

大数据背景下风险事件对证券价格 影响因子分析

指导教师：林路教授

项目组成员：王芳芳 石建栋 谭文米 谷豆豆 撒媛媛 车悦
李超楠 陆军 李劭珉 马文佳

目录

第一章 研究背景、目的和意义.....	1
1.1 研究背景.....	1
1.2 研究目的.....	2
1.3 研究意义.....	2
第二章 文献综述.....	4
2.1 事件研究法概述.....	4
2.2 事件研究法在金融领域中的应用.....	4
2.2 市场效率的检验.....	5
第三章 数据描述.....	9
第四章 模型方法.....	11
4.1 事件研究法.....	11
4.1.1 方法概述.....	11
4.1.2 模型步骤.....	11
4.2 支持向量机.....	14
第五章 模型思路.....	16
5.1 事件研究法.....	16
5.2 支持向量机方法.....	20
第六章 结论.....	28
参考文献.....	29

第一章 研究背景、目的和意义

1.1 研究背景

近年来随着经济的发展，人民的生活水平普遍得到了提高，个人财富得到了积累，但是同时通货膨胀日益加重，使我们积累的财富遭到贬值，因此选取合适的理财产品成为人民日益关注的话题。传统的理财方式以银行存款为主，银行存款的零风险性受到人们的青睐，但是银行存款利率并不能有效地规避通货膨胀风险，而投资股票作为一种高收益高风险的理财方式越来越被人们接受。

自我国股票市场成立以来，其对经济建设、社会发展提供了巨大的支持。与此同时，作为现代金融体系的重要组成部分，股票市场也在规模、结构、运行等方面实现了巨大进步。短短 20 多年的时间里，我国股票市场得到了迅速发展，特别是 2000 年以后的历次改革，更是加速了股票市场的规范化进程，推动了其与世界各主要金融市场之间的联系，从而促进了我国的股票市场逐渐融入国际金融市场，并成为其重要组成部分。从我国股票市场的实际作用看，其在各个发展阶段都能够显著地影响经济的发展与社会的稳定：一方面，股票市场作为企业融资的一个重要平台，其运行效率的高低直接决定了上市企业的融资能力，进而影响企业的经营与发展，并对宏观经济产生影响；另一方面，我国的股票市场从成立之初就是众多投资者广泛参与的投资场所，无论是机构投资者还是个人投资者，都在市场中频繁地进行投资活动，因此，股票市场具有较高的活跃性。市场的每一个发展过程与阶段都直接关系到众多投资者的利益并影响着社会的稳定。然而，从我国股票市场的实际发展过程看，其经历了多次较为深刻的改革，随之而来的是股票市场的巨大波动，从而影响整个金融市场，甚至影响国家的政治环境。

股票价格的波动是股票市场运行的基础，也是保持市场活力的必要条件，但是，当波动过于剧烈并呈现出异常状态时，就会破坏市场的均衡价值，带来大量的投机行为推动下的系统性泡沫或者资产价格的非理性下降，从而造成市场功能的丧失。目前，大量的投资者涌入金融市场，但是如何在众多的不确定因素中

分析有利因素并获得利润并非易事，显然，如果能够找到科学的方法来预测股票价格的变动方向，更甚者，如果能够预测股票价格变动的幅度，则将会给投资者带来不可估量的福利，同时还能够使监管者更好地监督管理金融市场，使市场管理者更好地进行管理资产。然而大量研究表明，影响证券市场的因素种类繁多，错综复杂，且证券市场是一个高度复杂的非线性系统，股票价格的变动受到这些复杂因素的影响，尤其是随着金融市场的发展以及我国政府对新兴技术产业的支持，证券市场上市公司不断增多，特别是中小企业上市公司更是如雨后春笋，高风险事件愈发呈现出频发性和多样化，这些高风险事件的发生轻者会直接导致部分证券价格大幅波动，对公司的经营与发展造成不良影响，重者会直接导致企业发生大规模违约风险，恶化公司债权安全，特别是中小企业或者新兴产业，其面对风险事件时的应对能力相对较弱，风险事件发生时更加容易产生信用违约风险，甚至使得企业面临倒闭的风险，因此在发生风险事件时对于股票价格的预测更是一个十分艰巨而又必不可少的任务。

