证券研究报告—深度报告

金融工程

数量化投资

数量化投资系列报告之四十二

2011 年 8 月 30 日

相关研究报告:

专题报告

基于模式聚类的短线选股模型

本文将传统的 K-均值聚类算法修改为层次的 K-均值聚类算法；并使用这一方法对中证 800 历史成分股， 05 年 3 月 28 日至 11 年 3 月 14 日的 15 分钟 k 线以滚动 3 个交易日为时间长度进行成交价格和成交量的模式聚类。在观察的样本期内，以价格时间序列中变化最大的时间点作为关键点，并将第一个关键点的成交价格归一化为 100，其它的关键点类似处理，用欧式距离度量拟合后的两条折线段的相似度。对于成交量来说，我们计算两个关键点之间成交量的均值，得到成交量随时间变化的曲线。

经过折线段拟合处理后，我们共计得到 121939 个成交价格时间序列样本。这些样本被划分到 198 个聚类结果中，即成交价格的不同模式。成交量样本被划分到 194 个聚类结果中并进行类似的分析。

预期上涨的价格模式中心在样本内解释力是最强的，伴随样本数量增加，收益普遍较为稳定，只有个别形态下降。预期上涨的成交量模式稳定性次之。预期下跌形态的价格模式解释力相对较弱，伴随样本数量增加收益普遍缓慢回升，预期下跌的成交量模式也呈现出同样的规律，收敛速度大于价格中心。

我们引入了 Average Precision 指标①对价、量模式进行组合。从排序在前 20 的组合中我们可以看到， AP 值的峰值基本出现在 lambda 值在 0.8-0.9 的区间，所有 20 个模式中的 lambda 值均大于 0.5。也就是说，在模式组合中，价格中心占据了绝对优势，量中心的权重相对而言要小得多。价量模式结合之后，无论对 top 还是 bottom 组合的稳定性和效率均有很大的提高。

两类组合的样本数量与指数活跃程度有一定的相关性，伴随指数长期上涨或下跌，样本数量会出现同步的增减，但波动程度远远大于指数的波动。平均而言每月出现的多头样本数量在 50-100 个之间。样本内 top 模式组合只有 05、06 和 10 年在 0.6%的交易成本下准确率未能高于 50%，其余指标均高于 50%。6527 个样本在 vwap 成交模式下持有 3h 的平均收益率 0.833%。Bottom 组合的效果与未去除重复样本时一致，显著差于 Top 组合。

由于样本外区间整体处于指数的下跌趋势中，Top 组合遭遇重创，表现不佳，无论是胜率还是准确率，都低于样本内的平均水平，但是 Bottom 组合反之；bottom 组合在每次交易中几乎没有误判，而 top 组合则在 3-6 月的时间出现了较大的反复，三季度则表现尚可。

未来我们将针对“模式识别”设计更加合理的目标函数来评价模式中心。我们也将借鉴“Learning to Rank”模型，利用机器学习的方法融合样本各种各样的特征，结合机器学习理论提高检索效果。后续应用将在以下几个方面展开：指数投资应用、基于模型选股结果的交易策略、频度扩展、模式伸缩、择时应用。

证券分析师：董艺婷
电话: 021-60933155
E-mail: dongyt@guosen.com.cn
证券投资咨询执业资格证书编码：S0980510120055
联系人：郑亚斌
电话: 021-60933150
E-mail: zhengyb@guosen.com.cn

内容目录

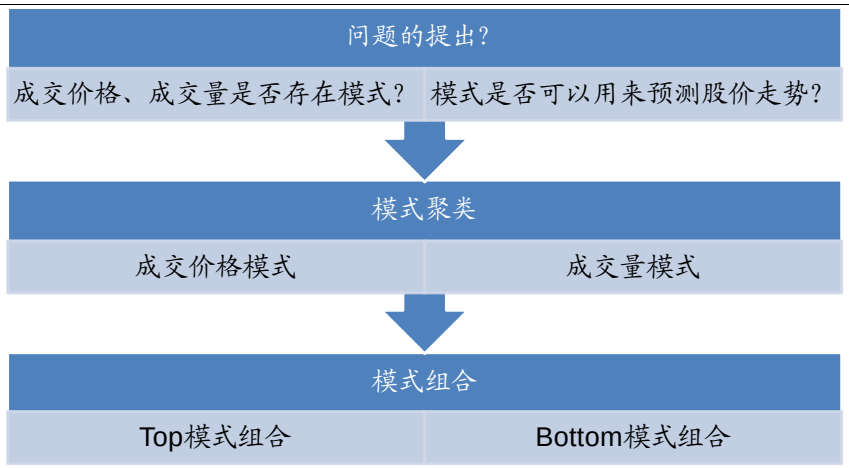
图表目录

前言

研究框架

大千世界，万事万物中都存在千丝万缕的联系。俗话说，物以类聚人以群分，说的也是这个道理。对于股票来说，我们也可以根据不同的特征对其进行聚类，将“相似”的股票样本放在同一个篮子里。相似性的度量可以有多个角度。从基本面角度来说，我们可以根据股票所属的行业对其进行划分，例如：金融行业、地产行业等；从技术面角度来说，我们可以根据股票的技术指标对其进行归类，例如：近期股价上涨、下跌、盘整，成交量放大、萎缩等。将股票样本分门别类有助于我们对它们进行更准确的分析。

图 1：报告研究框架



资料来源：国信证券经济研究所

本文的研究框架如上图所示。我们提出利用模式聚类方法构建选股模型，首先要确定的是技术指标（在本文中主要指的是成交价格、成交量）是否存在一定的可以用来预测股价走势的模式或模式组合。幸运的是，技术分析理论告诉我们这样的模式和模式组合是存在的，我们需要做的是把这些模式或模式组合从历史交易数据中挖掘出来。

接下来，既然这样的模式存在，我们如何找到成交价格、成交量的模式？我们利用经典的层次 K-均值聚类算法，结合合适的相似度度量方法，我们得到了大量的成交价格、成交量模式。我们需要的是对股价走势预测有帮助的模式，那些没有信息量的模式需要排除。

最后，技术分析表明，成交量与成交价格的结合可以进一步提高预测效果。因此，我们提出了一种模式组合的方法，可以很好地结合成交价格、成交量的模式并构建 Top、Bottom 模式组合。

模式聚类的提出

本文着重于从技术层面分析股票的模式，我们首先需要回答的问题是：技术指标是否存在某些模式可以用来对股票进行聚类？其次，这些模式是否可以帮助我们预测股价的走势？在回答了这些问题之后，我们才有必要对模式进行进一步的挖掘，找到我们需要的模式并依此进行决策和分析。

对于第一个问题，技术分析的理论基础提出了三个市场假设：市场行为涵盖一切信息；价格沿趋势移动；历史会重演。其中第三个假设，历史会重演是从人类的心理因素方面考虑。股票的价格由供求关系决定，而供求关系受到投资者心理行为的影

响。在这种心理状态下，市场的交易行为将趋向于一定的模式，过去出现的价格趋势变化方式，在今后也会不断出现。例如，某些成交价格模式由于在过去表现良好，投资者通常假设它在将来表现一样良好，因此他会做出买入的决策。从这个角度看来，未来的走势隐含在历史数据中，或者说未来是过去的重复。因此，我们可以回答第一个问题：即技术指标存在一定的模式可以用来对股票进行聚类。

对于第二个问题，技术指标的模式是否可以用来预测股价走势？不同的模式或模式组合对于股价预测给出的信息量是不同的。例如：股价的连续小阳线上涨，同时伴随着成交量的温和放大，这种成交价格、成交量的模式组合往往预示着未来将保持上涨趋势。而股价在小范围箱体内盘整，同时成交量出现不稳定的变化，这种模式组合对于股价预测没有指导意义。在后续的内容中我们将分别分析不同的成交价格、成交量模式对于股价预测的有效性，并验证成交价格、成交量的模式组合可以进一步提高股价预测的准确性。

基于模式聚类的短线选股模型

模式聚类算法

如前所述，基于模式聚类的短线选股模型的思想是对成交价格和成交量进行聚类，进而得到常见模式或模式组合。当需要对某只股票未来走势进行预测时，可以考察其当前成交价格和成交量符合哪一类模式或模式组合，然后根据样本内符合该模式或模式组合的样本后期走势给出具体的分类预测结果。

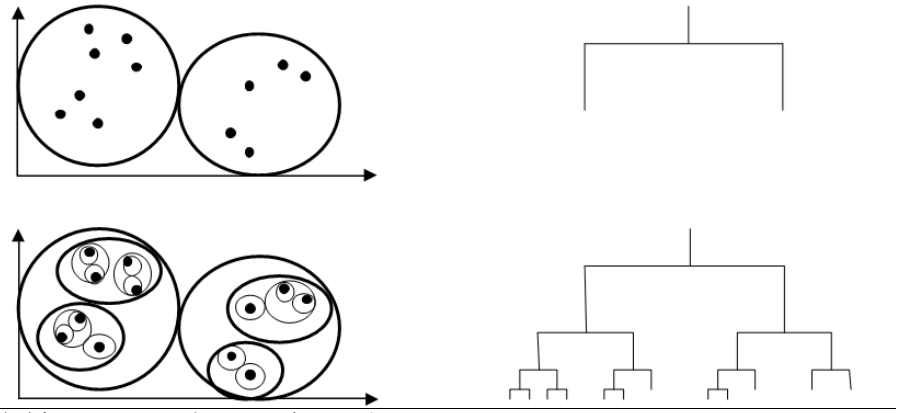
对于聚类算法，我们选择的是较为经典的 K-均值聚类算法，在给定聚类数目 K 和相似度计算方法的前提下，算法初始地选择 K 个聚类中心，并将所有的样本点划分到与其距离最近的聚类中心中，完成一次迭代过程。在更新 K 个聚类中心后，算法进行下一步的迭代过程，直到达到收敛条件。

与其它聚类算法相比，K-means 聚类算法仅需要计算每个样本点与聚类中心的相似度，而不需要两两地计算两个样本点之间的相似度。从而在样本数较大的时候，K-means 聚类算法可以提高算法的效率。

为了进一步提高聚类效果，我们在传统的 K-均值聚类算法上做了稍微的修改，并称之为层次的 K-均值聚类算法。在每一层的聚类过程中，我们将样本点分为 2 个聚类结果，这样做的好处是在每一次分裂的过程中，可以尽可能地扩大聚类与聚类之间的差异，同时降低计算复杂度。在下一层的聚类过程中，我们对上一层得到的聚类结果进行类似的进一步划分，直到达到我们需要的层数或当前聚类大小小于某个阈值为止。

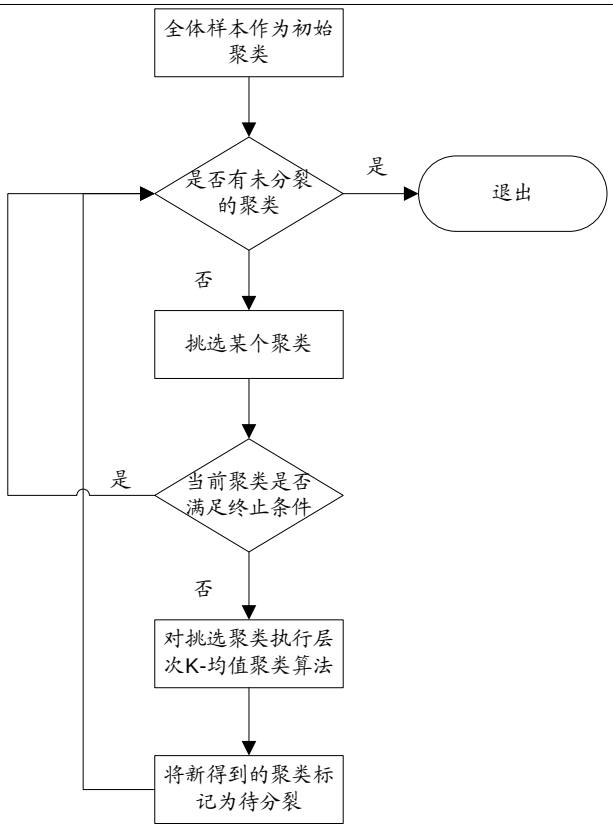
图 2 是一个简单的层次 K-均值聚类的例子，在第一层的聚类过程中，样本被划分为两个聚类（左右各一个圆圈）。在第二层的聚类过程中，左边圆圈内的样本继续分裂成两个聚类，右边圆圈内的样本也进行同样的分裂。最后我们可以得到图 2 中的层次聚类结果。层次 K-均值聚类算法的流程图如图 3 所示。

