

分类号_____

密级_____公开_____

UDC _____

编号_____20140210105_____

广东外语外贸大学

硕士学位论文

事件驱动型策略下的股票价格预测

——基于支持向量机的方法

申请人姓名 _____ 杨念念 _____

导师姓名及职称 _____ 杨励，教授 _____

申请学位类别 _____ 国际商务硕士 _____

学科专业名称 _____ 国际商务硕士 _____

培养单位 _____ 经济贸易学院 _____

学位授予单位 _____ 广东外语外贸大学 _____

2016 年 6 月 6 日

分类号_____

密级_____公开_____

UDC _____

编号_____20140210105_____

广东外语外贸大学硕士学位论文

事件驱动型策略下的股票价格预测

——基于支持向量机

Event Driven System Stock Price Forecast Model

—— Based on SVM

申请人姓名_____杨念念_____

导师姓名及职称_____杨励，教授_____

申请学位类别_____国际商务硕士_____

学科专业名称_____国际商务硕士_____

论文提交日期_____2016年6月6日_____

论文答辩日期_____2016年5月24日_____

答辩委员会_____黄新飞 教授_____

_____张金艳 教授 肖奎喜教授_____

学位授予单位：_____广东外语外贸大学_____

独 创 性 声 明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 广东外语外贸大学 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：

签字日期：

年 月 日

学位论文版权使用授权书

本学位论文作者完全了解 广东外语外贸大学 有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权 广东外语外贸大学 可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本授权书）

学位论文作者签名：

导师签名：

签字日期： 年 月 日

签字日期： 年 月 日

摘 要

事件驱动投资策略已经发展成与行业配置并驾齐驱的策略研究和资产配置方法。在提前挖掘和深入分析可能造成股价异常波动的事件基础上,把握交易时机可以让投资者在事件驱动型股票上获取超额投资回报。

SVM 因为能出色地处理非线性关系,近年来被越来越多地被使用在股价预测上。本文用资金流向作为体现事件驱动型股票的观测指标,应用支持向量机分类方法,对个股价格的涨跌进行预测。经过实证分析,该模型的预测准确率约为80%以上。

本文第一章是绪论,主要阐述了本文的写作背景和创新点,以及国内外关于相关课题的研究。

第二章是本文写作的前提,主要介绍了股票市场上的两个常用指标,价格和股票价格指数,并简单总结了目前股票市场常用的预测分析方法。

第三章是有关事件驱动策略的理论分析,解释了事件影响股价的原因,并举出了两只事件驱动策略的基金案例。

第四章包括了支持向量机的数学原理以及它是怎样预测股价的。

第五章本文的关键,建立了预测股价的模型,并通过实证分析证明了模型的准确性。

第六章提出了本文研究的长处和不足。

关键字: 支持向量机; 资金流向; 涨跌预测

Abstract

Event-driven strategy has already developed into and keeps abreast of industry configuration and asset allocation strategy research method. In the early mining and analyzing the possible causes of the abnormal stock price fluctuations event, based on the grasp the trading time can let investors get excess returns on the event-driven stock.

SVM has been applied to predict the stock price more and more frequently in recent years due to its great capability to deal with the non-linear relation. This thesis predicts the price of certain stocks by using the support vector machine and the capital flow as the index of event driven stock. The predictions of this model is above 80% accurate based on empirical analysis.

In this paper, the first chapter is introduction, mainly elaborated the writing background of this article and the innovation points, as well as the domestic and foreign research on related topics.

The second chapter is the precondition of the paper, mainly introducing the two of the usual indicator in security market, price and price index. This chapter also summarizes several common methods to predict the price of stock.

The third chapter is about the event-driven strategy theory analysis, explained the reason why event affect the stock price, and the two event-driven strategy fund case.

The fourth chapter includes the mathematical principle of support vector machine (SVM) and how it is to predict stock price.

The fifth chapter is the key of the article, to predict stock price model is established, and through the empirical analysis shows that the accuracy of the model.

The sixth chapter put forward the advantages and disadvantages in this paper.

Key Words: SVM; capital flow direction; advance-decline prediction

目 录

摘 要.....	I
Abstract.....	II
目 录.....	III
第 1 章 绪论.....	1
1.1 研究背景及研究意义.....	1
1.2 文献综述.....	1
1.2.1 股票市场的可预测性和非线性.....	1
1.2.2 基于支持向量机的股价预测的研究.....	2
1.2.3 事件驱动型策略的相关研究.....	3
1.3 创新之处.....	4
第 2 章 股票市场预测理论.....	5
2.1 股票市场的基本指标.....	5
2.1.1 股票价格.....	5
2.1.2 股票价格指数.....	5
2.2 股票市场的分析方法.....	6
第 3 章 事件驱动策略的理论分析.....	7
3.1 事件驱动型策略的含义.....	7
3.2 事件对股价的影响路径.....	9
3.2.1 金融市场的羊群效应.....	9
3.2.2 预期对资金流向产生影响.....	10
3.2.3 资金流向对价格产生影响.....	11
3.3 事件驱动策略的相关案例.....	13
第 4 章 支持向量机的原理.....	16
4.1 机器学习原理.....	16
4.2 支持向量机的分类原理.....	16
4.2.1 最优分类面.....	16
4.2.2 支持向量分类机.....	18
第 5 章 实证研究.....	20
5.1 建立 SVM 模型.....	20
5.1.1 概述.....	20
5.1.2 模型的数学表达.....	21
5.2 实证工具.....	22

5.3 实证数据.....	22
5.3.1 指标的选取.....	22
5.3.2 样本的选取.....	24
5.4 实证结果.....	25
5.4.1 预测结果.....	25
5.4.2 检验结果.....	30
5.4.3 简化结果.....	30
第6章 总结和展望.....	32
参考文献.....	33
致谢.....	35

第 1 章 绪论

1.1 研究背景及研究意义

自上世纪 90 年代第一次提出了事件驱动投资策略这个概念以来，该策略已经成长为与行业配置相提并论的策略之一。将资金合理地配置到不同行业或者不同主题中，可以有效对冲股票的系统性风险。事件驱动投资策略属于主题投资的另一种形式，是根据受到不同事件影响进行分类并且组合投资的策略。而在全球化的经济形势下，单纯用行业或者区域的定义已经不能准确划分成千上万的股票，这时主题投资应运而生。它并不是按照一般的逻辑使用基本面分析或者技术面分析进行选股，而是从宏观到行业，再到公司个体地挖掘可能会对行业或公司产生影响的各种有发生可能性的事件，通过这一系列的事件形成投资主体，然后在该主题中寻找可以受益的，并且基本面技术面都良好的公司进行投资，最后构建一揽子的事件型主题投资股票组合。

“事件驱动型交易策略”是在经过对整体宏观经济的分析以及对各行各业的把握之后，投资者认为事件有一定概率发生，对股价异常波动产生预期的基础上，通过择时等一系列有效的操作获取超额投资回报的交易策略。“事件驱动型交易策略”的“事件”一般不会有明确的定义，但是具有较为明确的事件窗口和内容。通常会影响投资者对股票的判断，产生或者看涨或者看跌的预期，进而对部分投资者的投资行为产生一定的影响，从而造成股价短期的异常波动。例如重大资产重组事件、宏观政策重大利好、突破性的技术创新事件。研究“事件驱动型交易策略”成为我们在股票市场获得超额利润的重要途径。

1.2 文献综述

1.2.1 股票市场的可预测性和非线性

早在上世纪 60 年代，经济学家们就已经开始了关于股票市场是否有效及是否可预测的研究。1965 年，Paul Samuelson 提出了价格随机游走理论。该理论认为，在信息完全流通的市场中，信息的传播不需要任何成本，价格的变化是无法预测的。随之，Fama（1970）在其“有效市场理论”中更清晰的阐述了有效市

市场假设 (EMH) 理论, 对股票市场不可预测性提供了权威性的阐述。假定市场上的投资者都是理性的追求利益最大化, 并且投资者可以轻而易举获得市场上的各种信息, 或者说信息的传播并不需要任何成本。这里“有效”一词的意思并非经济学中通常所指的减去成本的收益最大化, 而是指公开的信息会被市场快速消化并反映在股价上。之所以会发生这样的情况, 是因为市场竞争非常激烈, 多空双方会根据市场状况迅速做出调整, 追求更大化的利益。再加上科学的发展, 使得更智能化、程序化的机器应用到市场中, 因此股价对消息的反应不是在几秒内做出的, 而是在千分之一秒内做出的。

之后有效市场理论获得了广大的认可, 在学术以及实际中都得到了广泛的应用。但随着探讨的深入, 越来越多的实证研究却表明, 随机游走模型并不契合股票市场的实际现象, 股票市场具有一定的可预测性。1963 年, Mandelbrot 首次提出了股票价格行为并不遵循随机游走模式, 而是存在“诺亚效应 (Noah Effect)”和“约瑟效应 (Joseph Effect)”。其中, “诺亚效应”证明了股票的收益分布存在“尖峰厚尾现象”, 服从稳定的帕累托分布, 是非有效市场假说所说的正态分布。“约瑟效应”说明了股票收益存在长期相关性, 是非有效市场假说所说的“股票收益是独立的”。最具影响力的文献是 Fama and French (1988a, 1989) 和 Campbell and Shiller (1988b)。他们讨论了证券组合的股息收益 D_t/P_t-1 , 尤其是股息价格比 D_t/P_t 能够成功预测期望收益, 这种预测方法偏向基本面的分析, 对市场情绪的讨论是不很到位。