1.2 研究目的

在世界经济不断演进和发展的过程中，股票市场扮演者重要角色，特别是几年来世界经济大幅震荡、国际形势错综复杂以及自然灾害频发的今天，股票市场面临着更多的威胁，股票价格的暴涨暴跌仍然时有发生。几年来，众多中小型企业兴起，其对于突发风险事件的抵抗力依旧很薄弱，在风险事件的影响下，其股票价格受到市场冲击的可能性更大，对于利好事件，其股票价格会出现明显的上涨，然而对于利空事件其反应也很强烈，会出现股票价格大幅度下跌，公司效益直线下降，甚至会出现破产倒闭的情况，给投资者以及自身带来巨大的风险。因此研究风险事件对于股票价格的影响不仅能帮助公司更好地进行管理，做出正确的管理决策，还能帮助投资者规避风险，避免财产的大量亏损，同时还能够使市场监管者更好地监督管理证券市场，使证券市场健康地运行，维持相对稳定。

1.3 研究意义

在以往的研究中，研究股价走势通常采用时间序列分析法和影响因素分析法。时间序列分析法一直是研究股价过去价格走势的一种重要工具，时间序列模

型研究的主要是股票价格曲线走势本身，模型的目的是从股票价格走势曲线中找到趋势和波动率信息，并基于此预测股价。但是时间序列分析是预测经济指标的重要工具，对于股市这种行情的变化受到国家宏观经济形势、公司运营、股民情绪等等诸多因素相关联的情况，预测结果并不尽如人意。尤其是当上市公司有突发消息发生时，时间序列分析的预测方法就更显得无能为力了，而我国的证券市场由于发展时间短、投机性强、散户居多等特性更容易受到事件风吹草动的影响，有时一个市场传闻都会影响股价的大幅波动。时间序列分析的方法由于没有将影响股价走势的外部因素考虑进去，致使其存在了极大的局限性。因此在本文中我们采用另一种研究风险事件下股票价格变动的方法，即影响因素分析法。影响股价的因素包括国家宏观经济、政治环境、国内经济形势、行业发展情况、公司竞争情况、盈利状况、经营管理情况、投资者情绪等方面，其中国家宏观经济情况和行业情况相对比较稳定，短期内变动不大，而公司情况则随着公司自身的治理发展状况有较大的变动，通过对不同的风险事件进行分析，并选取公司的一些财务指标来刻画这些影响因素，找到不同发生风险事件时这些指标的变化，进而找出其对股票价格的影响。通过结合阅读参考文献，我们得出结论首先应对风险事件进行分类，因为风险事件不仅有利好事件还存在利空事件，它们对于股票价格的影响是不同的，甚至同样是利空事件其对于股价影响的幅度也存在很大差异，因此结合事件研究法对风险事件下股票价格的影响变动做出判断是可行的。同时，通过事件研究法，我们还能鉴定一个突发事件是否是风险事件。

另外，由于公司的发展状况不同，在面对突发风险事件时，其敏感性是不同的，这时我们采用支持向量机的得分分类方法对公司进行分类，通过大量的训练数据，可以通过机器学习实现在一个风险事件来临时，判断其是利好事件还是利空事件。利用事件研究法和支持向量机分类方法的结合，就能较好地预测风险事件对于股价变动方向和变化幅度的影响，这样就可以在发生风险事件时做出正确的决策，不致使股票市场发生大幅度变动，影响经济环境的稳定，甚至社会环境的稳定。

第二章 文献综述

事件研究方法被广泛运用于金融经济领域中，本部分主要是对国内外目前金融经济领域的事件研究法相关文献进行梳理，主要包括三部分，首先对于事件研究法进行概述，梳理国内外文献对于事件研究法的介绍；其次，梳理事件研究法在金融经济领域的发展状况的相关文献；最后，对于事件对股票价格的影响的研究予以概述，其中事件包括互联网财经新闻、利率、突发事件、公告等。

2.1 事件研究法概述

事件研究最早是由Ball & Brown (1968) 提出的，目前它已经发展成为资本市场中研究长期收益异常现象的主要工具。事件研究法是用于测度某事件发生后的影响效果的一种实证研究方法，主要集中于探讨事件（如收购、合并、增持公告或再融资行为等）的发生对股票价格（或企业价值）带来的冲击，冲击可体现于以下几个方面：平均股价效应、股票成交量的变化、市场收益方差的变化（反映股价波动性）、经营（会计）绩效的变化等。

目前国内有关事件研究的文章也有很多：段瑞强（2004）介绍了事件研究法在金融市场中的应用；白仲光和蓝翔（2003）简述了金融事件研究法的方法和步骤；袁显平和柯大岗（2006）详述了事件研究法应用于实证研究的过程。本文通过对国内外相关文献的梳理回顾，发现事件研究分析方法一般包括为：事件定义、样本选择、正常收益率计算、异常收益率评估、数据分析与检验、实证结果解释等步骤。其中以样本选择、异常收益率评估为方法的核心。