图 2：层次 K-均值聚类示例



资料来源：Google 图库，国信证券研究所整理

图 3：层次 K-均值聚类算法流程图



资料来源：Google 图库，国信证券研究所整理

股价和成交量模式的处理

股票池中选择中证 800 历史成分股，时间区间为 2005 年 3 月 28 日至 2011 年 3 月 14 日。对于每一只股票样本，选择截止到当天为止，3 个交易日内的 15 分钟线数据作为成交价格和成交量的采样。我们采用的交易策略为：以第二天开盘前 1 小时的交易量加权平均价格（**V**olume **W**eighted **A**verage **P**rice）买入，以第三天开盘前 1 小时的 VWAP 卖出，共计持有三个小时。如果第三天开盘前 1 小时的 VWAP 大于第二天开盘前 1 小时的 VWAP，那么对应的样本是获利，反之则亏损。

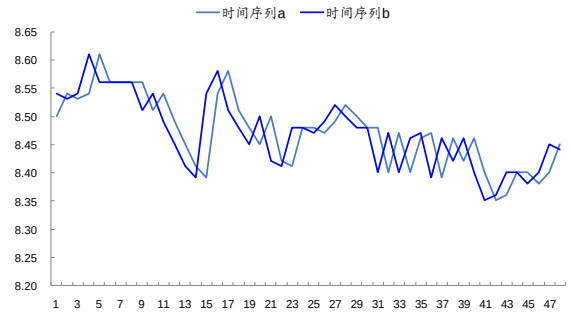
后续我们也将总结不同的交易策略下成交价格、成交量的模式。

对于成交价格，我们采用的是 15 分钟线的收盘价，后续我们将考虑开盘价、最高价、最低价等对于聚类算法的影响。通过这种数据处理方式，我们可以得到共计 48 个数据点，横坐标为时间，纵坐标表示当前的成交价格或成交量，最终形成一条成交价格或成交量随时间变化的曲线。

我们最终的目的是总结成交价格和成交量走势的常见模式，通过模式聚类算法可以将走势较为相近的样本聚集在同一个聚类结果中。在聚类过程中我们需要度量两个时间序列的相似度，直接利用传统的欧氏距离公式存在一定的缺陷。例如图 4 中的时间序列 a 和时间序列 b，如果我们直接利用传统的欧式距离度量方法，两者之间的距离较大。而我们仔细观察发现，它们的走势是非常接近的，唯一的区别在于时间序列 b 的“起始”时间点比时间序列 a 晚了一点。

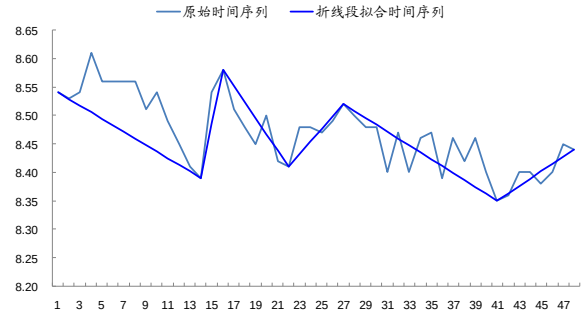
基于上述的讨论，我们有必要对数据进行预处理，提高聚类算法的效果。直观上，我们可以选取时间序列中较为重要的关键点作为对原始时间序列的拟合。这样可以在一定程度上避免由于时间相位差造成的不精确问题。如图 5 所示，对于原始时间序列，我们选取了其中的 7 个关键点，将这些关键点连成折线段作为对原始时间序列的拟合。

图 4：时间序列相似度计算



资料来源：天软科技，国信证券经济研究所整理

图 5：时间序列关键点提取



资料来源：天软科技，国信证券经济研究所整理

关键点的选取方法有很多种，在本文中，我们找到时间序列中变化最大的时间点作为关键点。例如，在图 5 中，从时间点 1 到时间点 14 大致呈现下跌的走势（从时间点 1 到时间点 4 虽然有小幅的上涨走势，但由于变化不大，因此我们不认为时间点 4 是关键点）。而从时间点 14 开始，呈现上涨的走势。因此，时间点 14 被选择为关键点。其他的关键点也可以通过类似的方法得到。为了避免成交价格绝对值的不同对于时间序列相似度计算的影响，我们将第一个关键点的成交价格归一化为 100，其它的关键点依此进行类似的处理。

通过这样的折线段拟合方法，图 4 中的两个时间序列 a 和 b 都可以被两条折线段拟合，我们再用欧式距离度量拟合后的两条折线段的相似度，作为对原始时间序列 a 和 b 相似度的近似。通过折线段拟合的方法，可以提高时间序列相似度计算的准确性。

对于成交量来说，我们计算两个关键点之间成交量的均值，得到成交量随时间变化的曲线。需要指出的是，由于原始时间序列走势的不同，最后我们得到的拟合折线段数目也会有所差异。例如，单边上涨的走势仅用一个折线段就可以进行拟合。如无特殊说明，在本文后续中的样本均包含 6 个折线段数目。后续我们也将考虑折线段数目对于选股效果的影响。

到这里，对于每个股票样本，我们可以得到它在成交价格和成交量上的折线段拟合后的时间序列。我们用欧式距离度量折线段拟合后的时间序列相似度，分别对成交

价格和成交量数据进行层次的 K-均值聚类，得到常见的成交价格和成交量模式。下面我们将分别对成交价格模式和成交量模式进行聚类分析。

股价模式中心的聚类分析

经过折线段拟合处理后，我们共计得到 121939 个成交价格时间序列样本。利用层次 K-均值聚类算法可以将这些样本划分到 198 个聚类结果中，每一个聚类结果对应一个成交价格聚类中心。我们把这 198 个聚类中心当做是成交价格的不同模式。如前所述，我们关心的是在这些成交价格的模式中，是否对于股价后期走势有判断价值。例如：某些成交价格模式预示后期将有较大概率上涨，而某些成交价格模式预示后期将有较大概率下跌。

由于短线选股模型经常遭遇持有时间参数的问题，因此我们首先对模式识别选择股票后的持有时间进行了简单分析，下表列举了选股后次日开盘一小时 vwap 买入，n（3<=n<=14）小时后同样通过一小时 vwap 成交卖出，所实现的收益之间的相关系数。从表中我们看到，持有时间从 3 小时延长到 7 小时（持有 1 个交易日至 2 个交易日时），收益相关系数衰减到 0.7，为了减少计算量，我们首先以 3 个小时为例进行模式聚类和统计分析。

表 1：持有时间变化后的实现收益相关系数矩阵

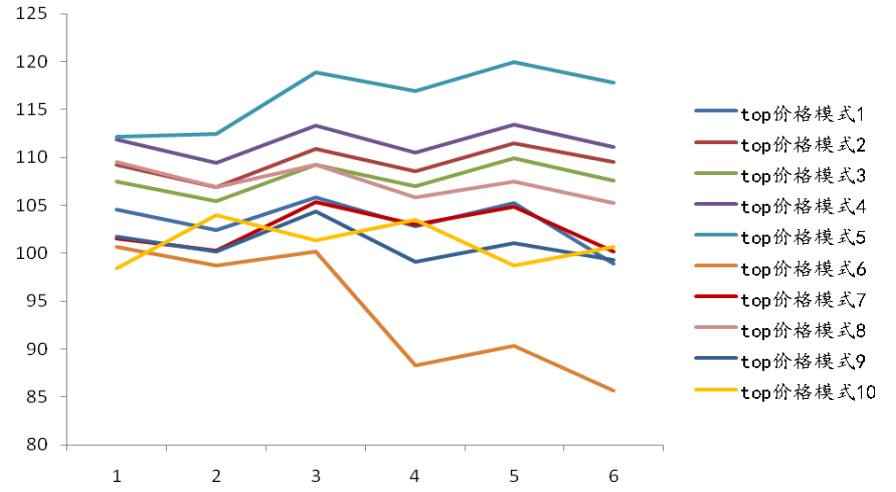
相关系数	持有 3h	持有 4h	持有 5h	持有 6h	持有 7h	持有 8h	持有 9h	持有 10h	持有 11h	持有 12h	持有 13h	持有 14h
持有 3h	1.00	0.94	0.88	0.82	0.71	0.67	0.65	0.63	0.58	0.55	0.54	0.54
持有 4h	0.94	1.00	0.95	0.89	0.78	0.74	0.71	0.70	0.64	0.61	0.60	0.60
持有 5h	0.88	0.95	1.00	0.96	0.85	0.81	0.77	0.75	0.68	0.66	0.64	0.64
持有 6h	0.82	0.89	0.96	1.00	0.91	0.87	0.84	0.81	0.73	0.70	0.68	0.67
持有 7h	0.71	0.78	0.85	0.91	1.00	0.97	0.93	0.90	0.82	0.79	0.77	0.76
持有 8h	0.67	0.74	0.81	0.87	0.97	1.00	0.97	0.94	0.86	0.83	0.81	0.80
持有 9h	0.65	0.71	0.77	0.84	0.93	0.97	1.00	0.97	0.90	0.87	0.85	0.83
持有 10h	0.63	0.70	0.75	0.81	0.90	0.94	0.97	1.00	0.94	0.91	0.89	0.87
持有 11h	0.58	0.64	0.68	0.73	0.82	0.86	0.90	0.94	1.00	0.98	0.95	0.93
持有 12h	0.55	0.61	0.66	0.70	0.79	0.83	0.87	0.91	0.98	1.00	0.98	0.95
持有 13h	0.54	0.60	0.64	0.68	0.77	0.81	0.85	0.89	0.95	0.98	1.00	0.98
持有 14h	0.54	0.60	0.64	0.67	0.76	0.80	0.83	0.87	0.93	0.95	0.98	1.00

资料来源：天软科技，国信证券经济研究所整理

在给定某个成交价格模式后，我们可以计算全体样本与该成交价格模式的距离。距离越近的样本被认为是与该成交价格模式越为相似。如若这些样本后期都呈现较为稳定的上涨或下跌，那么，这样的成交价格模式是我们需要的。图 6 为仅考虑价格模式的情况下，次日开盘一小时 vwap 买入，再次日开盘一小时 vwap 卖出实现收益最高的 10 种价格模式。

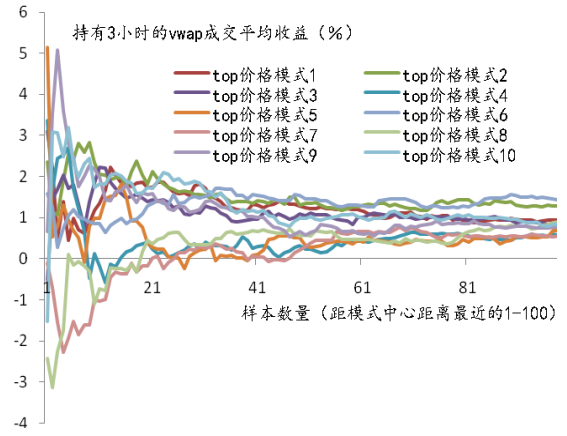
这 10 种模式仅从直观上观察并未体现出很强的价格动量，有一种模式甚至在观察期中持续下跌，股价从未高于起始点（100）。