1.2.2 基于支持向量机的股价预测的研究

虽然支持向量机 (又称 SVM) 在理论上具有非常突出的优势, 但其应用到实际生活中还处于起步阶段, 近几年才在经济、金融等方面的研究开始变得广泛。Trafalis.T 和 Ince.H (2000) 用支持向量机和 MLP 和 ARIMA 进行对比研究, 证实了 SVR 的预测精度远高于 MLP 和 ARIMA, 这时 SVM 的优势得以体现。L.J.CAO 和 Tay (2001) 主要在金融预测领域做了较多的研究, 他们对支持向量机和 BP 神经网络进行比较研究, 得出 SVM 比神经网络具有更出色的性能, 而且参数的选取对预测精度的影响很大。Kim (2003) 用支持向量机进行股指趋势预测, 并与 CBR (case-based reasoning) 进行了对比, 结果显示 SVM 具有明显的性能优势和预测精度。Huang.W (2005) 等人用 SVM 对日经 225 指数每周走势进行预测, 并与线性判别分析 (LDA)、二次判别分析 (QDA)、BP 神经网络进行了对比, 结果表明 SVM 预测率远高于其它几种方法。Chen. S (2007) 等人将基于 GARCH 模型的 SVR 与单个 GARCH 模型、MA 模型、神经网络、

SVR, 对纽约证券交易所综合指数、英镑对美元汇率波动率预测进行了对比, 结果表明新的模型具有更好的预测性能。自此, 国外的研究已经充分证实了支持向量机在预测效果上高于常用的其他几种方法。

国内关于 SVM 的研究基本上从 2005 年之后才开始进行。在研究的早期, 学者们经常讲支持向量机和人工神经网络进行对比。他们对比后认为, 神经网络经常具有学习速度慢、网络性能差、局域极小值等缺点, 而支持向量机正好能克服这些缺点。杨一文和杨朝军(2005)提出了神经网络的局限性, 引出支持向量机对上证指数较为准确的多步骤预测。汤凌冰和盛焕焯(2009)将小波分析与支持向量机结合构造了一个小波支持向量机, 实证研究后证明该方法的预测精度高于单纯使用小波分析和支持向量机。随着对资本市场认识的加深, 金融时间序列的观点也被学者带入到股价预测中。吴江和李太勇(2010)提出了金融时间序列的加权 SVM 预测方法。研究表明, 加权 SVM 可以有效地提高预测的准确度。姚萧和余乐安(2012)将模糊隶属度的思想引入 SVM 中, 提出了一种模糊近似 SVM, 既保持了原有 SVM 的优点, 而且可以降低离群点和噪声对模型的干扰, 该 SVM 能显著提高信贷风险分类的准确性, 具有较高的实用价值。赛英和张凤廷(2013)用粒子群优化的线性核函数 SVM 对沪深 300 股指期货进行了回归预测, 并取得良好效果。

1.2.3 事件驱动型策略的相关研究

主题投资策略最早是由 Edward M. Kerschner 提出, 20 世纪 90 年代开始盛行于美国、英国等国。肖玉航(2013)认为 A 股“主题投资”具有必然性, 但是投资者并不是都能保持清醒的投资理念, 因此国内外的学者普遍用行为金融学来研究主题投资策略, 也就是研究该策略中的“羊群效应”。比较有代表性的研究是 Lakonishok et al(1990)对于美国证券市场养老基金的羊群效应的研究。该文提出了用买卖双方交易量的不均衡来测度羊群行为, 这种方法被后面的许多学者引用, 并得到了一些改进。例如, Wermers(1999)对他们的羊群行为的定义进行了修正。宋军、吴冲锋(2001)验证了中国市场也具有羊群效应, 并推断: 在市场大幅下挫的时候给投资者带来的抛售压力明显大于在市场上升时的给投资者带来的购买冲动。李雪等(2008)选取的是申银万国证券研究所在 2006 年和 2007 年间提出的四个主要的投资主题, 证明了中国 A 股市场主题投资策略中存在较为明显的羊群行为, 很多券商机构或者网站都会根据行业动态发布各种投资主题, 通过互联网等媒介的传播迅速成为投资者追捧的热点, 一旦客观环境中出现了符合投资主题的因素, 则 Shiller 所定义的兴趣传染的条件将会被满足, 羊

群行为将随着投资主题的散播而形成和延续。

事件驱动投资策略是主题投资（Thematic Investing Strategy）的一种，国内外现在对事件驱动投资策略单独的学术研究较少。李备友等（2012）分析了基于事件驱动的证券市场波动生成机制，结论认为在大盘走势与事件驱动方向相同的情况下，大盘指数将放大事件驱动的影响，加速股票价格波动率的上升；在大盘走势与事件驱动方向相反的情况下，大盘指数将打压事件驱动的影响。杨阳等（2014）借鉴事件分析法以东方证券研究部发布的主题策略库中，2005年初至2011年6月的17个事件驱动投资主题作为研究样本，实证检验了事件驱动投资策略的超额收益率。

1.3 创新之处

SVM因为在解决小样本、非线性表现出许多特有的优势，常被应用在股票价格预测中。但是国内外研究中，绝大部分都只是用价格指数本身去预测未来的价格。例如，张玉川、张玉泉（2007）在文中使用的指标是MA（移动平均线）、MACD（指数平滑异同平均线）、KDJ（随机指标）、RSI（相对强弱指标）、BIAS（乖离率指标）、OBV（能量潮指标），以上的六个指标仅涉及到了股票的开盘价、收盘价、成交量等基本观测数据。本文认为，仅用此类基本观测数据的预测模型依旧存在很大的运气成分，并不能从股价波动形成机制的理论上说服读者。因此本文的创新点就在于：研究事件影响股票价格的路径，并且使用SVM来预测事件驱动型股票的价格。

第 2 章 股票市场预测理论

2.1 股票市场的基本指标

2.1.1 股票价格

股票是有价证券的一种，有价证券是指一类可以证明所有权或债券的凭证，没有实际价值，但是有市场价格。股票还有票面价值，例如 A 股股票面值一般为固定的 1 元，港股面值大部分为 0.1 元或者 0.01 元，但票面价值在实际交易中没有任何意义。股票价格则主要指股票在二级市场交易的价格。

虽然股票也算一种特殊的商品，但影响股票价格的因素和影响普通商品价格的因素大有不同，这决定了股票具有它特殊的地方。首先，股票的价格可能不仅仅是根据简单的供需来决定，政治环境、宏观经济、心理因素等等都有可能影响到股票价格，因此股票价格具有复杂性。其次，商品在一定时间内可能因为供需达到平衡而使得价格也趋于一个稳定，然而因为频繁的交易频率，使得股票价格很难维持稳定，而是具有波动性。最后，规律性，这是基于投资者行为的一种特殊性，投资者可能会根据以前的经验做出相同的判断，使得股票价格具有规律性。

一般来说，股票的价格会围绕其内在价值波动，因此股价首先反映的是投资者对公司的估值。市盈率 PE 是体现估值高低最直接的指标， $PE = \text{股价} / \text{每股盈利}$ 。投资者会赋予不同行业不同的市盈率，成长性低的制造业可能平均市盈率只有个位数，成长性高的高科技行业平均市盈率也有可能高达三位数。因此，在估值大幅度偏离股票内在价值时，市场会自发地进行估值修复，反映在股价就是股票交易价格向股票内在价值的趋同。

2.1.2 股票价格指数

股票价格指数描述的是股票市场总体价格的变化，或者是一揽子股票的总体价格变化，反映了一个时间段股市或者行业的综合性动态。股票价格指数的编制工作由证监会完成，比如在中国最重要的上证指数和深证成指。上证指数属于综合指数，包含了在上海证券交易所上市的所有股票；深成指数属于成分指数，是选取了在深圳证券交易所上市的股票中具有一定代表性的股票。

另外证监会还会将上市的几千只股票进行行业的分类，每一个行业也会编制

其行业指数。在各种概念出台时，例如智能穿戴、国企改革，即使证监会对这类股票没有明确的分类，但是各种炒股软件或者券商也会将概念的一揽子股票价格编织成指数。

股票价格指数可以很全面、很精简地概括出市场的总体情况，是投资者分析市场的重要参考指标。

2.2 股票市场的分析方法

即使研究表明了股票价格确实具有可预测性，但是想要非常精确地预测股票价格是很难的，因为影响股票价格的因素太多、太复杂。但是仍然有一些方法可以让我们对股价未来的趋势做出一定成的判断。在上世纪 60 年代以前，人们可能更多使用较为基础的分析方法来判断股价走势，常用的有基本面分析法和技術面分析法。在科学快速发展的当代，一部分人更倾向于用统计，甚至于人工智能的方法来分析股价。对股市的分析方法介绍如下：

（1）基本分析法

股票的基本分析是以股票的内在价值为依据，忽略掉市场上的短期情绪，由市场的供需情况分析股价趋势。

（2）技术分析法

道氏理论是技术分析的鼻祖，该理论认为：市场消化和反应一切信息；市场行为是按趋势推行的；历史会重演。随后威廉·彼得·汉密尔顿所著的《股市晴雨表》继承并且丰富了道氏理论。总结而言，技术分析就是通过历史的经验和走势来预测未来的趋势。技术分析发展到今天已经分离出了数量繁多的流派，但是技术分析在预测精度上仍然存在很大的缺陷，并且缺乏可靠的理论支持。

（3）传统统计学

该方法主要是基于建立各种回归模型和最小二乘原理进行预测。

（4）时间序列分析法

该方法的研究对象的数据都是随时间变化的，利用数学工作处理历史数据并建立模型，找出数据的变化规律和趋势，进而预测未来时间的状态。

（5）人工智能法

自从进入电子信息技术时代，金融股票市场中也出现了许多人工智能的新方法，并且发展十分迅速。其中最常用的是人工神经网络。

第3章 事件驱动策略的理论分析

3.1 事件驱动型策略的含义

在中国股市，投资者可能更关注股票短期价格的异常波动，也就是价格可能大幅地偏离其应有的价值。但股票价格的异常变动并不一定是异常的风险，有可能也伴随着巨大的超额收益。而股价的异常波动，往往是受到某些事件的驱动，这些事件有可能大到是宏观经济的影响，小到是业绩的公布。在这样的客观基础上，一种新的交易策略产生了，那就是事件驱动型交易策略（Event-Driven Strategy）。