2.2 事件研究法在金融领域中的应用

事件研究迄今为止，已有很长的历史。早在20世纪30年代，Dolly (1933) 首次对股票分割的股价效应采用事件研究法，这是“事件研究法”首次在金融领域中的应用。随后，学者们在Dolly的基础上进一步发展完善了事件研究法，并且伴随着越来越多的有关事件研究的文献，最终，Ball & Brown (1968)、Fama (1969) 奠定了较为成熟的“事件研究法”：关注度量事件期间样本证券的异常

收益均值与累积异常收益率。在事件研究方法发展和完善的整个过程中，形成了事件研究方法论这一体系。前期Fama（1969）的研究奠定了事件研究在金融领域中的位置，但它还存在一些不足，包括其研究的深度与广度，以及其应用的方法（Blume, 1971; Gonedes, 1985）。

不光是国外学者，我国国内学者也有大量的事件研究相关文献，内容范围与国外大致相似，但明显时间上滞后于国外文献。白仲光（2003）首次以中文文献介绍了事件研究法的概念，并梳理了其研究步骤，以及事件研究法的局限性。随后，国内学者以微观企业数据为基础，将事件研究法应用到金融研究领域（陈汉文和陈向民，2002；陈信元和江峰，2005）。事件研究方法论发展至今，简而言之，主要向两个方面发展：月度数据被交易日数据所取缔；估计异常收益与校对其统计显著性的方法越来越显著。事件研究在金融经济领域的应用，也不外乎在于以下四个方面。

2.2 市场效率的检验

最初，事件研究法被用于检验市场的有效性。Fama（1965）指出，如果市场上的股票价格能够充分反映所有可以获得的信息，则该股票市场是有效的。即在一个有效的市场中，市场能够迅速对信息进行反应，并且市场上的价格能够充分反映所有可得信息，不能够根据目前的信息去预测其资产价格，因此导致任何人都无法持续在金融市场获得超额收益。

其中，交易日数据较月度数据不仅更为准确，而且更精准度量了股价对事件的反应速度，同时有利于减少甚至消除市场有效性和资产定价模型共同检验的问题（Fama, 1991）。因此，日收益率数据在运用事件研究法检验市场有效性的研究中有着举重若轻的位置，并且日收益率数据的应用保证了事件研究的准确性。Brown & Wamer（1985）指出，当股价对某一事件反应强烈并发生在短短的几天内时，不同的期望收益的估计方法（如常均值模型、市场模型与资本资产定价模型等）几乎对最终的推论没有影响。以日收益率为基础的事件研究结论是，股价对于事件信息的反应与市场有效性相一致，并且是快速有效的。

然而，事件研究仅限于探讨股价对事件信息的反应，并未揭露“有多少离均差产生的剩余方差”是合理的。当股价对于事件信息的反应变得迟钝时，此时事

件研究还需处理市场有效性和资产定价模型共同检验的问题。我国证券市场，同时存在股票市场的过度反应，某一事件引起股票价格剧烈变动，超过了预期的理论水平。张人骥（1998）认为，如果股票市场存在过度反应，那市场是非有效的。我国主要是通过巨额交易、股票分割等事件研究，来进行市场有效性的检验（耿军会和石会娟，2006），有关股票市场异常现象的研究，如规模效应、季节效应也成为了中外学者研究关注的领域。

2.3 事件对股票价格的影响研究

一般来说，研究事件对于股票价格的影响，这里的事件主要包括互联网财经新闻、兼并并购、突发事件、利率等。国内外学者对有关事件对于股票价格的影响研究也大致包含了这几类事件，本部分对于事件对股票价格的影响研究，进行了梳理总结。

随着互联网媒体的兴起，互联网信息对于公司股价波动的影响是显而易见的，有关这方面的研究也比较多。学者主要是将互联网财经新闻梳理作为解释变量，并没有考虑新闻内容。最早Niederhoffer（1971）通过采用《纽约时报》重大新闻标题，并根据其内容进行分类，对标题于股市影响评级，以此来考察都有哪些新闻事件能够影响股票市场。结果表明，股市在新闻发布后的第一天时间里，反应最为强烈。杨继东（2005）首先对有关文献进行了梳理总结，探讨了媒体力量对于投资行为的影响。王攀等（2012）也采用了新闻数量作为研究指标，主要是用百度搜索结果来度量，通过实证检验发现新闻媒体关注度越高的股票，其发行价格也相对较高。