图 6：预期上涨的前 10 种成交价格模式



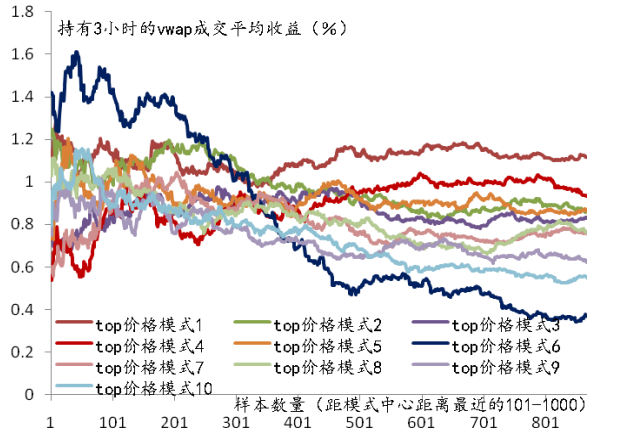
资料来源：天软科技，国信证券经济研究所整理

图 7：预期上涨的前 10 种价格模式前 100 个样本收益分布



资料来源：天软科技，国信证券经济研究所整理

图 8：预期上涨的前 10 种价格模式 100-1000 个样本收益分布



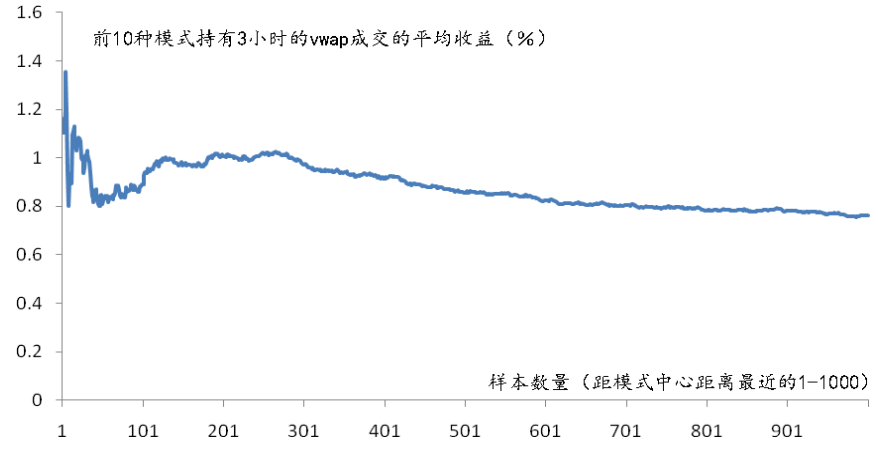
资料来源：天软科技，国信证券经济研究所整理

从图 7 中可以看到，在样本数少于 100 的情况下，仅考虑价格模式，即便是统计上最强的 10 种模式，仍然存在少量负收益样本，图 8 显示，在样本数量上升到 100 之后，从 100-1000 个样本数量的区间，只有一种模式的平均收益缓慢下降至跌破 0.6%的交易成本线，其余均稳定在成本线之上。

（本文中所使用的样本数量，均按照与聚类中心的距离排序，x 个样本数量，即：全样本内与该聚类中心的距离排序在前 x 的样本）

图 9 显示了预期上涨的前 10 种成交价格模式的平均收益曲线图，从图中我们可以看出，平均收益曲线较为稳定地保持在成本线之上。

图 9：预期上涨的前 10 种成交价格模式前 1000 个样本的平均收益曲线



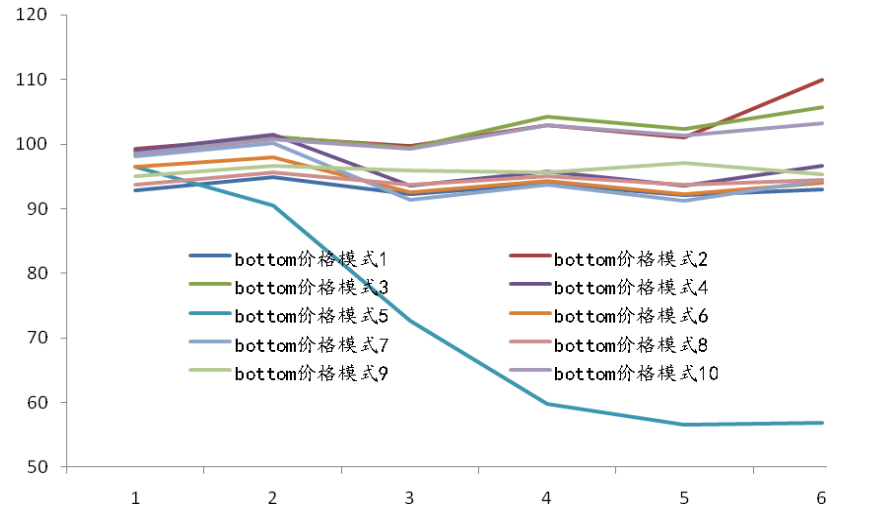
资料来源：天软科技，国信证券经济研究所整理

对比图 6 和图 10，我们可以看到，预期下跌的前 10 种模式仅从直观上观察呈现出与图 6 相反的形态，其中也有一种模式甚至在观察期中持续下跌，股价从未高于起始点（100）。

我们同样对这 10 种模式的前 100 个样本、101-1000 个样本统计了持有 3 小时的收益统计（持有期之前 1 小时 vwap 买入，持有期之后 1 小时 vwap 卖出），从图 11、12 我们观察到，伴随样本数量的增加，下跌形态对收益的解释程度下降较快，其中大多数都没有在 3 小时内实现大于交易成本（0.6%）的跌幅。从平均线（图 13）上看更为显著，在每种模式 50 个样本数量以上，平均收益曲线就未能低于交易成本线（0.6%）。

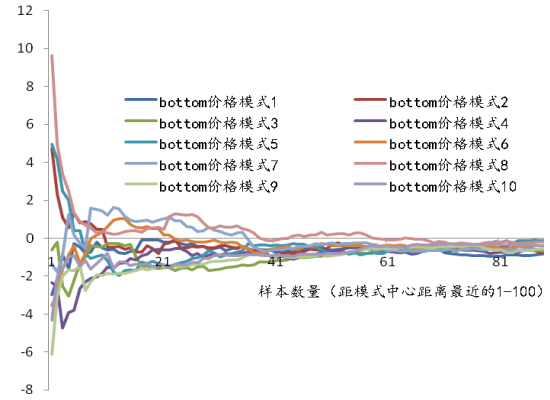
关于预期上涨和预期下跌的模式中心在统计结论上有如此显著的差异，成因无从猜测，但我们相信这一现象与个股无法实现短线卖空有较大的关系。

图 10：预期下跌的前 10 种成交价格模式



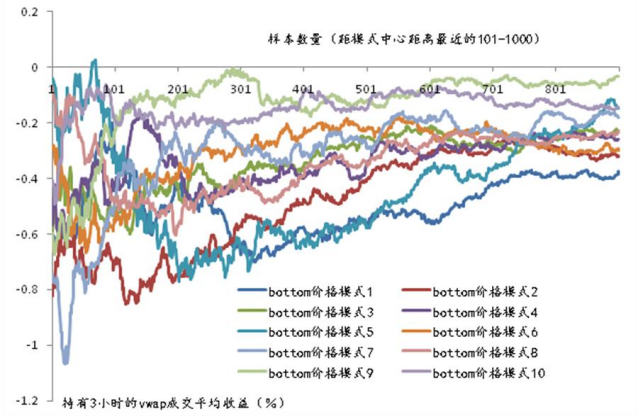
资料来源：天软科技，国信证券经济研究所整理

图 11：预期下跌的前 10 种价格模式 100 个样本收益分布



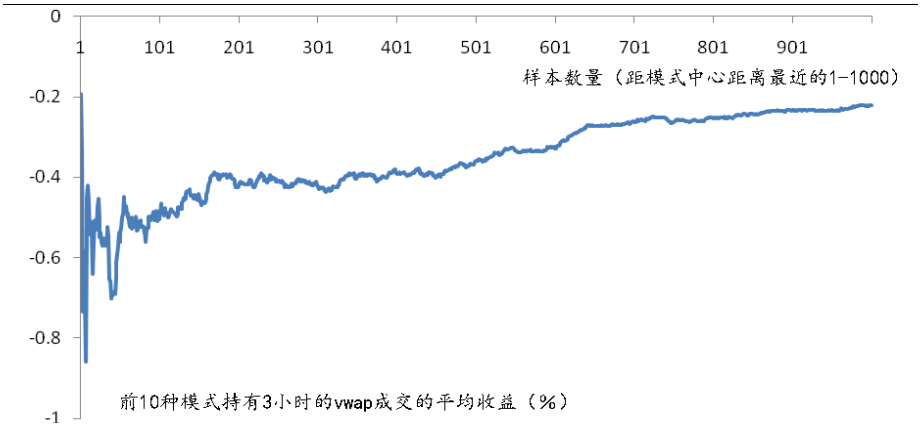
资料来源：天软科技，国信证券经济研究所整理

图 12：预期下跌的前 10 种价格模式 100-1000 个样本收益分布



资料来源：天软科技，国信证券经济研究所整理

图 13：预期下跌的前 10 种成交价格模式前 1000 个样本的平均收益曲线



资料来源：天软科技，国信证券经济研究所整理

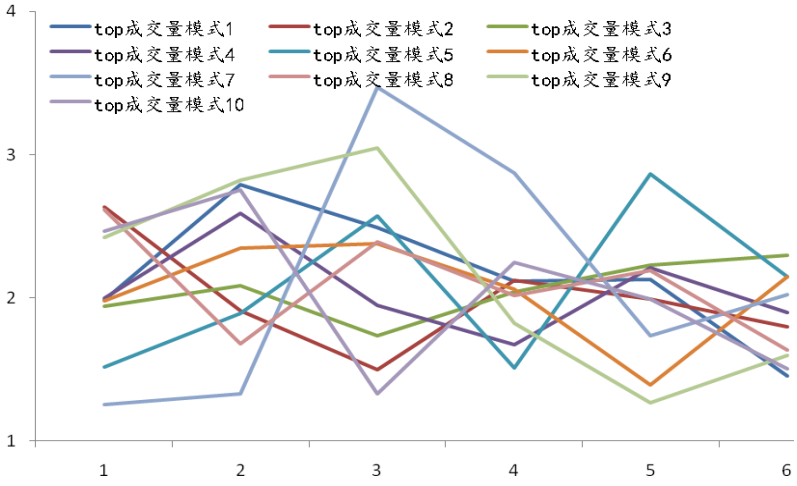
成交量模式中心的聚类分析

经过折线段拟合处理后，我们共计得到 121939 个成交量时间序列样本。我们对折线段两端内的时间进行成交量平均，形成新的成交量线段，同样利用层次 K-均值聚类算法，可以将这些样本划分到 194 个聚类结果中。

与成交价格的初步研究类似，我们希望通过聚类分析，找到一些对后市持有收益具有较强解释力的模式。

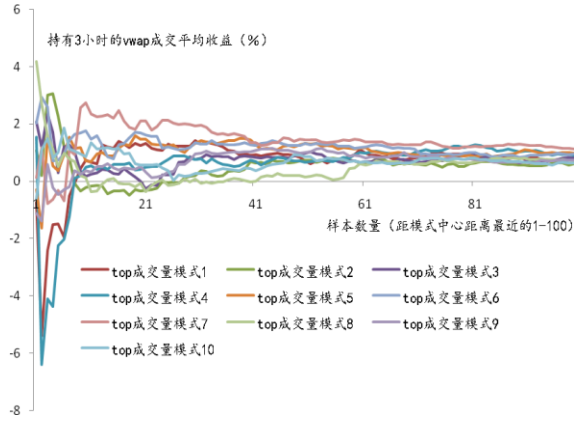
与价格中心相比，图 14 给出的信息非常混乱。如果说图 6、图 10 中所示的 10 种价格形态在直观上至少存在一定的类似之处，那么图 14 完全没有类似的暗示。这一结果似乎与投资人的习惯或直觉有关——异常的交易量确实会引起注意，但是往往是在某些特殊的时点，或是在价格形态形成了某种其他暗示之后，这一信号才会确实推动投资者进行买卖，很少有完全基于成交量做出的决策。