自上世纪 90 年代，来自瑞银华宝的分析师爱德华科斯纳第一次提出了事件驱动投资策略这个概念，如今该策略已经成长为与行业配置相提并论的策略之一。将资金合理地配置到不同行业或者不同主题中，可以有效对冲股票的系统性风险。事件驱动投资策略属于主题投资的另一种形式，是根据受到不同事件影响进行分类并且组合投资的策略。而在全球化的经济形势下，单纯用行业或者区域的定义已经不能准确划分成千上万的股票，这时主题投资应运而生。它并不是按照一般的逻辑使用基本面分析或者技术面分析进行选股，而是从宏观到行业，再到公司个体地挖掘可能会对行业或公司产生影响的各种有发生可能性的事件，通过这一系列的事件形成投资主体，然后在该主题中寻找可以受益的，并且基本面技术面都良好的公司进行投资，最后构建一揽子的事件型主题投资股票组合。

“事件驱动型交易策略”是在经过对整体宏观经济的分析以及对各行各业的把握之后，投资者认为事件有一定概率发生，对股价异常波动产生预期的基础上，通过择时等一系列有效的操作获取超额收益。对事件一般不会有明确的定义，但是具有较为明确的事件窗口和内容。通常会影响投资者对股票的判断，产生或者看涨或者看跌的预期，进而对部分投资者的投资行为产生一定的影响，从而造成股价短期的异常波动。例如重大资产重组事件、宏观政策重大利好、突破性的技术创新事件。研究“事件驱动型交易策略”成为我们在股票市场获得超额利润的重要途径。

在有的文献中，它们定义事件驱动型交易策略的“超额投资回报”为一揽子股票的组合由于某类事件（可以是相同的事件也可以是不相同的事件）的发生导致股价出现异常波动，偏离了其原来的行走轨迹，其投资组合的实际涨幅超过同期大盘指数涨幅的部分称之为“超额收益”。比如，在一段时间内，根据事件驱

动型策略投资的一揽子股票组合的总收益率为 20%，同期沪深 300 指数上涨 10%。由此可以得出，该组合由于“重大事件”带来的超额收益是 20%。但在金融事件研究法中，一般认为拿来作比较的基准收益是该股的预测股价。常使用的预测方法有但不局限于 CAPM 模型、套利定价模型、三因素模型等等。得到预测收益以后，定义异常收益为：

$$AR_t = R_t - R_p$$

即为，超额收益=实际收益-预测正常收益。

事件驱动交易策略是欧美资本市场较多使用的主流投资策略之一，投资标的一般为发生某些重大事件的公司。在国外对冲基金行业，按照关注事件的不同，事件驱动策略可以分为三类：并购套利、困境投资、重组投资。并购套利即投资于收到收购要约的目标公司，从市场价格和收购要约价格的差价中获利。近几年，在美国上市的中概股正经历着私有化浪潮，中星微电子、奇虎 360 等优秀的公司都开始发起私有化要约，准备回到国内市场上市。一般私有化要约收购的价格会比现价高出 10%到 20%不等，具有极为稳定的套利空间。困境投资就是以低价买入陷入经营困境的公司，在行业反转或者公司取得重大突破时获得巨大收益。比如在熊市中，投资破净、破增发价的公司都属于困境投资的一种。重组投资，即投资于宣布资产负债表将发生重大变动的公司证券，比如在 2015 年炒作火爆的国企改革概念，就是凭借行业整合、资产重组等形式，改变公司的经营状况，从而使投资者有了看多预期。

在大智慧软件中，事件被分为了自然事件、交易事件、财务事件。自然事件就包括干旱、地震等人类无法阻止的自然现象对上市公司的影响。交易事件包括新闻舆论、资产重组等对上市公司有直接的重大影响的重大事件。财务事件指增发、发行票据等一系列对上市公司的财务状况有影响的事件。

以上两种对事件的分类都不可能很详尽的包括所有的可能，分类的标准也不是很明确，因此在“事件”的定义和分类上，不必追求教科书般的概念，符合大致逻辑即可。

事实上，影响股价走势的因素包括各种宏观、微观和潜在的因素。因此，本文将凡是影响投资者对上市公司的判断从而影响股票价格的各种市场内外因素称为事件，包括宏观政策、行业形势、公司状况等上到宏观下到微观的所有事件。这些事件或多或少地影响股票价格的变化。

3.2 事件对股价的影响路径

事件能够对金融市场有所影响，主要是因为市场并非完全有效的，事件的发生以及借助各种媒介的传播，会迅速传达给投资者，并且对投资者的判断产生影响。投资者对事件的预期会使其采取相应的行为来消化市场上流通的信息，于是就会产生金融资产现金流的变化，从而在根本上影响金融资产的价格。事件对股价的影响实际来源于信息的内在价值，在非完全有效的市场上，信息的传播并不像有效市场上那样容易，信息的内在价值有一个被发现的过程。并且，由于在传播过程中，投资者的非理性，或者说是盲从心理，或者说是趋同心理，会伴随着信心的变化和预期的调整，最终会反映在资金流向和价格波动上。

3.2.1 金融市场的羊群效应

事件驱动投资策略的理论基础在于金融市场的非完全有效性。

假定市场上的投资者都是理性的追求利益最大化，并且投资者可以轻而易举获得市场上的各种信息，或者说信息的传播并不需要任何成本，每一个投资者都尝试用自己获得的信息来预测股票价格。在 1970 年，法玛提出了有效市场假说（efficient markets hypothesis），对有效市场的定义如下：当证券价格能够充分地反映投资者可以获得的信息时，证券市场就是有效市场，换句话说在有效市场中，无论随机选择何种证券，投资者都只能获得与投资风险相当的正常收益率。公开的信息会被市场快速消化并反映在股价上。之所以会发生这样的情况，是因为市场竞争非常激烈，多空双方会根据市场状况迅速做出调整，追求更大化的利益。再加上科学的发展，使得更智能化、程序化的机器应用到市场中，因此股价对消息的反应不是在几秒内做出的，而是在千分之一秒内做出的。

但是大量的心理学和行为科学证据显示，投资者行为并不像有效市场假说假设的那样理性，而是具有情绪化的有限理性。并且，投资者的非理性行为并不一定相互抵消，而是经常以同样的方式偏离。例如在 2016 年初，中国股市第一次实行“熔断机制”时，其实正是“熔断机制”强化和叠加了大家看空的情绪，因此造成股市大跌。事件驱动投资策略就是建立在这样一个由非理性投资者组成的非完全有效性市场。在信息环境不确定的情况下，投资者的决策行为往往不完全是根据已有的信息，而是容易收到其他投资者的影响并模仿他人的决策进行投资判断，就形成了羊群效应（herd behavior）。

因此在事件发生之前，部分投资者就会对事件未来对股价的影响做出预期，市场的羊群效应使得更多的投资者抛弃原先自身对股价的判断，也对事件的影响

做出了相同的预期。

3.2.2 预期对资金流向产生影响

在中国的 A 股市场上，由于内幕交易者的普遍存在，媒体释放的消息不一定能推动股票涨跌，因为在事件出现之前，也许股票已经涨过或者跌过了。因此我们能观测到的有效事件驱动指标就是资金流向。

资金流向测算的是推动指数涨跌的力量强弱，这反映了是更多的资金看多还是更多的资金看空。指数处于上升状态时产生的成交额是推动指数上涨的力量，指数下跌时的成交额是推动指数下跌的力量。特别是机构投资者操控的大额资金，会对股价造成更大的影响。例如，公司发布了资产重组的公告后，市场上的资金为了取得良好的投资效益就会争先流入到该股中，造成个股价格迅速攀升，长期投资效益也体现在了短期价格暴涨上。另外还有资金流出的例子。正如重庆啤酒（600132）在 2011 年 11 月 25 日开始的一轮暴跌，起因是重庆啤酒研发乙肝疫苗失败，研发投入颗粒无收。投资者认为这起事件将严重影响到该公司的业绩，于是大量资金外逃，造成了重庆啤酒股票近 300% 的亏损。

以中国中车（600176）为例。中国南车于 2014 年 10 月 28 日发布停牌公告，直到当年 12 月 31 日复牌并发布了与中国北车合并的预案。但是市场对南北车合并的反映并不是从发布公告时才开始，实际上我们从中国南车 28 日停牌前的 20 个交易日的成交次数以及资金净流入额便可以看出，未发生的事件已经驱动了投资者的交易。本文观测了中国南车 2014 年 9 月 22 日至 2014 年 10 月 24 日的 20 个交易日的成交次数。前七个交易日，成交次数基本在 2500 次/日上下震荡，之后成交变得频繁，在停牌前的最后一个交易日 10 月 24 日，成交次数达到 4674 次。从资金净流入额来看，临近停牌日的几个交易日，资金流向发生异动，特大单、大胆、中单、小单的净流入额均出现峰值。



图 3-1 中国南车 2014 年 9 月 22 日至 2014 年 10 月 24 日 成交次数

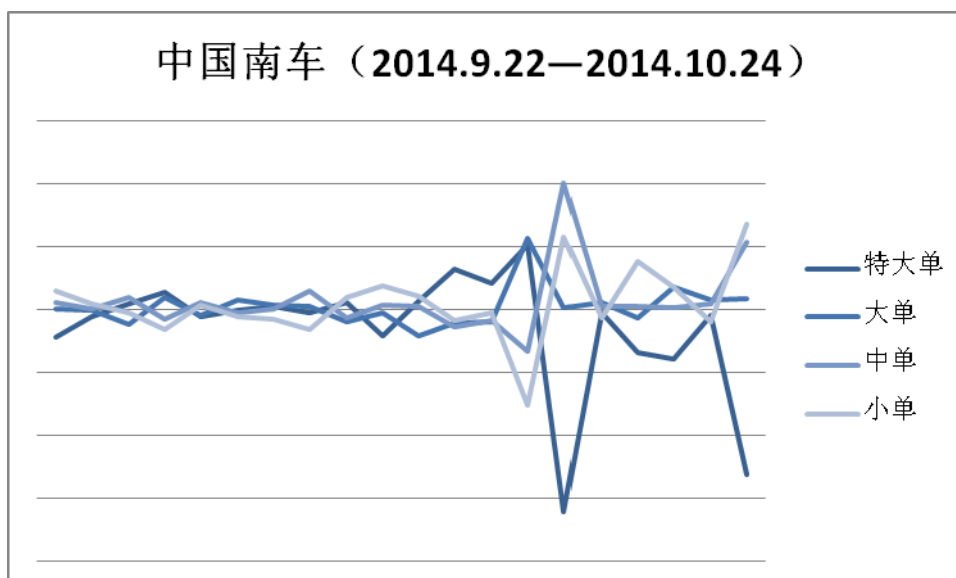


图 3-2 中国南车 2014 年 9 月 22 日至 2014 年 10 月 24 日 资金净流入

3.2.3 资金流向对价格产生影响

商品价格反映了商品的内在价值，在出现了货币这个概念之后，人们会用货币来衡量商品的价值，从而衍生出商品价格这个概念。从根本上来说，商品的价格是内在价值的体现，也是市场上供需关系的体现。股票作为一种特殊的商品，同样也表现出这样的属性。股价就是股票价值货币化的表现，反映了股票的内在价值，也体现了股票的供需关系，并随着股票的价值而波动。