还有大部分学者关注兼并收购事件对股票市场的影响，然而结果并不一致。Cheng & Leung（2004）将1986-1994年香港联交所的36件兼并并购事件为标的，通过分析发现兼并收购事件能够促使股价上涨，并且异常收益率大于零。李善民等（2006）选取了1955年后三十年的并购事件作为研究样本，结果发现并购事件发生后的长期异常收益率并不会提高，相反小于零。张新等（2013）选取了1216起1992年至2002年这十年的兼并事件作为研究标的，并同时运用事件研究法和会计分析法对于兼并事件对于股价的影响进行对比分析，结果表明异常收益率小于零，但并不显著。

突发事件对于金融市场的影响，也是一大热点研究方向。David & Betty (2004) 指出“9.11”事件强烈冲击了美国经济，并且对股票市场产生了重大的影响。Yang & Chen (2008) 针对日本股市，以灾难伤亡为事件标的，发现该类事件对于股票市场并没有显著影响。张国清等 (2008) 以松花江事件为案例，发现事件过后，明显抑制了样本公司股价的累积异常收益。

利率水平对于股票价格的影响也不尽相同。Durham (2003) 对美国基金利率水平与股票国债收益之间的关系进行了分析，发现利率水平的调整，会对十年期的国债收益率造成一定的影响，但却对股票收益率的影响并不显著。刘行 (2010) 针对我国同业拆借利率以及沪深300指数作为样本，结果发现同业拆借利率和股票指数之间并不存在显著的相关性。

2.4 支持向量机研究背景

现在，全球很多主要金融市场中都存在股票指数交易，一些重要股票指数的变化甚至代表着整个经济运行方向。市场上，大量投资者需要对冲潜在的市场风险，套利者希望通过股指交易获利，监管者希望市场平稳运行，所有这些都依赖于股票指数的精确预测。显然，无论对金融市场研究者还是实践者，股票指数预测都具有重要意义。

然而，由于金融时间序列存在诸多混杂因素以及本身的不平稳性，金融市场预测一直是一个极具挑战性的任务。不平稳性意味着金融时间序列的分布会随着时间变化。一方面，很多偶然因素如经济政治状况、投资者预期、灾难或者战争等，都可能导致股指、汇率等金融时间序列的变化，这些外部混杂因素是我们无法用模型刻画的，同时也使得我们很难从过去金融市场行为中提取到完整信息从而完全捕捉未来价格和过去价格之间的依赖关系；另一方面，任何金融时间序列之间的关系以及金融时间序列与其他相关数据之间的关系，也可能随着时间变化。这些都造成了我们金融市场预测的困难。早期的金融时间序列分析主要是技术分析[11]，依靠图标和技术指标进行主观判断，后来随着及经济金融学研究的不断发展，研究人员开始尝试对金融时间序列进行建模分析，例如投资组合理论，套利定价理论等，再到后来时间序列的ARIMA模型、GARCH模型等。

近年来，人工神经网络已经成功地用于金融时间序列建模[20]。然而，这些

研究也同时表明人工神经网络也因为股票市场数据的噪声性和维度的复杂性而表现出不一致和不可预测性。最近,一种新颖的算法称为支持向量机(SVM)[10]用于金融时间序列预测[7-9, 21-27]与神经网络寻求局部最优不同,支持向量机是建立在结构风险最小化基础上的全局最优,不太可能发生过拟合问题。

以前大多数的交易策略依赖于金融市场价格水平的精确预测[17, 18, 29-31],然而,最近一些研究提出,根据股票价格变化的方向预测做出的交易策略,有时可能更为有效,带来更高的收益[19]。同时,对股价变化方向的预测,也直接影响投资者对金融产品的买卖决定及交易策略。

Kim, K[1](2003)用SVM预测韩国KOSPI指数的变化方向,文中选取了包括随机指标(%D, %k)、动量指标(Momentum)、相对强弱指标(RSI)、威廉指标(%R)、累积派发指标(A/D Oscillator)、顺势指标(CCI)等12个技术指标作为输入变量, SVM模型分类预测指数的变化方向,预测准确率达到57%以上,并且验证了SVM对核函数及参数选取的敏感性及SVM相对于神经网络在预测表现上的优越性。Kara Y等(2011)[2],在研究土耳其伊斯坦布尔100指数的变化方向时,同样选取了类似的10个技术指标,比较了SVM与人工神经网络的预测表现,预测准确率超过了70%。Wei Huang, Yoshiteru Nakamori等(2005)[4]将SVM模型用于预测日经225指数的周变化方向,同样验证了SVM模型在股指方向预测上相对于其他分类预测模型的优越性。