图 14：预期上涨的前 10 个成交量模式



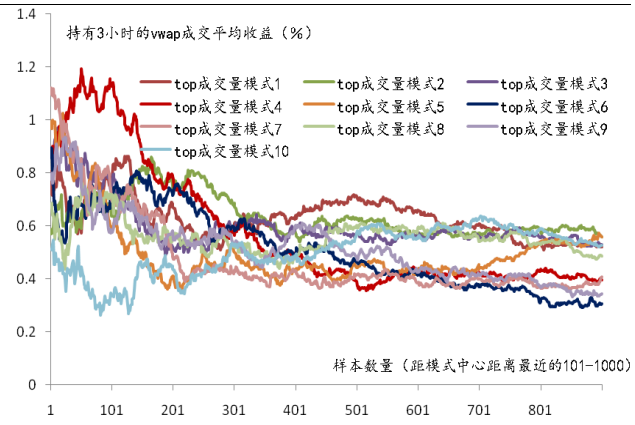
资料来源：天软科技，国信证券经济研究所整理

图 15：预期上涨的前 10 种量模式 100 个样本收益分布



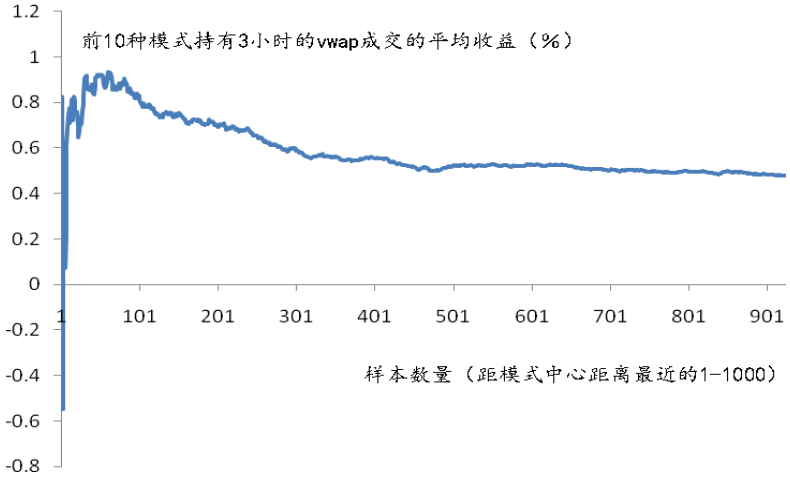
资料来源：天软科技，国信证券经济研究所整理

图 16：预期上涨的前 10 种量模式 100-1000 个样本收益分布



资料来源：天软科技，国信证券经济研究所整理

图 17：预期上涨的前 10 种成交量模式前 1000 个样本的平均收益曲线

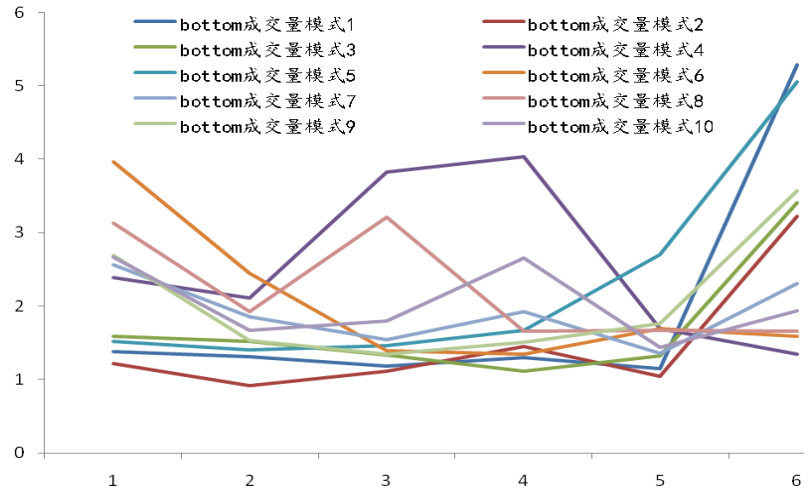


资料来源：天软科技，国信证券经济研究所整理

图 15、16、17 给出的结论与上述直观感觉是一致的，在聚类数量差不多的情况下，成交量中心的前十对后市持有收益的解释度明显低于价格中心。在样本数量少于 10 的情况下，竟然出现了平均的负收益，而在 300 个样本之外，10 种模式的平均持有收益就已经低于成本线（0.6%）。由此我们可以定性地得到一个结论——即便是经验上对于后市收益有很强解释的成交量变化，也可能有很大的偶然性，在样本数不足的情况下，尤其不能以成交量作为后市走势的唯一判断依据。

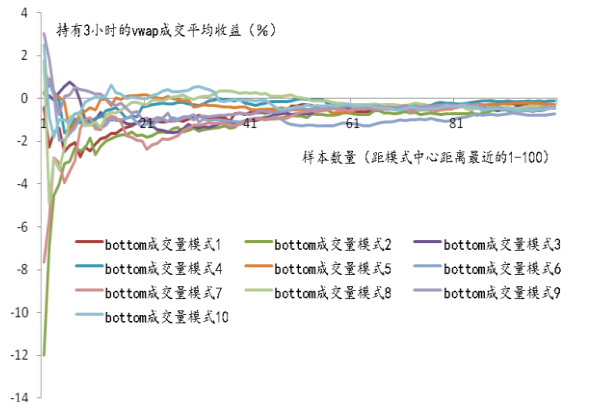
较为意外的是，预期下跌的前 10 个成交量模式出现了较为一致的共同特征——观察期末端的成交量放大。对比预期下跌的前 10 个价格中心，我们从图 21 中可以看到，在 50 个样本以外，10 种模式的平均持有跌幅就已经小于交易成本（0.6%），并且伴随着样本数量增加，跌幅迅速收敛，至 1000 个收益时，跌幅已经基本接近于 0（全样本的平均涨跌幅是 0.18%）。

图 18：预期下跌的前 10 个成交量模式



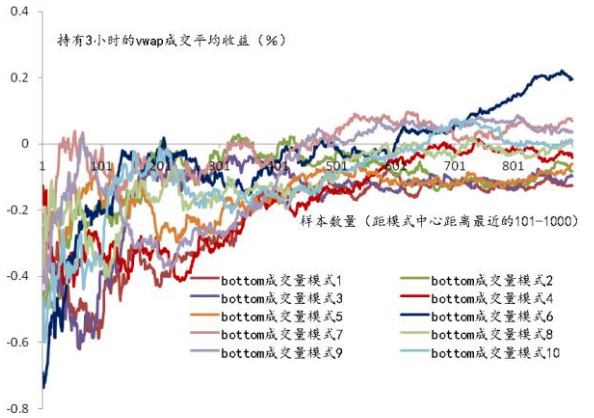
资料来源：天软科技，国信证券经济研究所整理

图 19：预期下跌的前 10 种量模式 100 个样本收益分布



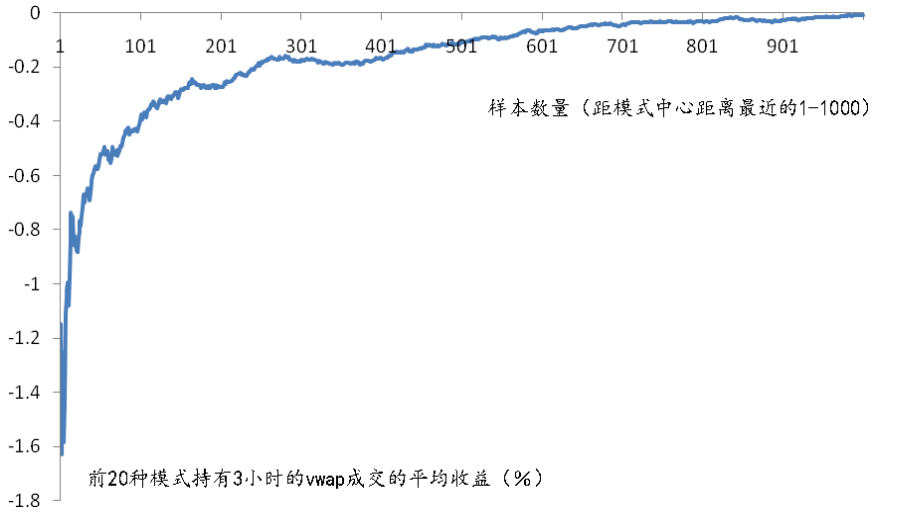
资料来源：天软科技，国信证券经济研究所整理

图 20：预期下跌的前 10 种量模式 100-1000 个样本收益分布



资料来源：天软科技，国信证券经济研究所整理

图 21：预期下跌的前 10 种成交量模式前 1000 个样本的平均收益曲线



资料来源：天软科技，国信证券经济研究所整理

总结上述四类中心（分别为预期上涨和下跌的价格、成交量聚类中心），我们可以看到，预期上涨的价格模式中心，解释力是最强的，伴随样本数量增加，收益普遍较为稳定，只有个别形态下降。预期上涨的成交量模式稳定性次之。预期下跌形态的价格模式中心解释力相对较弱，伴随样本数量增加，收益普遍缓慢回升，预期下跌的成交量模式也呈现出同样的规律并且收敛速度大于价格中心。

模式组合

以距离为参数进行的价量模式组合

前文中我们通过聚类算法得到了单一中心情况下对后市收益解释最高的各 10 种成交价格和成交量的模式。实际应用中，价格分析往往是与成交量相结合的，但是，两种模式各以多大的权重影响决策？为了解决这一问题，我们引入了 Average Precision 指标[①]。

首先，我们需要穷尽式地枚举出所有可能的量价组合，并分析每种组合条件下的收益率情况。具体地来说，在给定某个成交价格模式 $Price_i$ 和成交量模式 $Volume_j$ 的前提下，我们可以对每一个股票样本 k 进行排序打分，其依据为股票样本 k 与模式 $Price_i$、模式 $Volume_j$ 的相似度，假设我们得到的排序分别为 $Rank_{ik}$ 和 $Rank_{jk}$，那么在这个模式组合的前提下，该股票样本的最终得分可以定义为：

$$Score = lambda * Rank_{ik} + (1 - lambda) * Rank_{jk}$$

其中：0<=lambda<=1，当 lambda=1 时，即完全只考虑样本与价格模式中心的距离；lambda=0 时，即完全只考虑样本与成交量模式中心的距离。

如果某个模式组合根据这个组合打分公式可以将上涨（下跌）概率大的样本尽早返回，那么我们认为这个模式组合是 top（bottom）组合（分别预示上涨和下跌）。我们可以根据得分的大小得到最有效的 20 个 top（bottom）组合（以 Average Precision 指标为排序依据）。

表 2：价量模式结合 top20 组合相关信息

成交价格模式编号（1-198）	成交价格模式的 Average Precision 值	成交量模式编号（1-194）	成交量模式的 Average Precision 值	lambda	Average Precision 值
177	0.296	95	0.069	0.9	0.852
157	0.685	126	0.353	0.8	0.799
112	0.513	79	-0.031	0.9	0.792
179	0.339	174	-0.002	0.9	0.790
111	0.608	79	-0.031	0.8	0.787
98	0.312	40	0.132	0.7	0.775
98	0.312	39	0.336	0.6	0.775
157	0.685	121	0.315	0.9	0.775
157	0.685	53	0.272	0.9	0.775
157	0.685	39	0.336	0.8	0.773
157	0.685	125	0.171	0.8	0.770
157	0.685	117	0.011	0.9	0.766
157	0.685	37	0.092	0.8	0.764
157	0.685	144	0.320	0.9	0.764
112	0.513	71	0.134	0.8	0.764
157	0.685	120	0.073	0.9	0.763
177	0.296	93	-0.025	0.9	0.759
179	0.339	175	0.076	0.8	0.758
95	0.441	40	0.132	0.5	0.755
157	0.685	116	0.236	0.8	0.752