虽然股票也算一种特殊的商品，但影响股票价格的因素和影响普通商品价格

的因素大有不同，这决定了股票具有它特殊的地方。首先，股票的价格可能不仅仅是根据简单的供需来决定，政治环境、宏观经济、心理因素等等都有可能影响到股票价格，因此股票价格具有复杂性。其次，商品在一定时间内可能因为供需达到平衡而使得价格也趋于一个稳定，然而因为频繁的交易频率，使得股票价格很难维持稳定，而是具有波动性。最后，规律性，这是基于投资者行为的一种特殊性，投资者可能会根据以前的经验做出相同的判断，使得股票价格具有规律性。

而且，股价的变动往往并不是只受到供需关系的影响。股票的投资者因为想要追逐利益最大化，时常会根据市场信息做出过分的反应，这使得股票价格脱离了内在价值的约束，在短期内有着巨幅的变动。这种巨幅的变动就是来自于投资者的投机行为。

正是因为普通商品买卖与股票投资存在本质差异，投资者可以通过购买份额较多的股票来获取更多的收益。已经持有一定份额股票的投资者，因为有上涨预期，因此会坚定持有手中的份额。这时候，短期内股票的供给会减少，需求会增多。在合理价位买不到份额的投资者会在更高的价位寻找机会，股票持有者也会想要在更高的价位实现收益。理论上，这种现象会使得股价上涨。因此，供求关系的变化决定了股票的短期价格。

股数份额乘上股票价格就是是成交金额，所以在伴随着供求变化的交易过程中，成交资金在数量也产生了变化，因此资金进出成为供求关系变化的直接体现。在流通股不变的情况下，需求来自与股票的看多者，供给方来自于场上的流通股以及股票的看空者。如果需求大于供给时，反映在资金流向上就是资金的净流入；如果供给大于需要，那么反映在资金流向上就是资金净流出。由于市场上众多投资者对于股价的不同预期，流通股就会在投资者之间进行份额的转换，换手率这一指标就是这一现象的体现。



图 3-3 事件驱动对股价的影响路径

3.3 事件驱动策略的相关案例

本文将列举两个事件驱动策略型基金的例子来证明超额收益的存在，分别是华宝事件驱动混合基金（001118）和嘉实事件驱动股票基金（001416）。

华宝事件驱动混合基金（以下简称“华宝”）成立于2015年4月8日，截止2015年第三季度末总份额为51.5亿。从华宝的重仓股中我们可以看到，该基金主要将资金配置在通信、电子、传媒、休闲服务、医药生物等行业，这几个行业都是2015年国家重大政策性行业。相同的，嘉实事件驱动股票基金（以下简称“嘉实”）也同样将资金主要配置在了这几个行业，两只基金都试图积极发掘可能对上市公司当前或未来价值产生重大影响的各类事件，把握各类事件创造出的投资机会，在控制风险的前提下为投资人谋求长期、稳定的资本增值。嘉实成立于2015年6月9日，截止2015年第三季度末总份额为147.2亿份。

本文将以上两只基金自成立以来以来的累计净值与沪深300指数做了一个比较。虽然一直处于下行状态，但嘉实的收益明显优于沪深300，最大超额收益接近20%。华宝的净值走势大致与沪深300相同，收益线多时处于粘合状态，但华宝的走势偶尔会弱于大盘。这与华宝属于混合类基金有关，60%以上的资金用于购买债券或者以现金的形式存在，不足40%的资金用于投资股票，因此华宝的超额收益不明显，但单从风险来说，也比沪深300指数的风险小。

从华宝和嘉实两只基金的案例我们确定了事件驱动策略是受到了实践证实的，超额收益确实存在，并且在合理控制仓位和资金灵活配置的情况下，组合投资可以有效分散风险。

表 3-1 华宝和嘉实重仓前二十的股票

华宝			嘉实		
品种简称	占比（%）	所属行业	品种简称	占比（%）	所属行业
电广传媒	2.50	传媒	利亚德	0.81	电子
一汽轿车	2.39	汽车	天壕环境	0.66	公用事业
峨眉山 A	2.13	休闲服务	东诚药业	0.63	医药生物
中兴通讯	2.05	通信	长春高新	0.62	医药生物
中国联通	1.73	通信	吴通控股	0.58	通信
丽江旅游	1.62	休闲服务	中国中冶	0.56	建筑装饰
广汽集团	1.46	汽车	网宿科技	0.54	通信
乐普医疗	1.43	医药生物	武钢股份	0.52	钢铁
华策影视	1.38	传媒	信维通信	0.48	电子

欧菲光	1.38	电子	中国巨石	0.48	化工
湖北广电	1.34	传媒	红宇新材	0.43	机械设备
尚荣医疗	1.28	医药生物	涪陵电力	0.43	公用事业
中国医药	1.25	医药生物	吉艾科技	0.39	机械设备
燕京啤酒	1.24	食品饮料	新钢股份	0.34	钢铁
铁龙物流	1.22	交通运输	猛狮科技	0.33	汽车
建发股份	1.15	交通运输	烟台冰轮	0.2	机械设备
中国国旅	1.15	休闲服务	同大股份	0.19	化工
东软集团	1.13	计算机	中成股份	0.17	商业贸易
阳普医疗	1.12	医药生物	浩宁达	0.16	电气设备
锦江股份	1.10	休闲服务	赛为智能	0.15	计算机

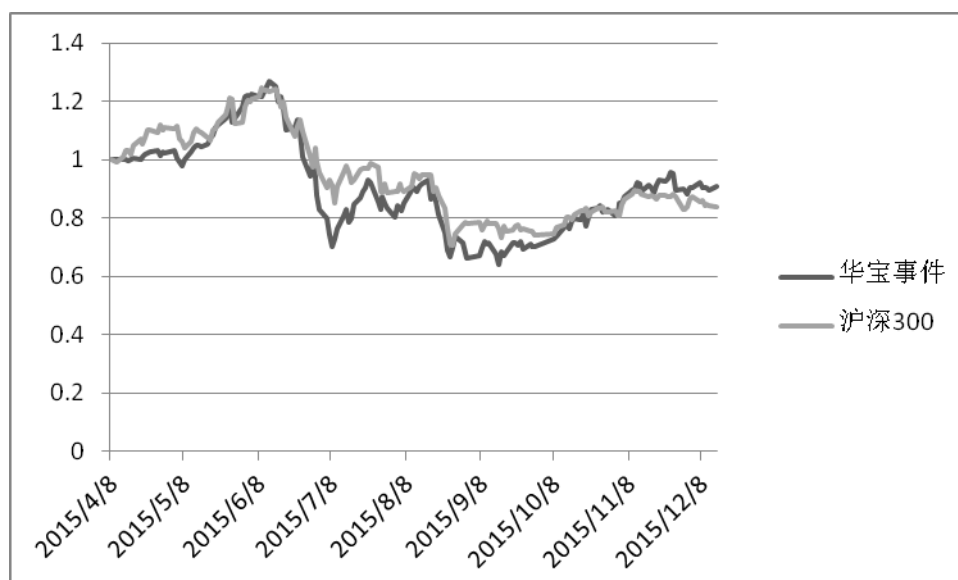


图 3-4 华宝与沪深 300 收益对比

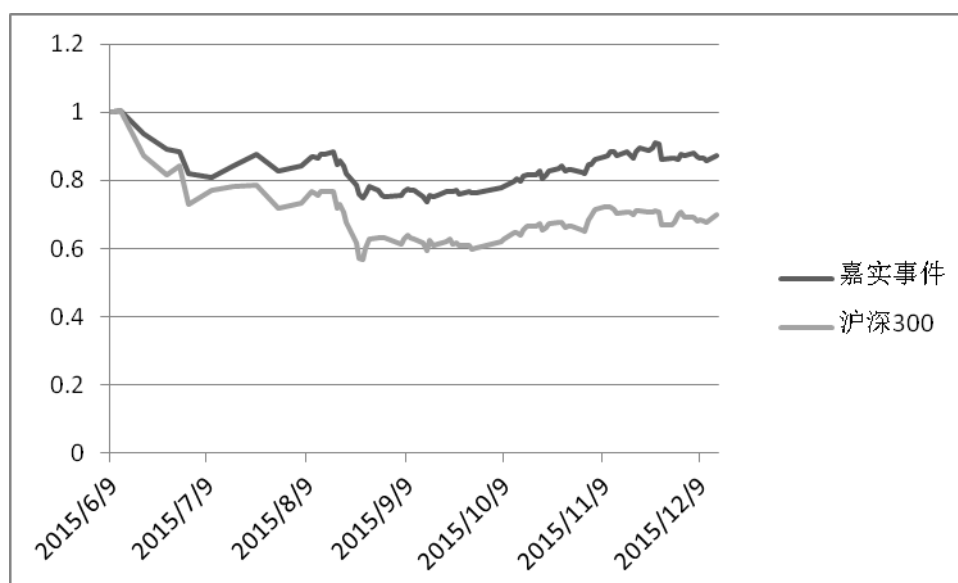


图 3-5 嘉实与沪深 300 收益对比

第 4 章 支持向量机的原理

4.1 机器学习原理

支持向量机是上个世纪 90 年代中期被提出的，它是一种机器学习方法，企图在结构风险最小的原则上，尽可能地减少样本的数量，获得同样的统计效果。因此它在解决小样本、非线性及高维模式识别中表现出许多特有的优势，是克服“维数灾难”、“过学习”等传统困难的有利武器。众多研究表明，支持向量机确实在预测精度上超越了人工神经网络等方法。支持向量机最初的逻辑就是如何找到一个函数将一组数据进行有规律的分类，这种应用随后发展到了很多领域的研究中。