目前,针对我国股票市场指数变动方向研究预测的文章不多,王帅、尚伟(2014)[15]对沪深300指数变动方向做了类似研究,文中选取了Kara Y(2011)同样的10个技术指标做为输入解释变量,预测模型选用最小二乘支持向量分类机(LSSVM),结果沪深300变化方向预测准确率接近90%。我依文中所述,选取风险事件下股票数据及其中6技术指标,并将支持向量机结合事件研究法用于风险事件下股票收盘价变动方向预测,准确率接近于85%。

第三章 数据描述

本次项目的目的是预测风险事件发生后，公司股票日收盘价变动的方向，即预测下一个交易日的收盘价与前一日相比是涨还是跌，本文中我们用二分类变量 y_t 表示该变动方向， $y_t = 0$ 表示 t 时刻的收盘价低于前一日的收盘价， $y_t = 1$ 表示 t 时刻的收盘价高于前一日的收盘价。数据来源于 wind 数据库，选取一段时间深圳市场发生过资产重组事件的公司股票数据，提取各公司在事件发生日前后的收盘价、最高价、最低价以及市场行情数据，数据清理包括去空值、去无效值，经过多步清理程序，最终可以作为有效数据进行模型拟合的深市股票有#只。参考国内外之前的研究，选取了 6 个对股价变动代表性较好的技术指标作为输入变量，下图表给出了所选取的技术指标及相关说明。

数据提取

数据提取基于以下两个目的：

第一，计算事件发生日前后股票的累计异常收益率；

第二，利用相关数据计算模型中用到的技术指标。

步骤：

1 通过 Wind 资讯终端，得到自 2015 年一月至 2015 年 12 月深圳市场发生资产交易事件的 125 个公司以及发生人事变动事件的 558 个公司的股票代码以及事件发生日；

2 基于各公司股票代码以及事件发生日，提取出事件发生日前 50 个交易日以及后 20 个交易日的最高价、最低价以及收盘价，并相应提取该时间段代表整个深证成指的涨跌幅；

3 去除无效数据后得到对应人事变动事件 124 支股票和资产交易事件 519 支股票的有效数据，利用现有数据计算所需的 6 个技术指标。

Indicator	Formula	Description
SMA	$\frac{C_t + C_{t-1} + \dots + C_{t-9}}{10}$	<p>C_t 表示 t 时刻（事件发生日）的收盘价</p> <p>L_t 表示 t 时刻（事件发生日）的最低价</p> <p>H_t 表示 t 时刻（事件发生日）的最高价</p>
WMA	$\frac{10 \times C_t + 9 \times C_{t-1} + \dots + C_{t-9}}{10 + 9 + \dots + 1}$	
MTM	$C_t - C_{t-1}$	
CCI	$\frac{M_t - SM_t}{0.015D_t}$	
WPR	$\frac{H_t - C_t}{H_t - L_t} \times 100$	
A/D Oscillator	$\frac{H_t - C_{t-1}}{H_t - L_t}$	$M_t = (H_t + L_t + C_t) / 3$ $SM_t = \frac{1}{10} \sum_{i=1}^{10} M_{t-i+1}$ $D_t = \frac{1}{10} \sum_{i=1}^{10} M_{t-i+1} - SM_t $

第四章 模型方法

4.1 事件研究法

4.1.1 方法概述

事件研究法是一种统计方法，是在研究当市场上某一个事件发生的时候，股价是否会产生波动时，以及是否会产生“异常报酬率”，借由此种资讯，可以了解到股价的波动与该事件是否相关。具体来讲是根据研究目的选择某一特定事件，研究事件发生前后样本股票收益率的变化，进而解释特定事件对样本股票价格变化与收益率的影响，主要被用于检验事件发生前后价格变化或价格对披露信息的反应程度。事件研究法是基于有效市场假设的，即股票价格反映所有已知的公共信息，由于投资者是理性的，投资者对新信息的反应也是理性的，因此，在样本股票实际收益中剔除假定某个事件没有发生而估计出来的正常收益就可以得到异常收益，异常收益可以衡量股价对事件发生或信息披露异常反应的程度。