资料来源：天软科技，国信证券经济研究所整理

表 3：价量模式结合 bottom20 组合相关信息

成交价格模式编号（1-198）	成交价格模式的 Average Precision 值	成交量模式编号（1-194）	成交量模式的 Average Precision 值	lambda	Average Precision 值
120	-0.288	37	0.082	0.9	-0.793
120	-0.288	39	0.277	0.8	-0.790
120	-0.288	141	0.040	0.9	-0.757
120	-0.288	38	0.168	0.9	-0.752
120	-0.288	144	0.260	0.9	-0.736
120	-0.288	31	0.238	0.9	-0.714
120	-0.288	30	0.261	0.9	-0.692
120	-0.288	32	0.228	0.9	-0.688
120	-0.288	40	0.114	0.8	-0.671
120	-0.288	7	0.141	0.9	-0.646
120	-0.288	147	0.026	0.9	-0.636
120	-0.288	145	-0.026	0.9	-0.636
120	-0.288	52	0.217	0.9	-0.612
120	-0.288	56	0.241	0.9	-0.612
120	-0.288	116	0.178	0.9	-0.601
120	-0.288	53	0.230	0.9	-0.595
120	-0.288	138	0.216	0.9	-0.591
120	-0.288	43	0.180	0.9	-0.585
120	-0.288	125	0.138	0.9	-0.584
120	-0.288	142	0.101	0.9	-0.583

资料来源：天软科技，国信证券经济研究所整理

Top20 价量模式组合

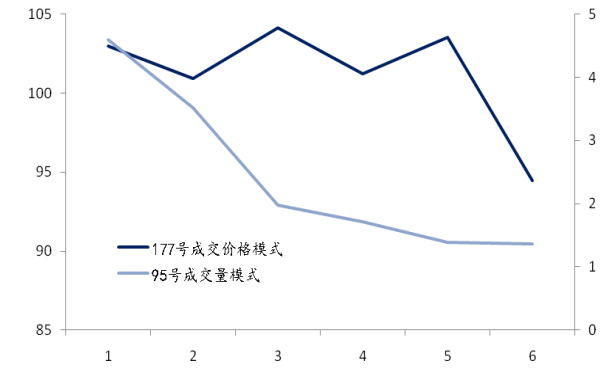
以表 2 中排序第一的模式组合（177 号价格中心、95 号成交量中心，lambda=0.9）为例，以下几张图表列举了价量中心结合的研究过程。

首先，针对任何一种价格中心和量中心，通过调整 lambda，我们可以得到 11 个

Average Precision 值，其中 AP 值最大的 lambda 告诉我们在这种模式组合中，价格中心和量中心距离的权重分别是多少。从排序在前 20 的组合中我们可以看到，无论价量中心分别是什么，AP 值的峰值基本出现在 lambda 值在 0.8-0.9 的区间，所有 20 个模式中的 lambda 值均大于 0.5。也就是说，在模式组合中，价格中心占据了绝对优势，量中心的权重相对而言要小得多。

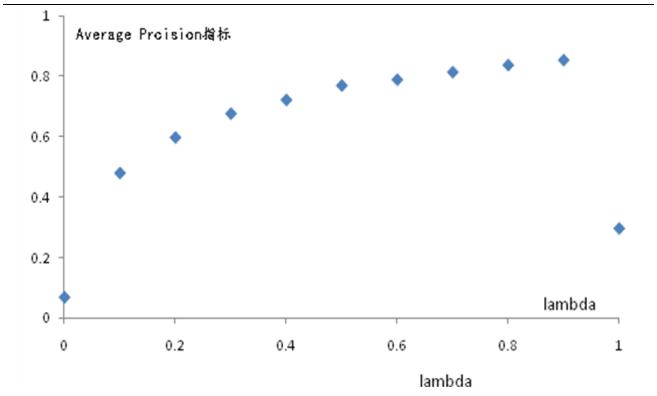
通过 198 个价格中心和 194 个量中心的排列组合，我们得到了 38412 种排列。再通过 11 个 lambda 值的参数调整，排列数量变为 422532。表 2 所示的 20 种模式，即上述 422532 种模式中 AP 值最高的 20 种。考虑到样本总数为 121939，我们可以得知每一个样本可能同时属于 3-4 种排列组合的结果。原因很简单：每一种中心在不同的距离半径和离散程度下，可能聚合到重合样本，后文我们将对它们的合集进行分析。

图 22：177 号价格中心和 95 号成交量中心



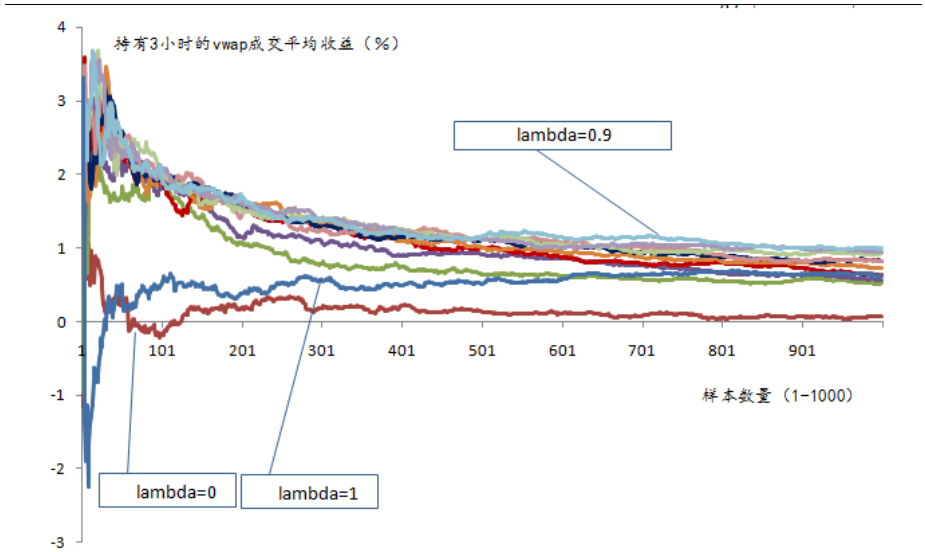
资料来源：天软科技，国信证券经济研究所整理

图 23：lambda 与 Average Precision 值



资料来源：天软科技，国信证券经济研究所整理

图 24：177 号价格中心和 95 号成交量中心在不同 lambda 值下的样本收益



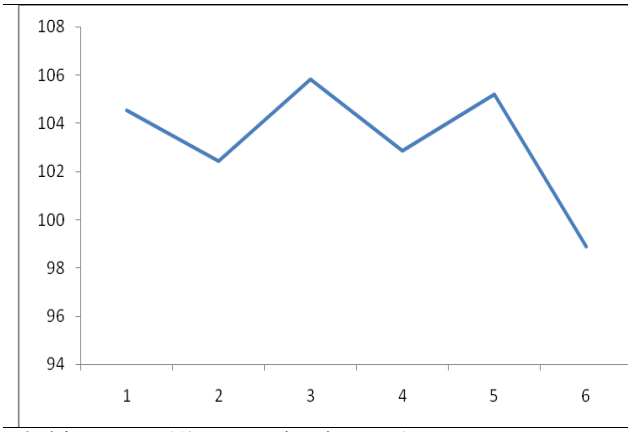
资料来源：国信证券研究所

这一案例是一种特别振奋人心的结果——从图 24 中可以看到 177 号价格中心和 95 号成交量中心的前 1000 个样本收益并不高，但通过 0.9 的系数结合之后得到了非常大的改善，证明价格形态和成交量变化在某种结合形式下可以得到更好的结果。

此外还有一个现象值得注意，我们观察到，价格中心的 AP 值分布区间为-0.216 至 0.685，其中 AP 值为 0.685 的 157 号价格中心频繁出现在前 20 的模式组合中，

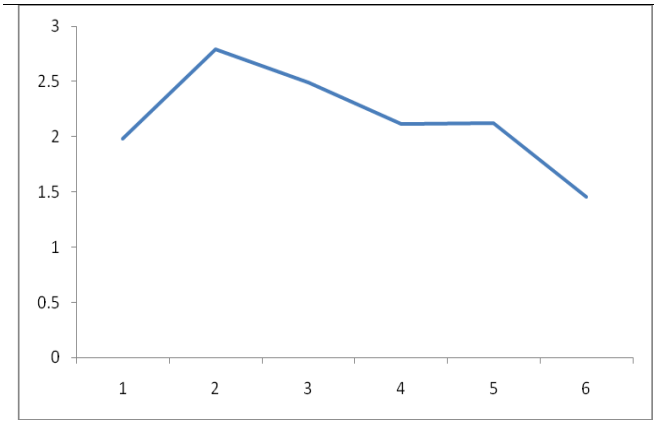
但成交量中心无论从排序还是从重复次数上，都未出现特别有凝聚力的模式中心（成交量中心的 AP 值分布区间为-0.115 至 0.353）。另外，AP 值最高的价格中心（157 号）和 AP 值最高的成交量中心（126 号）通过 0.8 的 lambda 值结合，也表现出非常强的收益解释度，AP 值在所有模式组合中排序第二。

图 25：157 号价格中心



资料来源：天软科技，国信证券经济研究所整理

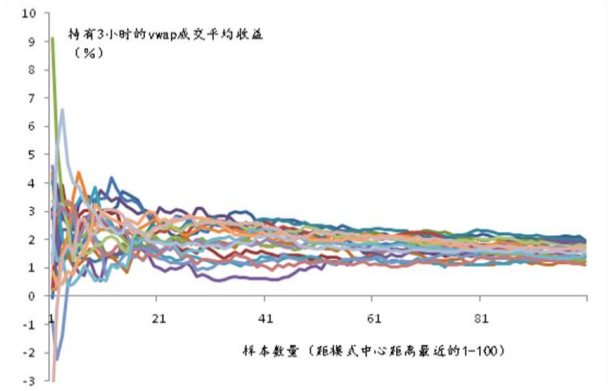
图 26：126 号成交量中心



资料来源：天软科技，国信证券经济研究所整理

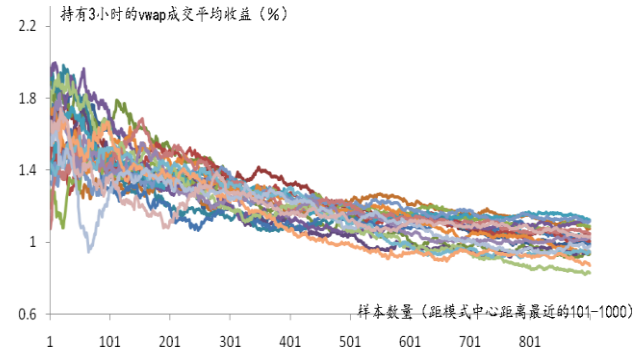
检查 top20 组合的收益情况，我们可以看到，无论是相对图 7-9 的 top10 价格中心，还是相对图 15-17 的 top10 成交量中心，均有了很大的改善。