SVM 对数据进行分类的基本原理如下。根据给定的训练集 $\{(x_1, y_1), (x_2, y_2) \dots (x_l, y_l)\}$ ，其中 $x_i \in \mathbb{R}_n$ ，表示第 i 个训练输入， $y_i = \pm 1$ 代表分类输出。另外给出待预测数据集： $x_{l+1}, x_{l+2}, \dots, x_m$ ($m > l$)，设计非线性分类器 $h(x)$ ，使得 $y_i = h(x_i)$ ，并且能对预测数据进行正确分类，即得到对应于 x_j ($1 \leq j \leq m$) 的输出值 y_j ($1 \leq j \leq m$)。

通常支持向量机学习问题表述如下。变量 y 与 x 之间给定一个未知关系 $F(x, y)$ ，支持向量机根据 m 个独立同分布的观测样本 $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ 学习并在一系列 $\{f(x, a)\}$ 中找出一个最佳函数 $\{f(x, a_0)\}$ 来使得预期风险

$$E(a) = \int Q(y, f(x, a)) dF(x, y)$$

最小。其中 $\{f(x, a)\}$ 称为学习函数集， $a \in \Lambda$ (Λ 为广义参数集)， $Q(y, f(x, a))$ 称为损失函数，在分类问题中通常损失函数取为

$$Q(y, f(x, a)) = 0, \text{ 若 } y = f(x, a)$$

$$Q(y, f(x, a)) = 1, \text{ 若 } y \neq f(x, a)$$

学习目标使 $E(a)$ 最小，称为期望风险最小化准则。

4.2 支持向量机的分类原理

4.2.1 最优分类面

我们首先来在二维空间中的最优分类线。假设现在有两类样本，分别是下图中的实心点和空心点。如果存在一条直线 H 能够将这两类样本分开，那么直线 H 就被称为分类线。直线 H_1 、 H_2 是存在在该平面中，与直线 H 平行的两条直线，分别经过两类样本，因此被称为支持向量，并且 H_1 、 H_2 一定是与直线 H 距离最近的直线，两条线间的间隔 $d(H_1, H_2) = 2/\|w\|$ 称为分类间隔。能够使分类间隔最大的那条直线 H 就是最优分类线。

延伸到高维空间中，最优分类线就是能够使分类间隔最大的分类面。

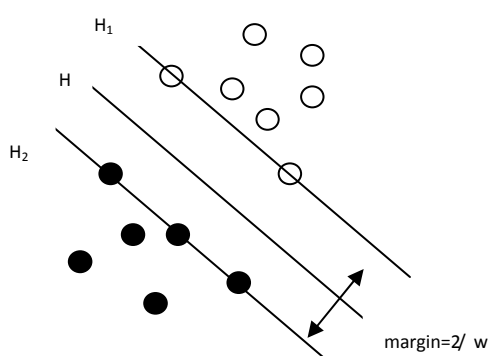


图 4.1 最优分类超平面

用拉格朗日泛函表示 $\|w\|^2/2$:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i \{y_i [(x_i, w) + b] - 1\}$$

其中 α_i 是拉格朗日因子。最小化 $L(w, b, \alpha)$ 。即使导数为零：

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0$$

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \Rightarrow \sum_{i=1}^l \alpha_i x_i y_i = w$$

由上式得到对偶优化问题：

$$W(\alpha) = \sum_{i=1}^l \alpha_i - \min_a \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j)$$

在约束条件下，最大化 $W(\alpha)$ ，即得最优解。使 α_i 的样本就是支持向量。得到

的最有分类函数是

$$f(x) = \sin \left(\sum_{i=1}^l y_i a_i K(x, x_i) + b \right)$$

其中错误！未找到引用源。错误！未找到引用源。和错误！未找到引用源。分别表示两个不同类中的支持向量。

4.2.2 支持向量分类机

上一节讲述的方法适用于线性问题，而非线性问题是可以映射为高维空间的线性问题的。该映射可以表示为：

$$x \rightarrow \Phi(x) = (a_1 \Phi_1(x), a_2 \Phi_2(x), \dots, a_n \Phi_n(x)), a_n \in R, \Phi_n(x) \in R$$

在 SVM 中，核函数错误！未找到引用源。就是用来实现这种映射的。核函数必须满足 Mercer 条件：

$$K(x, y) = \sum_{n=1}^{\infty} a_n \Psi(x) \Psi(y), a_n > 0$$

$$\iint K(x, y) g(x) g(y) dx dy > 0, \int g^2(x) dx < \infty$$

在无法直接计算错误！未找到引用源。的情况下，直接将低维空间变换大高维空间了。变换后的错误！未找到引用源。变为

$$\max_{\alpha} W(\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^l a_i - \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j K(x_i, x_j) \right\}$$

得到的最优分类函数为错误！未找到引用源。。

以下四种核函数在 SVM 中被使用的最为频繁：

- ①线性核函数：错误！未找到引用源。
- ②多项式核函数：错误！未找到引用源。
- ③径向基核函数：错误！未找到引用源。
- ④Sidmoid 核函数：错误！未找到引用源。

第 5 章 实证研究

5.1 建立 SVM 模型

5.1.1 概述

SVM 侧重研究的是针对模式识别中的线性可分的两类分类问题。而股票未来趋势即股票价格的涨或跌，其本质也是一种二分类问题。对该分类问题的应用可描述如下：

①输入训练集。在本文中，训练集由大量的股价历史参数组成，包括了一定时期内股票的收盘价、成交量、换手率等比较基本，容易获得，并且能充分反映股票市场交易情况的参数指标。其中每一个参数构成一个特征向量，训练集中的每条记录分别对应一个特定的类标签。该类标签通常表现为以往的一些经验数据。例如，一条具体训练数据的形式可设为样本向量： $(v_1, v_2, \dots, v_n; y)$ ，其中 v_n 表示特征值， y 表示类标签。

②分析输入数据。通过训练后的支持向量机能为每一个类标签形成一种准确的模型，由此得到可以对具有相同规律的样本进行预测的能力。

③准确率评估。将预测集（与训练集的格式相同）输入到上文得到的模型中能够得到相应的结果，再通过与实际观测值的对比得出该模型预测的准确率。

④选择满意的分类器对实际应用进行分类。

使用支持向量机进行股价预测主要是采用技术分析方法,通过选择输入变量，以及选择输入样本，然后通过历史数据进行训练，得出模型，并用模型对未来的数据进行预测估计。预测过程主要经过：选取样本，提取数据；确定输入向量，数据预处理；确定核函数、参数；输入训练数据，训练生成模型；输入预测数据,输出预测值；计算各个指标，并对比各个指标。整个过程见下图。

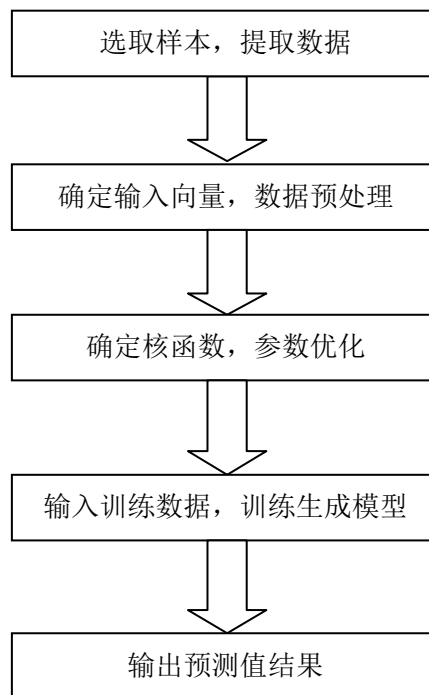


图 4-1 模型设计流程

5.1.2 模型的数学表达

设股价的涨跌 y 与资金指标 x 存在函数关系 $F(x,y)$, 则学习的目标就是在大量样本的基础上找到一个使得期望风险最小的函数 $f(x,a_0)$ 。然后输入是各种资金流向指标和收盘价, 根据函数关系 $f(x,a_0)$ 输出未来股价的涨跌。

用支持向量机寻找函数依赖关系 $F(x,y)$, 就是要求解以下二次规划问题,

1) 将数据分为训练集和预测集两个部分。设已知训练集 **错误! 未找到引用源。**, 其中 **错误! 未找到引用源。**, **错误! 未找到引用源。**, **错误! 未找到引用源。**。

2) 选取适当的核函数 **错误! 未找到引用源。** 和适当的惩罚参数 C , 构造并求解最优化问题

$$\begin{aligned}
 & \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^l \alpha_j \\
 & \text{s.t.} \quad \sum_{i=1}^l y_i \alpha_i = 0 \\
 & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l;
 \end{aligned}$$

得最优错误！未找到引用源。。

3)选取错误！未找到引用源。的一个正分量错误！未找到引用源。，并据此计算阈值

$$b^* = y_j - \sum_{i=1}^l y_i a_i^* K(x_i, x_j)$$

4)构造决策函数为

$$f(x) = \sin\left(\sum_{i=1}^l y_i a_i^* K(x, x_i) + b^*\right)$$

其中 $\sin(\cdot)$ 是符号函数。

选择最常用的高斯径向基函数作为核函数

$$K(x_1, x_2) = \exp\left(\frac{-\|x_1 - x_2\|^2}{\sigma^2}\right)$$

其中 σ^2 是参数。

我们在解最优化问题时用 MATLAB 遗传算法工具箱，通过 `crtbp` 生成种群编码，然后用 `bs2rv` 函数进行初始化，在循环中用 `reins` 函数进行重组操作找到 MSE 最小的参数。