4.1.2 模型步骤

4.1.2.1 累计异常收益率的计算

首先计算样本股票的实际日收益率，确定事件日后，计算样本股票 i 在事件期的第 t 天的实际收益率

$$R_{it} = \frac{A_t}{A_{t-1}} - 1$$

其中， A_t 指股票在公告事件发生后第 t 天的收盘价， A_{t-1} 指股票在公告事件发生后第 $t-1$ 天的收盘价。然后，计算能够代表所选样本股票的市场指数的日收益率：

$$R_{mt} = \left(\frac{I_t}{I_{t-1}} \right) - 1$$

其中， I_t 是股票市场指数在公告事件发生后第 t 天的收盘价； I_{t-1} 指市场指数在公

告事件发生后第t-1天的收盘价。由此计算第i支股票第t日的超额收益率：

$$AR_{it} = R_{it} - R_{mt}$$

其中， AR_{it} 为股票i在事件期内第t日的超额收益率。样本股票i在事件期第t天的累计收益率以及股票市场指数在事件期第t天的累计收益率分别为：

$$CAR_{it} = \frac{A_t}{A_1} - 1, \quad CAR_{mt} = \frac{I_t}{I_1} - 1$$

样本股票i在事件期的第t天的平均累计超额收益率是 CAR_{ct} ：

$$CAR_{ct} = \frac{1}{t} \sum_{j=1}^t CAR_{ij} - \frac{1}{t} \sum_{j=1}^t CAR_{mj}。$$

假设事件没有发生，股票市场指数价格应当反映个股走向，正常收益的计算可以有以下几种方法：

(1) 均值常数模型

$$R_{it} = \mu_i + \varepsilon_{it} \quad (E(\varepsilon_{it}) = 0, VAR(\varepsilon_{it}) = \sigma_{\varepsilon_i}^2)$$

该方法假设个股的收益率总是围绕一个常数值做上下波动，均值可以通过历史数据估算。

(2) 市场调整模型

$$R_{it} = \alpha_i + \beta_i * R_{mt} + \varepsilon_{it} \quad (\varepsilon_{it} \sim N(0, \sigma_{\varepsilon_i}^2) \quad i.i.d)$$

(3) 特征基准模型

$$R_{it} = CBB_t + \varepsilon_{it} \quad (E(\varepsilon_{it}) = 0, VAR(\varepsilon_{it}) = \sigma_{\varepsilon_i}^2)$$

它的计算方法是在全市场的所有股票中，对于公告事件发生日 t 的股票 i，按照 t-1 日数据，选取 n=10 只总市值和股票 i 最接近的股票构造等权组合作为股票 i 在 t 时刻的特征基准组合，其收益率 CBB 作为 t 时刻之后的预期收益。

本文中，我们选取了市场调整模型计算正常收益，那么异常收益则为：

$$AR_{it} = R_{it} - \hat{R}_{it}$$

其中， R_{it} 为股票 i 在事件窗口期内 t 日的实际收益率。

收益模型计算下的事件期第 t 天的累计正常收益率是：

$$C\hat{A}R_{it} = \frac{\hat{R}_{it}}{\hat{R}_1} - 1$$

其中 \hat{R}_{it} 指收益模型计算下的公告事件发生后第 t 天的收盘价， \hat{R}_1 指收益模型计算下的公告事件发生后第 1 天的收盘价。

样本股票 i 在事件期的第 t 天的平均累计异常收益率是 CAR ：

$$CAR = \frac{1}{t} \sum_{j=1}^t CAR_{ij} - \frac{1}{t} \sum_{j=1}^t C\hat{A}R_{ij}$$

4.1.2.2 显著性检验

首先标准化累积异常收益为

$$SCAR_i = \frac{CAR}{\sigma_{CAR_i}}$$

其中 σ_{CAR_i} 为累积异常收益的标准差。BMP-t 检验统计量可写为

$$t_{BMP} = \frac{\overline{SCAR}}{\sigma_{SCAR} / \sqrt{N}}$$

其中， $\overline{SCAR} = \frac{1}{N} \sum_{i=1}^N SCAR_i$ ， $\sigma_{SCAR}^2 = \frac{1}{N-1} \sum_{i=1}^N (SCAR_i - \overline{SCAR})^2$ ，则有零假设

$H_0: SCAR = 0$ 。若结果接受原假设，则认为公告事件对股价没有显著的影响；若结果拒绝原假设，则说明公告事件对股价有显著影响。

4.2 支持向量机

SVM分类的主要思想是将线性可分(或近似线性可分)的样本数据点看成是n维实数空间中的点,那么存在n-1维的超平面能够把这些样本点分开。在这众多分类超平面中,希望找到一个使分类间隔最大化或者错分情况最少的超平面,叫做最优分类超平面。样本数据点非线性可分时,我们通过变换,将原数据点映射到一个更高维的空间,使原低维空间中非线性可分的样本点在新的高维空间中成为线性可分或近似线性可分的样本点。