图 27：top20 模式组合前 100 个样本收益分布



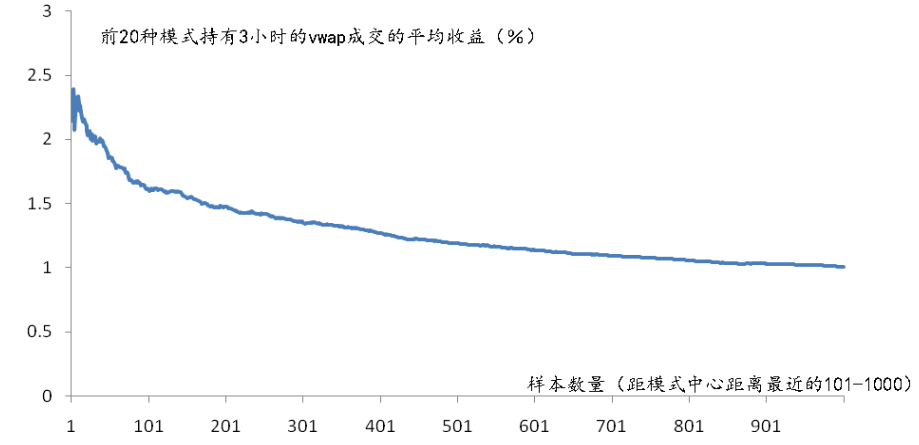
资料来源：天软科技，国信证券经济研究所整理

图 28：top20 模式组合 100-1000 个样本收益分布



资料来源：天软科技，国信证券经济研究所整理

图 29：top20 模式组合前 1000 个样本的平均收益曲线



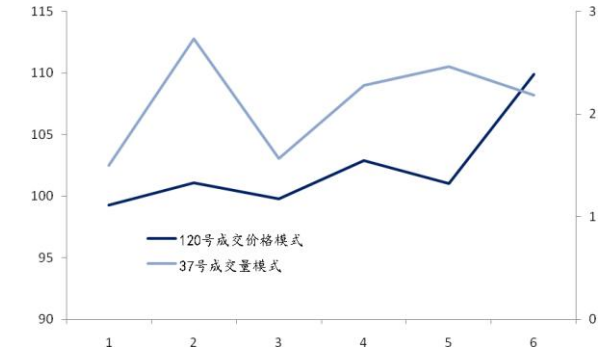
资料来源：天软科技，国信证券经济研究所整理

但是需要注意的是，这其中有大量的重复样本，后文我们将对合集（去除重复样本后的胜率和准确率进行详细分析）。

Bottom20 价量模式组合

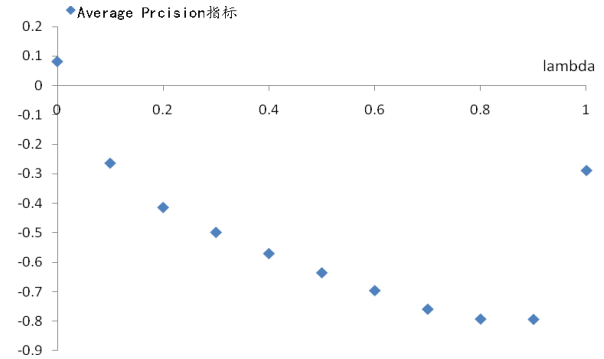
从表 3 所示的 bottom20 组合中我们看到了一个十分意外的现象：所有的组合竟然都选到了同一个价格模式：120 号。价格中心的 AP 值分布区间为-0.297 至 0.440，在 198 个价格中心中，120 号的 AP 值为倒数第二。但成交量中心无论从排序还是从重复次数上，同样未出现特别有凝聚力的模式中心（成交量中心的 AP 值分布区间为-0.167 至 0.288）。

图 30：120 号价格中心和 37 号成交量中心



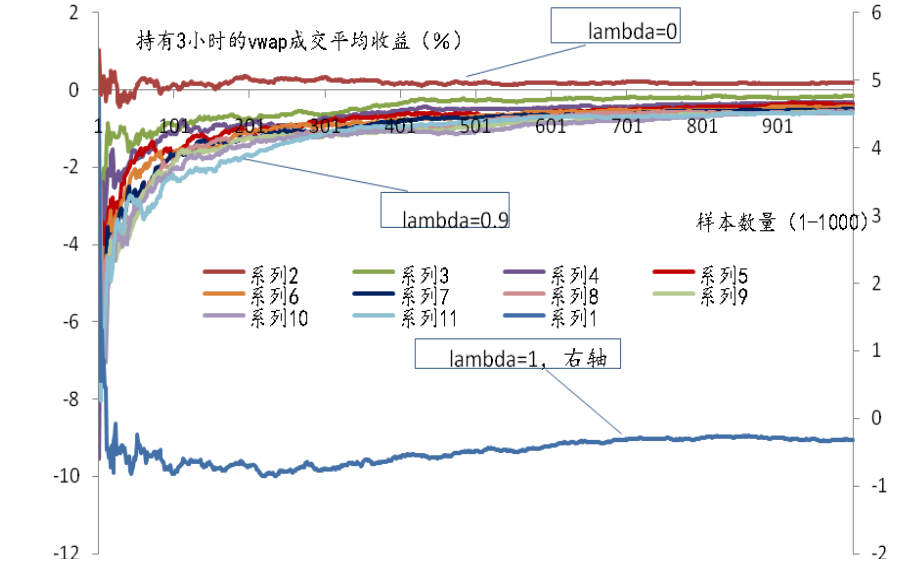
资料来源：天软科技，国信证券经济研究所整理

图 31：lambda 与 Average Precision 值



资料来源：天软科技，国信证券经济研究所整理

图 32：120 号价格中心和 37 号成交量中心在不同 lambda 值下的样本收益

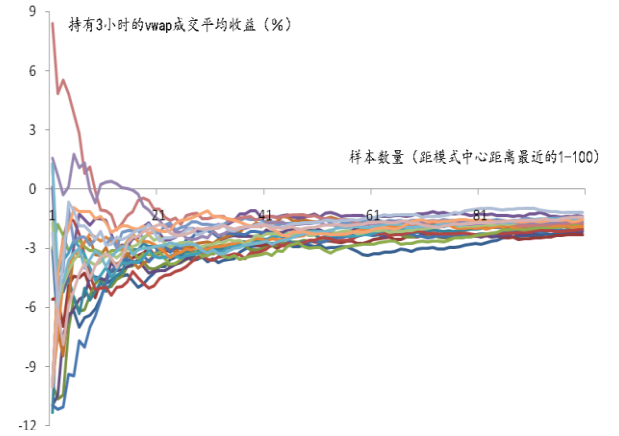


资料来源：国信证券研究所

我们观察到，在 bottom20 组合中，无论是哪一种成交量中心，与 120 价格中心结合的最佳 lambda 值均是 0.9（只有一个成交量中心是 0.8），可见 120 价格中心的形态表现出非常强的短线预跌能力。

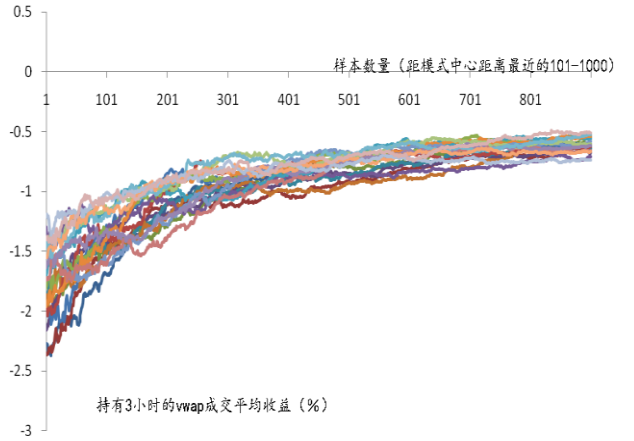
检查 bottom20 组合的收益情况，我们可以看到，无论是相对图 11-13 的 bottom10 价格中心，还是相对图 19-21 的 bottom10 成交量中心，均有了很大的改善，一直到样本数量接近 1000，3 小时实现跌幅均超过了成本线。

图 33：bottom20 模式组合前 100 个样本收益分布



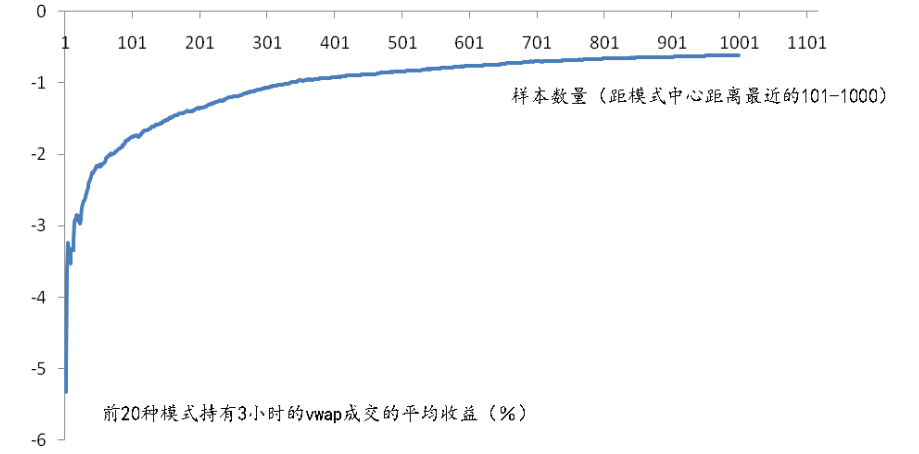
资料来源：天软科技，国信证券经济研究所整理

图 34：bottom20 模式组合 100-1000 个样本收益分布



资料来源：天软科技，国信证券经济研究所整理

图 35：bottom20 模式组合前 1000 个样本的平均收益曲线



资料来源：天软科技，国信证券经济研究所整理

综上所述，价格中心在本模型中占据着主导位置，价量模式结合之后，无论对 top 还是 bottom 组合的稳定性均有很大的提高，这一结论对我们未来将本模型的思路应用到各种实证检验的工作，均有不可忽视的指导意义。

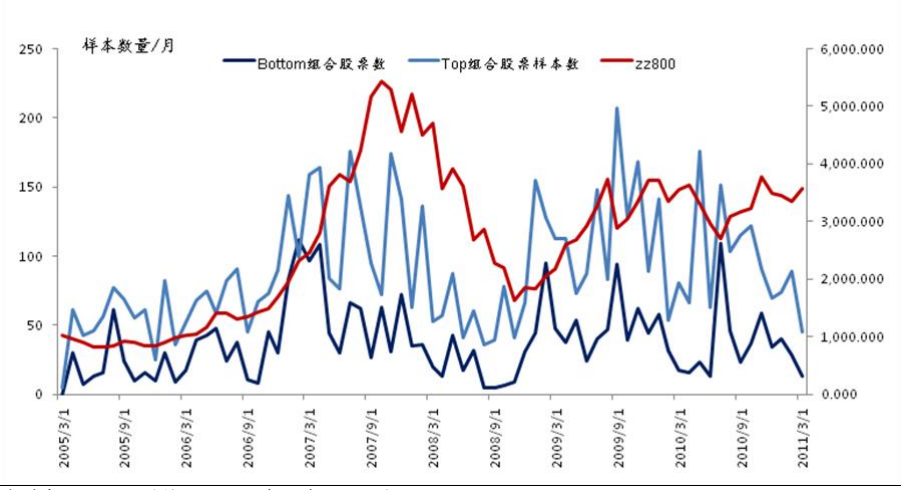
模型应用分析

Top、Bottom 组合在时间序列上的样本数分布

短线选股模型的特点之一就是：操作较为频繁，胜率预期不过高，以覆盖交易成本为首要目标。因此，我们统计了样本内 Top、Bottom 组合的样本数量在时间序列上的分布，以此观察该策略是否会出现间歇性的交易间断或数量过剩。下文所述样本数量均为去除了重复样本后的合集数量。

首先，我们按月统计了模式组合中 Top20 和 Bottom20 在样本内出现的次数。

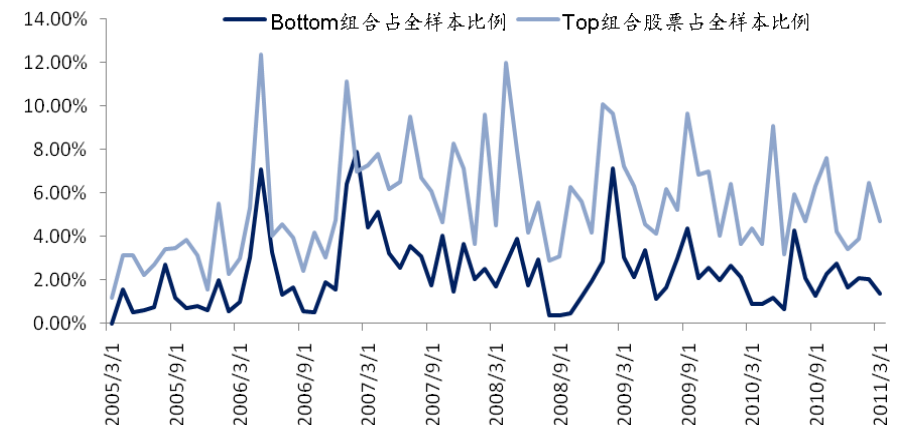
图 35：Top20 和 Bottom20 模式组合样本数时间序列分布



资料来源：天软科技，国信证券经济研究所整理

从图中看两类组合的样本数量与指数活跃程度有一定的相关性，伴随指数长期上涨或下跌，样本数量会出现同步的增减，但波动程度远远大于指数的波动。平均而言每月出现的多头样本数量在 50-100 个之间。而从两类组合样本数占全样本的比例来看，Top20 模式组合占全样本的比例在 6-8%左右，而 Bottom20 模式组合要相对小一些，普遍在 2%附近。