5.2 实证工具

本文中所有仿真实验的进行都是利用了 matlab 软件，而对于 SVM 预测部分的算法运行还需配合使用由台湾大学林智仁博士等开发的 `libsvm` 软件包。其中全部实验算法由 matlab 2012a 版本完成运行。

5.3 实证数据

5.3.1 指标的选取

1.成交量指标

本文选择了成交量（总手数）、换手率、成交笔数来衡量个股在交易日内成交的频繁程度。

成交量是一种供需的表现,指一个时间单位内对某项交易成交的数量,也就是当天成交的股票总手数(1手=100股)。成交量反映了一只股票的流动性,是判断市场对它的预期的重要指标。可以说,一个成交量极小的股票,也许市场是并不看好它的。反之在一般情况下,在上涨过程中还持续放大成交量的股票,趋势向好。成交量持续低迷时,一般出现在熊市或股票整理阶段,市场交易不活跃。

换手率也称周转率,指在一定时间内市场中股票转手买卖的频率,是反映股票流通性强弱的指标之一。由于市场上众多投资者对于股价的不同预期,流通股就会在投资者之间进行份额的转换,换手率这一指标就是这一现象的体现。 $\text{换手率} = \text{成交量} / \text{流通总股数} \times 100\%$ 。高换手率的股票属于热门股,说明市场上的交易很频繁,人们持有或者抛出这只股票的意愿高;如果股票的换手率低,我们称之为冷门股,其流动性差,通俗而言就是不受关注的股票。

成交笔数是依据5秒钟内所总计形成的这一单里面包含卖单的单数,反映了一定时间内成交的数量。以中国中车(600176)为例。中国南车于2014年10月28日发布停牌公告,直到当年12月31日复牌并发布了与中国北车合并的预案。但是市场对南北车合并的反映并不是从发布公告时才开始,实际上我们从中国南车28日停牌前的20个交易日的成交次数以及资金净流入额便可以看出,未发生的事件已经驱动了投资者的交易。本文观测了中国南车2014年9月22日至2014年10月24日的20个交易日的成交次数。前七个交易日,成交次数基本在2500次/日上下震荡,之后成交变得频繁,在停牌前的最后一个交易日10月24日,成交次数达到4674次。

2. 资金流向指标

资金流向指标主要选择了观测主力资金净流入、超大单净流入、大单净流入、中单净流入、小单净流入这五个指标。我们定义超大单是大于等于50万股或者100万元的成交单,大单是大于等于10万股或者20万元且小于50万股和100万元的成交单,中单是大于等于2万股或者4万元且小于10万股和20万元的成交单,小单是小于2万股和4万元的成交单。

资金净流入的计算公式: $\text{资金净流入} = \text{流入资金} - \text{流出资金}$,如果是正值表示资金净流入,负值则表示资金净流出。

资金流向测算的是推动指数涨跌的力量强弱,这反映了是更多的资金看多还是更多的资金看空。指数处于上升状态时产生的成交额是推动指数上涨的力量,指数下跌时的成交额是推动指数下跌的力量。特别是机构投资者操控的大额资金,会对股价造成更大的影响。例如,公司发布了资产重组的公告后,市场上的资金为了取得良好的投资效益就会争先流入到该股中,造成个股价格迅速攀升,长期投资效益也体现在了短期价格暴涨上。另外还有资金流出的例子。正如重庆

啤酒（600132）在 2011 年 11 月 25 日开始的一轮暴跌，起因是重庆啤酒研发乙肝疫苗失败，研发投入颗粒无收。投资者认为这起事件将严重影响到该公司的业绩，于是大量资金外逃，造成了重庆啤酒股票近 300% 的亏损。

5.3.2 样本的选取

1. 国企改革

国企改革是 2015 年改革的主旋律，由此派生的投资机会也贯穿全年，资本市场的热炒更是一大看点。通过比较 2008-2014 年各行业主营业务收入同比增速，建材水泥、钢铁有色、煤炭、航运、船舶制造、军工或因产能过剩的问题，或因经济发展阶段、产业生命周期、国家战略等原因，很有可能被纳入央企整合名单。本文选取了在 2015 年 12 月 8 日经国务院批准资产重组的五矿系，重组预期非常高并伴随股价异动的中粮系以及因重大重组事项而停牌的中远系作为国企改革事件的代表。

（1）五矿系

本文从五矿系中选取了五矿发展、金瑞科技、厦门钨业、株冶集团、赣锋锂业五只股票。国资委网站 12 月 8 日称，经报国务院批准，中国冶金科工集团有限公司整体并入中国五矿集团公司，成为其全资子公司。消息一出，五矿发展（600058）大涨 6.2%，金瑞科技（600390）涨停，株冶集团（600961）上涨 5%，五矿稀土（000831）上涨 3.4%。五矿系的资产重组主要是为了有效解决产能过剩的问题，使企业进一步集中资源，提高自身竞争力。

（2）中粮系

本文从中粮系中选取了中粮屯河、中粮地产、中粮生化三只股票。因为受到中远系、中海系资产重组消息的影响，2015 年 12 月 14 日，午后中粮系 3 股盘中直线拉升，均以涨停收盘。实际上在此之前中粮系的股票就经常出现异动，尤其是在传出央企资产重组的进一步消息时。虽然现在还并没有关于中粮系资产重组的确切消息，但人们对其的预期已经充分反映在了股价上。

（3）中远系

本文从中远系中选取了中国远洋和中远航运两只股票。这两只股票均已经在 2015 年 8 月 7 日因为中远系与中海系的重组而停牌，复牌时间暂时未知。2015 年 12 月 11 日，国务院国资委网站发布消息，经报国务院批准，中国航运业两大巨头中国远洋运输（集团）总公司与中国海运（集团）总公司启动重组。中远和中海重组合并后将成为世界最大的远洋集团，有利于中国远洋航运在世界范围的竞争，对“一带一路”，尤其是海上丝绸之路的发展将是有利的促进。

2.二胎政策

在中国逐渐进入老龄化社会的情况下，中央通过了《关于全面两孩政策改革完善计划生育服务管理的决定》，全面两孩将于 2016 年 1 月 1 日正式实施。在未来的 10 年内，我国将会形成一个规模适度的婴儿潮，尽管这次婴儿潮并不能彻底扭转我国人口结构老化的大趋势，但是放宽“二胎”政策有利于提升消费拉动经济的力度，还能培育服务业，创造大量就业机会，有助于经济转型升级，释放巨大的改革红利。

在二胎政策事件中，本文选取了奥飞动漫、贝因美、光明乳业、哈尔斯、江南高纤、仟源医药、山大华特、卫星石化、新海股份、伊利股份这十只股票，分布的行业包括：儿童医药、乳业、母婴生活用品、玩具行业、迪士尼概念等。

3.SDR

2015 年 12 月 1 日凌晨 1 点，国际货币基金组织正式宣布，人民币将于 2016 年 10 月 1 日加入 SDR(特别提款权)。人民币纳入 SDR 可以认为是人民币作为区域性结算货币的阶段性的成功，也标志着人民币国际化将更下一城，这是中国在国际市场上获得认可的表现。未来这一举动还将促进我国更多领域的资本开放和金融改革。人民币加入 SDR，受益的行业包括：纺织、出口贸易、汽车、化工、钢铁、金融、跨境电商等。本文选取 SDR 事件的 10 只股票为安洁科技、汇金股份、聚龙股份、跨境通、鲁泰 A、外运发展、物产中大、小商品城、长江投资、中国银行。

5.4 实证结果

5.4.1 预测结果

选取上文三十只股票事件发生前后 30 个交易的数据，将三个事件的数据分开处理。国企改革一共 300 个数据，210 个进行训练，90 个进行预测；二胎政策一共 299 个数据，210 个进行训练，89 个进行预测；SDR 一共 300 个数据，210 个进行训练，90 个进行预测。

表 5-1 国企改革预测值和实际值对照

预测值	实际值	预测值	实际值	预测值	实际值
7.534%	5.26%	2.691%	7.01%	2.762%	5.28%
1.075%	0.87%	-2.097%	-8.15%	1.755%	0.59%
0.137%	0.98%	-4.160%	-10.02%	0.972%	1.16%
10.977%	9.99%	-1.089%	-10.01%	-4.455%	-1.64%
0.837%	-1.87%	-3.138%	-9.97%	2.994%	5.85%
-2.123%	-0.63%	-1.421%	-9.96%	-2.516%	-2.60%
-3.317%	-0.53%	1.499%	-10.04%	7.336%	2.35%
-5.988%	-4.39%	5.773%	10.02%	8.145%	1.43%
-1.362%	1.06%	2.907%	10.04%	6.807%	4.53%
-6.858%	-2.33%	0.368%	9.97%	1.949%	-0.15%
-7.076%	-2.83%	-1.437%	-3.76%	1.137%	-0.15%
0.343%	2.04%	-2.981%	-9.96%	3.143%	2.55%
-3.838%	1.32%	2.916%	1.88%	-1.707%	-1.75%
-6.814%	-2.54%	4.516%	8.53%	-2.054%	-2.83%
1.758%	0.69%	2.039%	5.98%	-3.545%	-0.92%
-5.065%	-0.29%	0.057%	-1.18%	10.657%	4.63%
1.008%	0.75%	6.790%	6.56%	8.674%	9.08%
-10.477%	-9.90%	2.885%	7.76%	-0.620%	-8.05%
-4.382%	-3.94%	-3.468%	-3.79%	0.198%	-0.15%
-0.740%	2.25%	-1.983%	-10.03%	-1.089%	-1.92%
0.312%	0.78%	-2.624%	-8.06%	-1.751%	-2.55%
-0.065%	0.39%	2.670%	9.98%	-1.901%	3.24%
-3.316%	-2.17%	2.298%	0.17%	-1.780%	-2.99%
3.018%	1.31%	2.321%	-1.02%	7.123%	4.77%
-6.174%	-4.84%	3.261%	2.31%	5.810%	5.66%
-0.171%	0.41%	9.496%	10.03%	-6.080%	-2.64%
-4.944%	-2.36%	0.077%	-1.29%	-6.376%	-7.43%
-4.471%	-3.25%	-1.446%	-0.31%	0.782%	-0.23%
7.936%	10.01%	14.616%	10.04%	1.941%	2.17%
-5.279%	-2.60%	-8.725%	-9.96%	-1.511%	-2.12%