假定有大小为l的样本 $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$,其中 $x_i \in R^n$, $y_i \in \{1, -1\}$,当样本属于第一类时, $y_i=1$, 否则 $y_i=-1$ 。构造超平面

$$\omega^* x + b = 0$$

使得当 $\omega^* x_i + b \geq 1$ 时, $y_i=1$; 当 $\omega^* x_i + b \leq -1$ 时, $y_i=-1$ 。当样本非完全线性可分时,要尽量使错分的样本量最小,引入松弛变量 $\xi_i \geq 0$, $i=1, 2, 3, \dots, l$ 。此时就得到“软化”的约束条件 $y_i(\omega^* x + b) \geq 1 - \xi_i$ 。这里 $\xi = (\xi_1, \dots, \xi_l)^T$ 体现了允许训练集

被错分的情况,用 $\sum_{i=1}^l \xi_i$ 作为一种描述训练集被错分程度的度量。显然,我们一

方面希望分类间隔尽可能大,另一方面希望错分程度 $\sum_{i=1}^l \xi_i$ 尽可能小,为此我们

在目标函数里引入惩罚参数C作为综合这两方面目标的权重。

为了广泛解决各种线性不可分问题,要引入核函数,就是要把x所在的输入空间映射到另一个空间。通过从低维空间到高维空间的合适映射,使原来在低维空间中的非线性可分变为高维空间中的线性可分,从而把在低维空间中的寻找最优“超曲面”转化为高维空间中继续寻找最优超平面问题。常用的核函数有

1. Gauss 径向核函数:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2), \quad \gamma > 0$$

2. 多项式核函数

$$K(x, x') = ((x, x') + c)^d, \quad c \geq 0, d \in N^+$$

由于两类之间的最大“间隔”为 $\frac{2}{\|\omega\|}$ ，最优分类超平面使得分类“间隔”极大化的思想使问题转化为求解下面对变量 ω 和 b 的最优化问题。引入核函数和惩罚参数后，分类问题可以转化为

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

$$s.t. \sum_{i=1}^l \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C, i = 1, \dots, l$$

最后得到决策函数

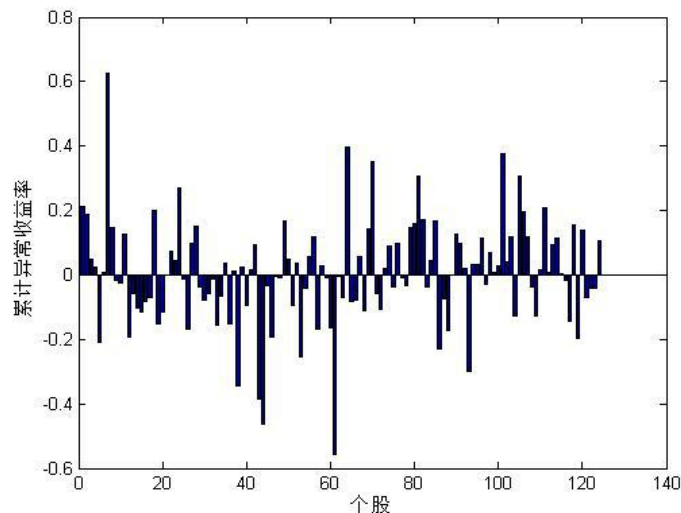
$$f(x) = \text{sgn}\left(\sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^*\right)$$

第五章 模型思路

5.1 事件研究法

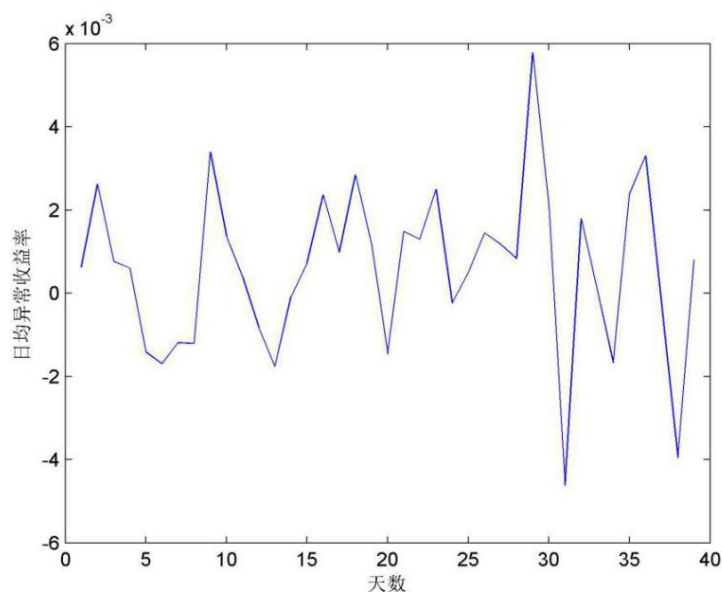
综合多篇文献，事件研究法是研究风险事件对股价影响的一种重要而有效的方法，对于我们这里的研究，首先一步需要检验该类风险事件是否对股价造成影响，也即对于指标累计异常收益率是否显著远离 0，在模型上转化为了统计学中的假设检验问题。第二步，我们希望通过一组股票累计异常收益率的均值得到该风险事件对股价的影响，即若该均值为正，则判断该风险事件对股价是利好的，反之，若异常收益率为负，则判断该风险事件对股价是利空的。然而通过对实际数据的计算，我们发现这种预测效果并不好，反思原因，我们发现虽然累计异常收益率的均值是正或负，但一组累计异常收益率中正负数量是差不多的或者说并没有明显的趋势，则我们给出这样一个一致的预测方向其效果一定是不好，由此我们希望得到一个根据不同股票事件发生前状态变化的不同预测，从而达到更好的预测效果。在此，我们考虑在下一部分中使用支持向量机的方法进行进一步的预测。

