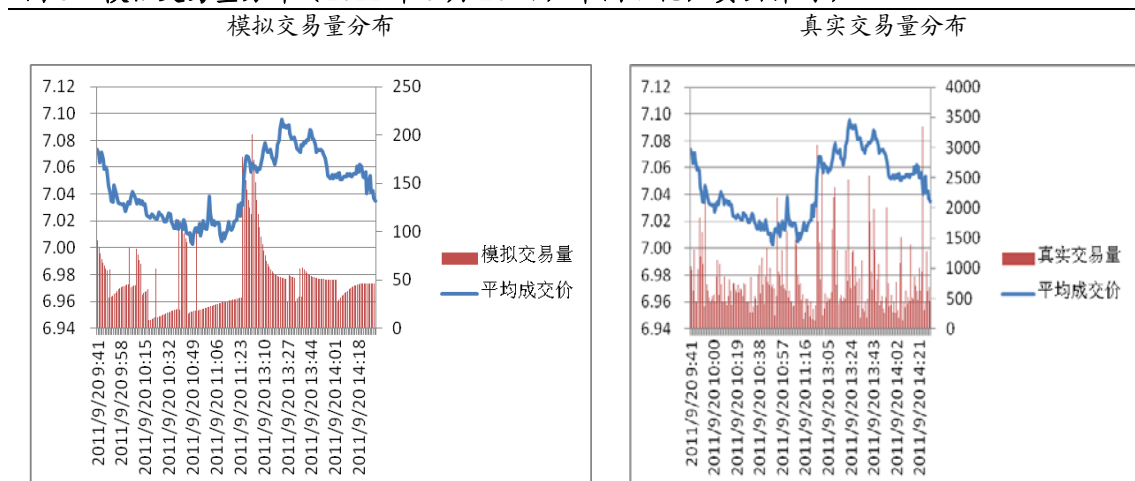
表 5、择时收益（中国石化，卖出择时）

时间	择时 买入点	择时 卖出点	总胜率	择时前 成交额	择时后 成交额	择时买入 收益(基点)
20110901	20	21	0.68	737.05	737.13	1.07
20110902	2	12	0.86	722.27	722.30	0.47
20110905	22	0	0.77	709.37	709.46	1.15
20110906	38	74	0.54	710.68	710.85	2.46
20110907	32	8	0.50	711.13	711.16	0.37
20110908	80	0	0.36	715.37	715.29	-1.13
20110909	32	1	0.67	715.08	715.11	0.43
20110913	0	16	0.63	707.58	707.56	-0.28
20110914	46	32	0.44	710.09	709.99	-1.42
20110915	0	0		715.14	715.14	0.00
20110916	0	18	0.72	719.46	719.47	0.10
20110919	2	0	1.00	707.02	707.03	0.02
20110920	59	73	0.57	704.50	705.11	8.57
20110921	0	0		710.94	710.94	0.00
20110922	0	0		709.15	709.15	0.00
20110923	82	103	0.75	694.73	695.27	7.78

数据来源：兴业证券研究所 天软科技

从上表可以看到，在中国石化 9 月 1 日-25 日的 16 个交易日内，只有 3 个交易日的择时收益为负；3 个交易日未监测到交易信号，按基准成交方法平均成交，无择时收益；其他 10 个交易日的择时收益皆为正，其中 2011 年 9 月 20 日有较大的正收益，达到 8.57BP。2011 年 9 月 20 日的模拟交易量分布图如下（交易量单位：手）：

图 9、模拟交易量分布（2011 年 9 月 20 日，中国石化，卖出择时）



数据来源：兴业证券研究所 天软科技

从择时点和模拟交易量分布看，作卖出交易时，模拟交易量在高平均成交价区域有明显的放大。

四、报告结论

我们通过对股票历史时段进行聚类，找出存在相似性的时段，研究其概率分布及后续胜率发现，根据相似时段的上涨概率进行交易决策，胜率在 60% 左右。

从样本测试结果来看，根据概率分布调整交易量的分配，不管是大盘股、中盘股、小盘股，也不管是买入择时还是卖出择时，从择时效果看都是成功的，相对基准可获得-2BP~53BP 的日内择时收益，且择时收益较为稳定，正择时收益比例占 80% 以上。从选取样本的择时收益看，小盘股高于中盘股，中盘股高于大盘股。

五、技术实现

算法交易在国外已广泛应用，国内虽处于起步阶段，仅在极少数机构实施，但这方面的研究工作正积极开展，同时技术方面也趋于成熟，以天软算法交易平台为例，程序化交易结构图如下，无论从方法还是技术，我们都有理由相信算法

投资评级说明

行业评级 报告发布日后的 12 个月内行业股票指数的涨跌幅度相对同期上证综指/深圳成指的涨跌幅为基准,投资建议的评级标准为:

推 荐: 相对表现优于市场
中 性: 相对表现与市场持平
回 避: 相对表现弱于市场

公司评级 报告发布日后的 12 个月内公司的涨跌幅度相对同期上证综指/深圳成指的涨跌幅为基准,投资建议的评级标准为:

强烈推荐: 相对大盘涨幅大于 15%
推 荐: 相对大盘涨幅在 5% ~ 15%之间
中 性: 相对大盘涨幅在-5% ~ 5%之间
回 避: 相对大盘涨幅小于-5%

机构客户部联系方式

上海市浦东新区民生路 1199 弄证大五道口广场 1 号楼 21 层
邮编: 200135
传真: 021-38565955

北京市西城区武定侯街 2 号泰康国际大厦 6 层
邮编: 100140
传真: 010-66290200

深圳市福田区益田路 4068 号卓越时代广场 15 楼 1502-1503
邮编: 518048
传真: 0755-82562090

信息披露

本公司在知晓的范围内履行信息披露义务。客户可登录www.xyzq.com.cn内幕交易防控栏内查询静默期安排和关联公司持股情况。

重要声明

兴业证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

兴业证券系列报告的信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，报告中的信息或意见并不构成所述证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。

在法律许可的情况下，兴业证券股份有限公司可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到兴业证券股份有限公司及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。