表 5-2 二胎政策预测值和实际值对照

预测值	实际值	预测值	实际值	预测值	实际值
0.443%	-0.38%	1.039%	0.51%	-0.023%	1.73%
2.027%	1.54%	0.900%	2.06%	0.755%	0.57%
1.627%	2.75%	-5.316%	-0.74%	0.366%	1.19%
-1.482%	-0.37%	3.460%	3.23%	-2.131%	-1.36%
1.524%	2.13%	0.465%	5.63%	-2.662%	-0.56%
-0.813%	-2.08%	1.932%	3.37%	1.237%	0.82%
-3.944%	-5.00%	-1.820%	-4.98%	-4.566%	-2.56%
2.096%	1.66%	4.358%	10.01%	-1.160%	0.06%
-1.128%	-0.86%	-1.704%	-3.07%	0.537%	0.64%
-2.907%	-2.61%	-0.774%	-1.91%	-0.954%	-1.21%
1.812%	2.68%	1.919%	2.66%	-1.203%	0.52%
-0.736%	0.00%	-0.801%	-0.22%	0.279%	-0.06%
3.335%	5.89%	1.256%	-3.98%	-2.840%	-0.77%
0.336%	-0.55%	1.136%	2.97%	-4.414%	-0.32%
0.352%	0.64%	8.027%	10.02%	-1.131%	-0.13%
-0.736%	-0.73%	2.204%	3.78%	0.110%	0.45%
-5.460%	-9.46%	-1.639%	-8.74%	-2.313%	-3.62%
-0.933%	-2.03%	2.945%	9.95%	-2.032%	-1.14%
-0.568%	0.00%	-0.806%	0.23%	-0.211%	0.41%
1.052%	1.24%	-2.118%	-4.34%	2.584%	3.92%
1.444%	2.56%	3.769%	7.05%	-2.089%	-0.32%
0.764%	-0.30%	-1.197%	-1.08%	-5.228%	-1.69%
0.960%	1.50%	3.288%	6.81%	0.953%	0.46%
-0.454%	-3.35%	1.308%	0.14%	-4.278%	-2.11%
0.100%	-0.61%	-2.878%	-2.04%	2.356%	-0.20%
-0.207%	-0.41%	-2.275%	-3.30%	0.216%	-0.47%
0.153%	0.00%	-0.906%	-0.41%	-2.402%	-0.88%
2.748%	2.37%	-1.928%	-0.71%	4.625%	2.05%
0.776%	4.12%	-1.371%	-0.30%	1.422%	0.34%
4.832%	7.03%	0.755%	3.16%		

表 5-3 SDR 预测值和实际值对照

预测值	实际值	预测值	实际值	预测值	实际值
-1.699%	-0.64%	-0.502%	-1.81%	2.596%	4.49%
-3.308%	-1.01%	-2.909%	-2.35%	0.543%	-0.95%
-1.620%	-2.05%	-0.960%	-0.04%	-1.033%	-1.20%
-3.809%	-3.80%	-3.961%	-3.55%	-0.091%	-1.22%
0.517%	0.00%	-1.110%	-1.27%	-0.693%	-0.74%
-3.458%	-0.20%	1.677%	2.16%	-0.655%	0.75%
-0.486%	-0.99%	-0.453%	-2.39%	0.770%	0.25%
0.686%	0.80%	-0.204%	-1.43%	0.853%	0.25%
1.087%	0.50%	0.739%	2.06%	0.806%	0.25%
-0.568%	-0.79%	-0.221%	-0.05%	-1.571%	-0.74%
3.542%	8.55%	2.506%	1.93%	-0.819%	-0.74%
-2.359%	-3.02%	2.314%	2.88%	0.249%	0.50%
0.197%	-0.57%	2.098%	0.22%	3.869%	0.99%
-0.377%	-1.33%	3.552%	2.27%	1.114%	-0.25%
-4.910%	-10.01%	-5.783%	-10.00%	-1.572%	-2.95%
-1.849%	-2.35%	-1.479%	-5.37%	1.419%	4.05%
-2.924%	-0.11%	-0.436%	0.35%	-1.633%	-2.19%
1.218%	2.96%	-0.743%	-0.75%	3.416%	3.98%
3.009%	4.47%	1.647%	2.62%	4.389%	3.83%
-3.221%	-1.43%	-4.084%	-3.53%	-1.646%	-3.92%
-2.412%	-0.83%	-0.216%	0.87%	-1.836%	-1.20%
-3.543%	-4.38%	-2.250%	-4.74%	-1.767%	-1.46%
-0.304%	0.22%	-0.092%	0.32%	1.008%	1.23%
-1.480%	-0.87%	1.512%	1.37%	0.732%	-0.97%
-0.342%	-0.55%	-0.531%	-1.09%	-1.298%	-0.74%
-0.146%	1.88%	1.753%	2.95%	2.005%	1.49%
-0.557%	-0.11%	1.313%	0.56%	-1.020%	-1.46%
-0.908%	-0.65%	1.734%	2.19%	0.466%	0.00%
3.632%	3.49%	1.019%	2.09%	-0.247%	0.74%
-0.242%	-1.79%	-0.062%	-2.05%	0.326%	0.25%

我们在解最优化问题时用 MATLAB 遗传算法工具箱，通过 `crtbp` 生成种群编码，然后用 `bs2rv` 函数进行初始化，在循环中用 `reins` 函数进行重组操作找到 MSE 最小的参数。从优化结果来看，模型预测的拟合度比较高，虽然在变动幅度上有一定的误差，但是趋势上基本是一样的。三个事件的 MES 均不超过 0.025，P 值也在 5% 以内，因此我们可以认为模型是符合标准的。