图 36：Top20 和 Bottom20 模式组合占全样本比例



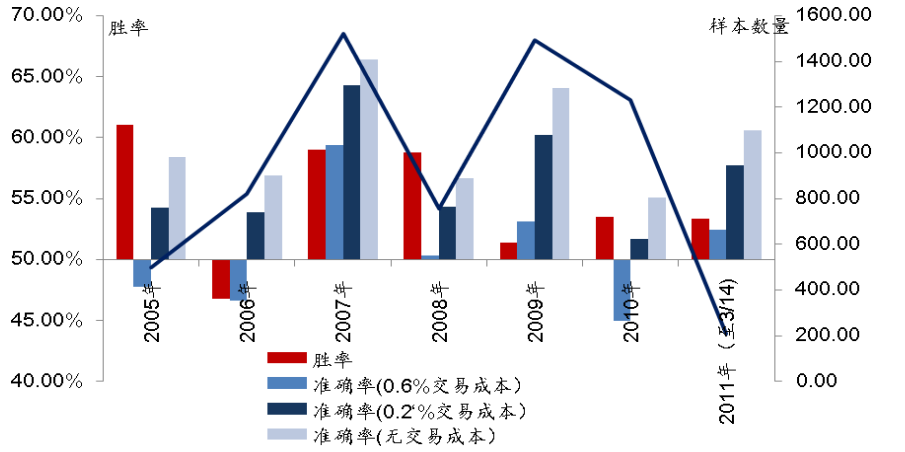
资料来源：天软科技，国信证券经济研究所整理

Top、Bottom 组合的胜率

在之前的分析中我们提到过，每一种中心在不同的距离半径和离散程度下，可能聚合到重合样本，因此我们对 Top20 和 Bottom20 在样本内的有效样本取了并集后，去除了重复样本，重新计算胜率。

对于 Top 组合，并集样本供 6527 个。我们分别统计了样本在 0.6%、0.2%、0 交易成本下的准确率，以及组合战胜同期中证 800 指数的胜率（此处为选股决策次日的指数涨跌幅）。从统计结果中可以看到，只有 2005、2006 和 2010 年样本在 0.6%的交易成本下准确率未能高于 50%，其余指标均高于 50%。6527 个样本在 vwap 成交模式下持有 3h 的平均收益率 0.833%。

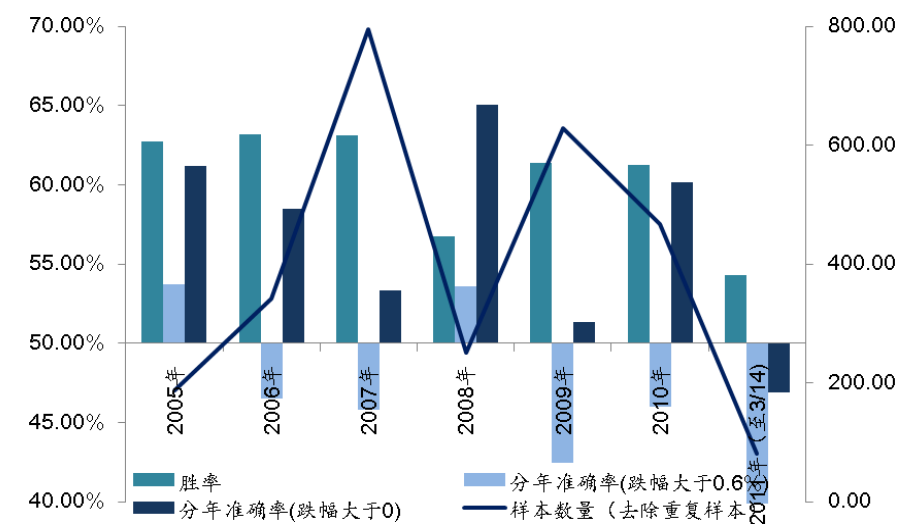
图 37：Top20 模式组合样本并集胜率



资料来源：天软科技，国信证券经济研究所整理

Bottom 组合的效果与未去除重复样本时一致，要显著差于 Top 组合。并集样本供 2753 个；在 vwap 成交模式下持有 3h 的平均收益率-0.380%。我们分别统计了样本在 0.6%、0 交易成本下的准确率，以及组合跑输同期中证 800 指数的胜率（此处为选股决策次日的指数涨跌幅）。从统计结果中可以看到，只有 2005、2008 年年样本在 0.6%的交易成本下准确率高于 50%。

图 38：Bottom20 模式组合样本并集胜率



资料来源：天软科技，国信证券经济研究所整理

样本外区间的检验结果

以 2011 年 3 月 15 日至 8 月 20 日中证 800 成分股的样本为样本外分析池，我们整理了 8617 个有效样本进行上述 Top 和 Bottom 模式组合的样本外分析。具体样本由每日滚动的三个交易日样本加入大样本池，保留其中进入各模式中心前 1000 名的样本，作为样本外策略的选中品种。

由于样本外区间整体处于指数的下跌趋势中，Top 组合遭遇重创，表现不佳，无论是胜率还是准确率，都低于样本内的平均水平，但是 Bottom 组合反之，具体样本统计见下表。

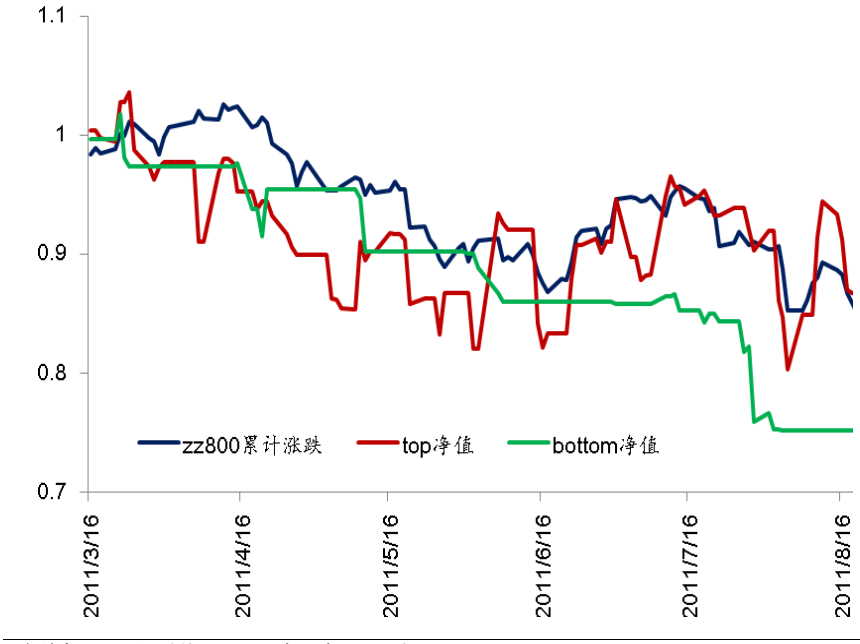
表 4：样本外 Top20 和 Bottom20 的样本统计

	样本总数	涨（跌）幅>0.6%	涨（跌）幅>0.2%	涨（跌）幅>0	有样本的交易日数	涨（跌）幅>zz800 指数
top20 模式组合	364	105	131	142	78	37
比例	100.00%	28.85%	35.99%	39.01%	100.00%	47.44%
bottom20 模式组合	89	44	50	57	42	31
比例	100.00%	49.44%	56.18%	64.04%	100.00%	73.81%

资料来源：天软科技，国信证券经济研究所整理

由于样本数量较少，因此我们通过策略净值对比的形式，来观察 top 和 bottom 组合的表现，从图中可以看到，bottom 组合在每次交易中几乎没有误判，而 top 组合则在 3-6 月的时间出现了较大的反复，三季度则表现尚可。

图 39：Top20 和 Bottom20 模式组合样本外策略净值对比



资料来源：天软科技，国信证券经济研究所整理

应用和扩展

优化讨论

本文提出了一种利用模式聚类的思路构建选股模型。在分别对成交价格、成交量进行聚类后，结合合适的评价指标，我们可以分析每一种聚类模式或模式组合的选股效果。本文中我们提出的评价指标是 Average Precision，后续我们也将考虑其他的评价指标，例如：准确率、胜率等。不同的评价指标对于模式或模式组合的评价有较为重要的影响。从评价指标角度，未来我们将针对“模式识别”设计更加合理的目标函数来评价模式中心。

从研究方法上，选股模型本质上是对于市场所有股票的排序，每个股票样本在成交价格、成交量等技术指标以及所属行业、股本大小等基本面指标上有不同的特征。从这个角度看来，选股模型在一定程序上与信息检索任务有天然的相似性。对于信息检索任务来说，我们需要返回与查询词相关的文档。在信息检

索领域，已经有很多的模型用于改进检索模型。近些年，有研究者着重研究“Learning to Rank”模型，其本质思想在于利用机器学习的方法融合样本各种各样的特征，结合成熟、深厚的机器学习理论提高检索效果。后续，我们也将从其中吸纳养分，改进我们的选股模型。

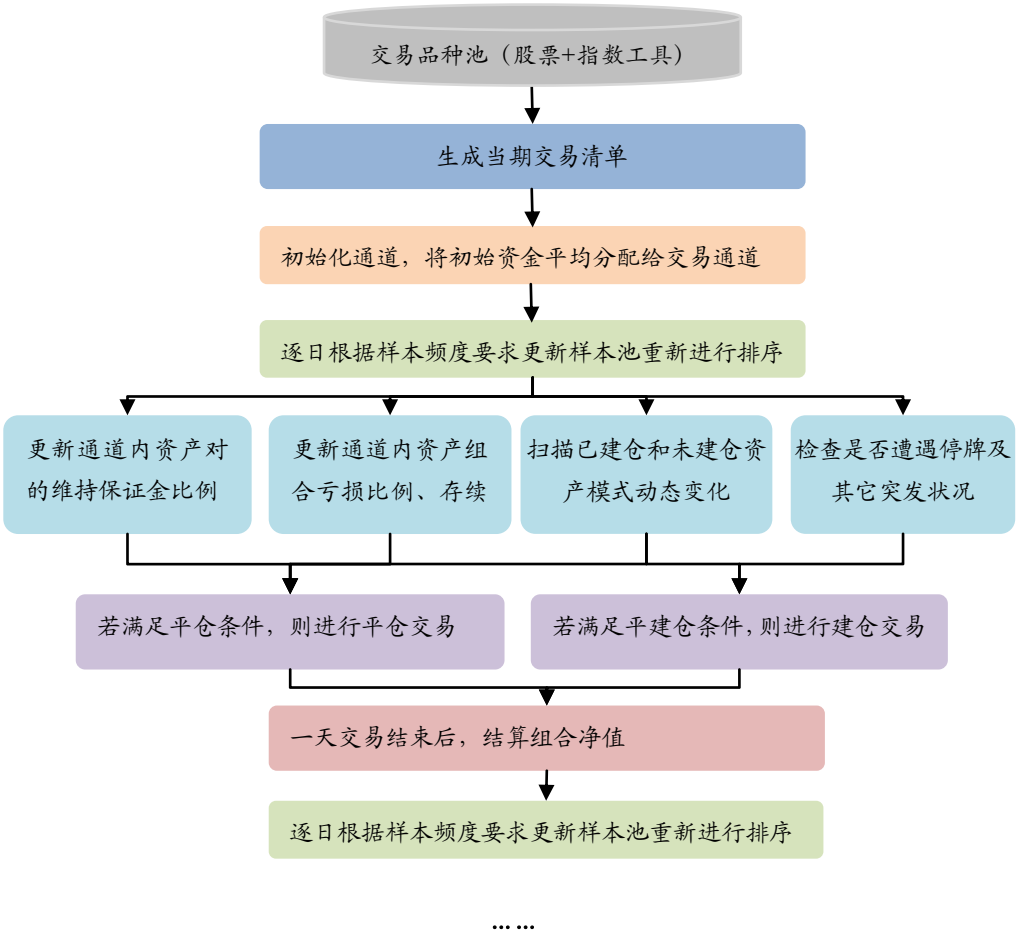
指数投资应用

由于指数工具在交易中有交易成本低、冲击成本低、无停牌、可带有杠杆等等特性，因此，可能比股票更适合用于基于模式识别的短线交易，在后续的实证报告中，我们将检验本方法用于指数工具交易的效果。