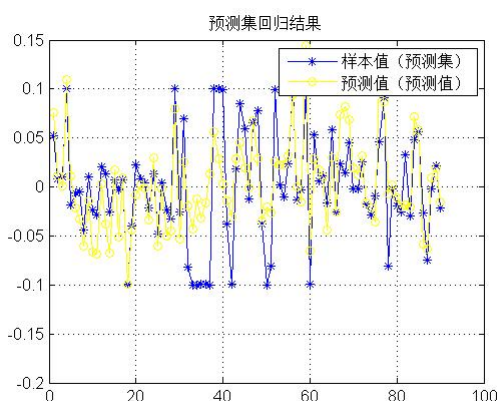


图 5-1 国企改革预测集回归

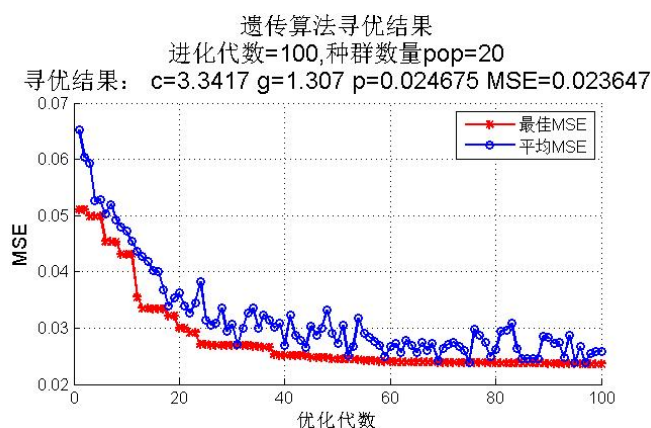


图 5-2 国企改革遗传算法寻优结果

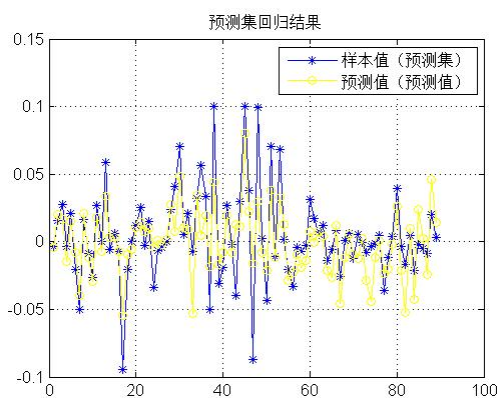


图 5-3 二胎政策预测集回归

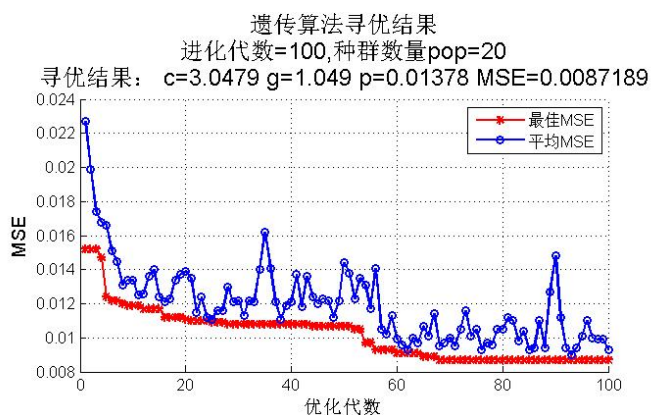


图 5-4 二胎政策遗传算法寻优结果

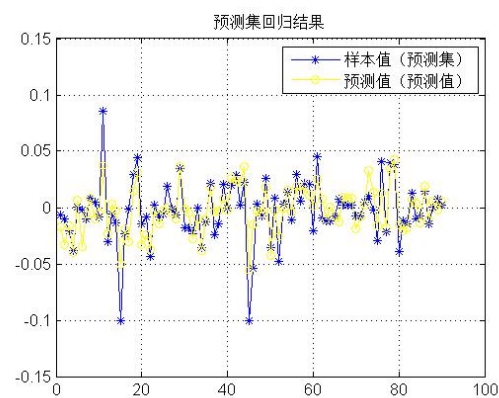


图 5-5 SDR 预测集回归

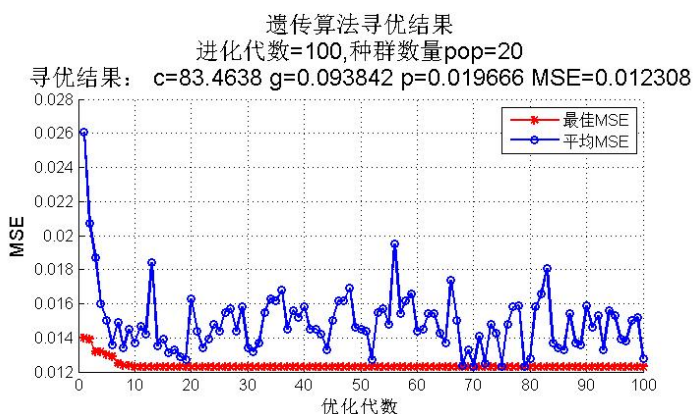


图 5-6 SDR 遗传算法寻优结果

5.4.2 检验结果

接下来，文章将得到的预测结果与观测的实际值进行了差异显著性检验。使用的检验工具是 Excel，检验方法的 t 检验和 F 检验。

表 5-4 预测结果 t 检验

	预测值	实际值
平均	-0.00027	0.000468
方差	0.001067	0.001638
观测值	269	269
假设平均差	0	
df	513	
t Stat	-0.23235	
P(T<=t) 单尾	0.408178	
t 单尾临界	1.647829	
P(T<=t) 双尾	0.816357	
t 双尾临界	1.964599	

表 5-5 预测结果 F 检验

	预测值	实际值
平均	-0.0027	0.000468
方差	0.001067	0.001638
观测值	269	269
df	268	268
F	0.651324	
P(F<=f) 单尾	0.000239	
F 双尾临界	0.817661	

从表 5.4 的 t 检验中，我们可以看到 P 值是大于显著水平 0.05，并且 t 值小于 t 临界值，因此预测值与实际值的差异是不显著的。

从表 5.5 的 F 检验中，我们得到的结果显示 F 值小于 F 临界值，表明两组数据没有显著差异。

没有使用平稳性检验是因为，本文将每日的资金流向对股价的影响看做是一次独立的事件，没有将时间序列因素考虑在内。

5.4.3 简化结果

如果我们把结果更简化一些，不预测具体的涨跌幅，仅仅预测是涨是跌（即涨幅为正还是为负），结果的准确度更高。也就是说，SVM 在仅预测股价短期趋势的功能上准确率可以达到 80%以上，这样的结果非常令人鼓舞。

在实际操作中，投资者可以将前一日资金流向的参数指标输入到模型中，会得到下一个交易日该股的预计涨跌幅，若为正值可以选择做多，若为负值则可以选择减仓或平仓。

表 5-6 简化预测结果

	国企改革	二胎政策	SDR
训练样本数	210	210	210
预测样本数	90	89	90
预测准确率	83.33%	85.39%	84.44%

第 6 章 总结和展望

自上世纪 90 年代，来自瑞银华宝的分析师爱德华科斯纳第一次提出了事件驱动投资策略这个概念，如今该策略已经成长为与行业配置相提并论的策略之一。将资金合理地配置到不同行业或者不同主题中，可以有效对冲股票的系统性风险。事件驱动投资策略属于主题投资的另一种形式，是根据受到不同事件影响进行分类并且组合投资的策略。而在全球化的经济形势下，单纯用行业或者区域的定义已经不能准确划分成千上万的股票，这时主题投资应运而生。它并不是按照一般的逻辑使用基本面分析或者技术面分析进行选股，而是从宏观到行业，再到公司个体地挖掘驱动产业或个股未来发展趋势的潜在因素，特别是依据某些重大事件或某种预期，通过分析和挖掘形成与这些重大事件或预期相匹配的投资主题，然后在该主题中寻找可以受益的，并且基本面技术面都良好的公司进行投资，最后构建一揽子的事件型主题投资股票组合。

股价预测一直以来是人们研究的热点，近年流行的技术分析指标如 MA、KDJ、MACD 等多被作为预测股价的输入变量。但是在中国 A 股市场，受事件驱动的个股比比皆是，传统的技术指标恐怕无法赶上内幕消息传播的速度。因此建立事件驱动策略模型很有必要，我们检测资金流的异动，通过模型计算做出量化买入或卖出策略。

从上文的结果中可以看出，支持向量机能够对股价有较为准确的预测，面对具有复杂性和波动性的股价能得出这样的预测结果已经是相当不易。通过国企改革、二胎政策和 SDR 三个事件中挑选的三十只股票的资金流向数据，进行训练和预测后的结果表明，在事件驱动型策略下的支持向量机股价预测模型可以在显著水平内较好的预测股价的涨跌幅，也可以比较准确地预测股票的多空方向，对股价的变化趋势判断较准，在股市中具有广泛的应用前景。并且，从实用角度来看，在误差范围内可以较为准确地预测股价的涨跌幅。

在实际的操作中投资者可以将前一日资金流向的参数指标输入到模型中，会得到下一个交易日该股的预计涨跌幅，若为正值可以选择做多，若为负值则可以选择减仓或平仓。

然而本文依旧没有突破的一点在于，忽略了时间序列的作用，没有考虑到资金的流入过流出会是一个持续变量的过程。

参考文献

- [1] Fama, Eugene F. Efficient Capital Market: A Review of Theory and Empirical Work [J] . Journal of Finance, 1970.
- [2] Mandelbrot, B. B., The Variation of Certain Speculative Prices, Journal of Business, 36, 1963.
- [3] French, K. and G. William Schwert. and Stambaugh, R. Expected Stock Returns and Volatility [J]. Journal of Financial Economics, 1986, (1987) 3- 29.
- [4] Trafalis T, Inee H. Sport vector machine for regression and applications to financial forecasting. 2000: IEEE: 2000. P. 348 — 353.
- [5] Tay F, Cao L. Improved financial time series forecasting by combining support vector machines with self-organizing feature map. Intelligent Data Analysis. 2001; 5 (4): 339 — 354.
- [6] Cao L, Tay F. Financial forecasting using support vector machines. Neural Computing & Applications. 2001; 10 (2): 184 — 192.
- [7] Kim K. Financial time series forecasting using support vector machines. Neurocomputing. 2003; 55 (1-2): 307-320.
- [8] Huang W, Nakamori Y, Wang S. Forecasting stock market movement direction with support vector machine. Computers and Operations Research. 2005: 32 (10): 2513 — 2522.
- [9] Chen S, Jeong K, Hrdle W. Support vector regression based GARCH model with application to forecasting volatility of financial returns. China Center for Economic Studies, Fudan University. 2007.
- [10] 杨一文, 杨朝军. 基于支持向量机的金融时间序列预测[J]. 系统工程理论方法应用. 2005: 14 (002): 176-181.
- [11] 汤凌冰, 盛焕烨, 汤凌霄. 基于小波支持向量机的金融预测[J]. 湘潭大学自然科学学报, 2009, 31(1): 12-15.
- [12] 吴江, 李太勇. 基于加权支持向量机的金融时间序列预测[J]. 商业研究, 2010 (1): 138-140.
- [13] 姚潇, 余乐安. 模糊近似支持向量机模型及其在信用风险评估中的应用[J].

系统工程理论与实践, 2012, 32(3): 549-554.

[14] 赛英, 张凤廷, 张涛. 基于支持向量机的中国股指期货回归预测研究[J]. 中国管理科学, 2013, 6(3): 35-39.

[15] 肖玉航. A 股“主题投资”具有必然性[J]. 股民之家, 2013, 39: 87.

[16] Lakonishok. J, Shleifer. A, Vishny. R . Do institutional investors destabilize stock prices? Evidence on herding and feedback trading[J]. NBER working paper, 1990: 3846.

[17] Wermers, R. Mutual fund herding and the impact on stock prices[J]. Journal of finance, 1999(54): 581-623.

[18] 宋军, 吴冲锋. 证券市场中羊群行为的比较研究[J]. 统计研究, 2001 (11): 23-27.

[19] 李雪, 孔凡航, 尹奇. 主题投资策略中羊群行为的实证研究[J]. 北京工商大学学报(社会科学版), 2008, 23(01): 87-91.

[20] 李备友, 张桂艳, 路英, 李守伟. 基于事件驱动的证券市场波动形成机制分析[J]. 华东经济管理, 2012, 26(3): 93-98.

[21] 杨阳, 李伟, 魏先华. 事件驱动投资策略及其影响因素的实证研究[J]. 数理统计与管理, 2014, 33(3): 519-530.

[22] 张玉川, 张作泉. 支持向量机在股票价格预测中的应用[J]. 北京交通大学学报: 自然科学版. 2007; 31 (006): 73-76

致 谢

随着毕业论文的完成，两年的研究生生涯也要结束了。两年的时间虽然算不上长，但却是 24 岁人生中最宝贵的收获和回忆。

首先感谢我的指导老师杨励教授两年来对我的关心、帮助和指导。这篇论文得以完成，承蒙杨励教授从论文选题、框架安排及论文写作各个环节的悉心指导。在学习过程中，杨励教授坦诚正直的为人、严谨务实的治学态度、诲人不倦的施学作风以及认真负责的工作态度深深地感染了我，在帮我提高论文写作的同时也为我今后的工作生活树立了良好的学习榜样。

感谢我的父亲母亲多年对我学业和梦想的信任与支持，你们永远在对我无私地奉献，你们永远是最坚强的后盾和最温暖的港湾。

感谢在两年的学习时间里教授各门课程的老师。感谢他们认真、负责的教学和指导。他们传授给我的专业知识是我完成这篇论文的基础，分享给我的人生经历是我不断成长的源泉。

感谢 MIB 中心的各位老师和同学。感谢老师们为我顺利完成学业提供了坚强的后勤保证和热情帮助，感谢同学们两年里的关心、帮助与鼓励。感谢你们给了我这么多一生难忘的美好回忆。

最后，向百忙之中抽出时间参加本次论文答辩的老师们致以深深的谢意！