基于模型选股结果的交易策略

在报告中我们分别检验了 Top 和 Bottom 模式组合的统计特性，实证检验中，将通过下图所示两种模式进行策略的回溯和跟踪。由于目前融券名单的覆盖面远远不够，因此 Long-Short 模式仍然只是一种理论上的可能，或是我们可以考虑在 ETF 融券实现后利用 ETF 短期融券成本低的特性，进行另一种变形的 Long-Short 策略检验。由于统计中 Bottom 组合的效果显著弱于 Top 组合，且样本数量显著减少，因此较为可行的策略可能是一个纯多头策略（或+融券或期指对冲 beta）的形式。

图 40 ： 基于模型选股结果的交易策略示意图



资料来源：天软科技，国信证券经济研究所整理

频度扩展

分形理论中有一种假设：如果该模式在某种频度下形成某种有效信息，那么在其他频度下也应该有类似效果，但应用频度需要同步放大或缩小。基于这一假设，未来我们将把本报告的分析频度扩展到小时和日 k 线上并进行下一步的实证研究。

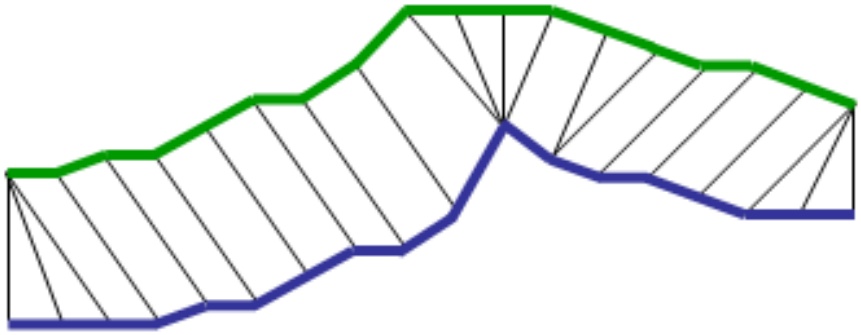
模式伸缩

本报告中所采用的样本均为固定长短（三个交易日的 15 分钟频度 K 线图），进行抽象及聚类，但在现实应用中，模式识别经常遭遇模式放大、缩小、发育不完全或有多余信息的情况，因此我们需要在未来的研究中加入模式的动态（或智能）识别，在这一研究中我们将引入一种新的研究方法：动态时间弯曲算法（dynamic time wrapping）。

最初应用在自动语音识别系统中，研究者利用动态时间弯曲算法处理不同语速给语音识别带来的影响，可以在一定程度上客服时间序列之间的相位差问题，而且可以用来度量等长或者非等长时间序列间的相似性。和传统的欧式距离相比，动态时间弯曲可以获得很高的识别、匹配精度，但是计算的时间复杂度比较大。

如下图，两个时间序列通过动态时间弯曲算法可以得到更为精确的匹配，时间点之间可以满足一对多、多对一的映射关系，最大限度的保留了两个时间序列间的相似性。

图 41：动态时间弯曲算法示意图



资料来源：天软科技，国信证券经济研究所整理

择时应用

由于我们的方法在不同频度下对品种未来的涨跌可做概率分析，因此综合这些信息似乎可以对市场同期的涨跌幅进行某种形式的预测。根据以往的研究经验，择时模型的应用有众多极限，因此我们更倾向于将这一研究成果用于指数工具的 Long-Short 策略。

附录

①Average Precision（AP）指标

在信息检索任务中，给定某个查询词，系统返回符合该查询词的文档。如果系统可以将更为相关的文档在较早的位置返回，那么系统的检索效果越好。这与我们选股模型的思想是一致的，针对 Top 模式组合来说，如果某个模式组合可以尽早将涨幅大的股票样本返回，那么该模式组合的得分越高。相反，针对 Bottom 模式组合来说，如果某个模式组合可以尽早将跌幅大的股票样本返回，那么该模式组合的得分越高。因此，我们借鉴信息检索中的 Average Precision（AP）指标来评价 Top、Bottom 模式组合。针对 Top 模式组合来说，AP 计算的是每个上涨股票样本返回后的涨幅平均值。针对 Bottom 模式组合来说，AP 计算的是每个下跌股票样本返回后的涨幅平均值。如下公式所示:

$$AveP_{Top}=\frac{\sum_{r=1}^{N}P(r)\times sgn(Stock_r>0)}{N}$$

$$sgn(Stock_r>0)=\begin{cases}0 & \text{如果第r个返回的股票样本下跌}\\ 1 & \text{如果第r个返回的股票样本上涨}\end{cases}$$

在公式中，r 表示股票样本的排序，N 表示返回的股票样本个数，在本文中 N=1000。而 sgn 函数是一个符号函数，当第 r 个股票样本涨幅 $Stock_r>0$ 时，其取值为 1，否则取值为 0。函数 P(r)计算的是前 r 个股票样本涨幅 $Stock_r$ 的平均值，如下公式所示。

$$P(r)=\frac{\sum_{i=1}^{r}Stock_i}{r}$$

从公式中可以看出，$AveP_{Top}$ 仅计算在上涨股票样本返回时的涨幅平均值，如果某个模式组合可以将涨幅大的样本尽早返回，那么将有较大的 $AveP_{Top}$ 取值，我们穷举所有可能的成交价格、成交量模式组合，并按照 $AveP_{Top}$ 排序得到前 20 个 Top 模式组合。

类似地，$AveP_{Bottom}$ 定义如下:

$$AveP_{Bottom}=\frac{\sum_{r=1}^{N}P(r)\times sgn(Stock_r<0)}{N}$$

$$sgn(Stock_r<0)=\begin{cases}0 & \text{如果第r个返回的股票样本上涨}\\ 1 & \text{如果第r个返回的股票样本下跌}\end{cases}$$

从公式中同样可以看出，$AveP_{Bottom}$ 仅计算在下跌股票样本返回时的涨幅平均值。我们也依据公式计算得到前 20 个 Bottom 模式组合。

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

风险提示

本报告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。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价或询价。我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。我公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。

证券投资咨询业务的说明

证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。

证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

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
崔　嵘 021-60933159

固定收益
李怀定 021-60933152
侯慧娣 021-60875161
张　旭 010-66026340

策略
黄学军 021-60933142
林丽梅 021-60933157

技术分析
闫　莉 010-88005316

交通运输
郑　武 0755- 82130422
陈建生 0755- 82133766
岳　鑫 0755- 82130432
周　俊 0755-82130833-6215
糜怀清 021-60933167

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

商业贸易
孙菲菲 0755-82130722
祝　彬 021-60933156
常　伟

汽车及零配件
左　涛 021-60933164

钢铁及新材料
郑　东 010- 66025270
秦　波 010-66026317
郭　莹 010-88005303

机械
郑　武 0755- 82130422
陈　玲 0755-82130646
杨　森 0755-82133343
后立尧 010-88005327

基础化工
刘旭明 010-66025272
张栋梁 0755-82130532
罗　洋 0755-82150633
吴琳琳 0755-82130833-1867
梁　丹 0755- 82134323

医药
贺平鸽 0755-82133396
丁　丹 0755- 82139908
杜佐远 0755-82130473
胡博新 0755-82133263
刘　勍 0755-82130833-1845

电力设备与新能源
杨敬梅 021-60933160
张　骁 010-88005311

传媒
陈财茂 010-88005322
刘　明 010-88005319

有色金属
彭　波 0755-82133909
龙　飞

电力与公用事业
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
唐俊杰 021-60875160

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
欧阳仕华 0755-82151833

电子元器件
段迎晟 0755- 82130761
高耀华 0755-82130771
熊　丹 0755-82133528

纺织服装
方军平 021-60933158

农业
张　如 021-60933151

建材
郑　东 010- 66025270
马　彦 010-88005304

旅游
曾　光 0755-82150809

食品饮料
黄　茂 0755-82138922

建筑
邱　波 0755-82133390
刘　萍 0755-82130678

新兴产业
陈　健 010-66022025
李筱筠 010-66026326
孙　伟 010- 66026320

研究支持
沈　瑞 0755-82132998
雷　达 0755-82132098
余　辉 0755-82130741
王越明 0755-82130478

数量化投资产品
焦　健 0755-82133928
阳　璀 0755-82133538
周　琦 0755-82133568
邓　岳 0755- 82150533

基金评价与研究
杨　涛 0755-82133339
康　亢 010-66026337
刘舒宇 0755-82133568
李　腾 0755-82130833-6223
刘　洋 0755-82150566
潘小果 0755-82130843
蔡乐祥 0755-82130833-1368
钱　晶 0755-82130833-1367

量化投资策略
葛新元 0755-82133332
董艺婷 021-60933155
郑　云 021-60875163
毛　甜 021-60933154
李荣兴 021-60933165
郑亚斌

量化交易策略与技术
戴　军 0755-82133129
黄志文 0755-82133928

数据与系统支持
赵斯尘 021-60875174
徐左乾 0755-82133090

彭甘霖	0755-82133259	李扬之	0755-82136165
秦国文	0755-82133528	陈爱华	0755-82133397
韦　敏	0755-82130833-3772	袁　剑	0755-82139918
张璐楠	0755-82130833- 1379		

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	魏　宁	0755-82133492 13823515980 weining@guosen.com.cn
王晓建	010-66026342 13701099132 wangxj@guosen.com.cn	马小丹	021-60875172 13801832154 maxd@guosen.com.cn	邵燕芳	0755-82133148 13480668226 shaoyf@guosen.com.cn
焦　戬	010-66026343 13601094018 jiaojian@guosen.com.cn	郑　毅	021-60875171 13795229060 zhengyi@guosen.com.cn	林　莉	0755-82133197 13824397011 linli2@guosen.com.cn
李文英	010-88005334 13910793700 liwying@guosen.com.cn	黄胜蓝	021-60875166 13761873797 huangsl@guosen.com.cn	王昊文	0755-82130818 18925287888 wanghaow@guosen.com.cn
赵海英	010-66025249 13810917275 zhaohy@guosen.com.cn	刘　塑	021-60875177 13817906789 liusu@guosen.com.cn	甘　墨	0755-82133456 15013851021 ganmo@ guosen.com.cn
原　祎	010-88005332 15910551936 yuanyi@guosen.com.cn	叶琳菲	021-60875178 13817758288 yelf@guosen.com.cn	段莉娟	0755-82130509 18675575010 duanlj@guosen.com.cn
		孔华强	021-60875170 13681669123 konghq@guosen.com.cn	徐　冉	0755-82130655 13632580795 xuran1@guosen.com.cn
				颜小燕	0755-82133147 13590436977 yanxy@guosen.com.cn
				赵晓曦	0755-82134356 15999667170 zhaoxxi@guosen.com.cn
				郑　灿	0755-82133043 13421837630 zhengcan@guosen.com.cn