以下是运用事件研究法对异常收益率进行检验的结果及分析：

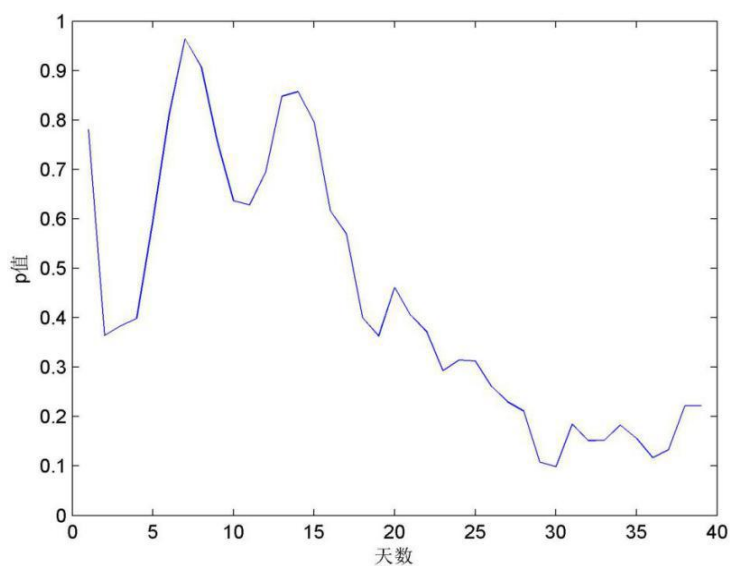


上图为人事变动风险中对应 124 个公司计算得到的累计异常收益率的分布情况。可以看出，这组数据中累计异常收益率正负值的个数是相当的，也即人事

变动风险对于不同的公司并没有一个较强的趋势；同时，大部分累计异常收益率都分布在区间 $[-0.2\%, 0.2\%]$ 中。

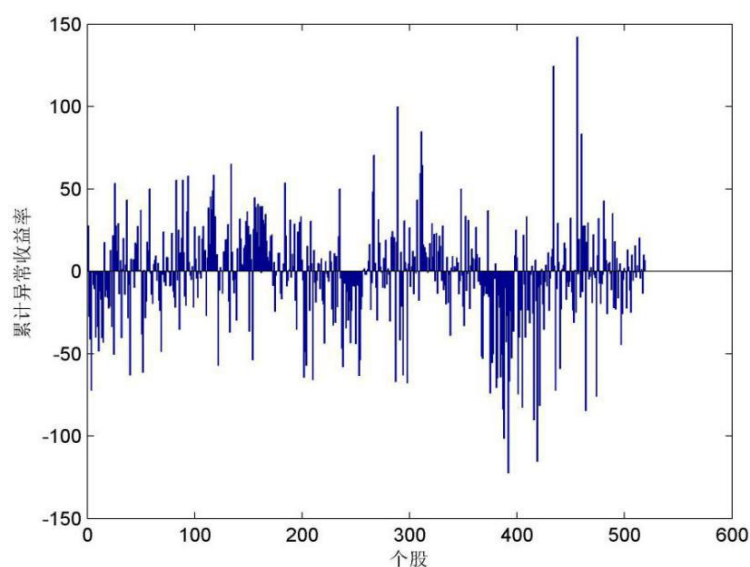


上图为人事变动风险中事件发生前后二十天异常收益率波动情况，我们认为风险事件往往使得股票价格波动更剧烈，这里我们使用异常收益率作为指标进行衡量，得到该风险事件确实对股价波动产生影响，结合下图累计异常收益率显著性检验的结果，我们可以得到人事变动风险事件对于股价波动的影响是显著的。

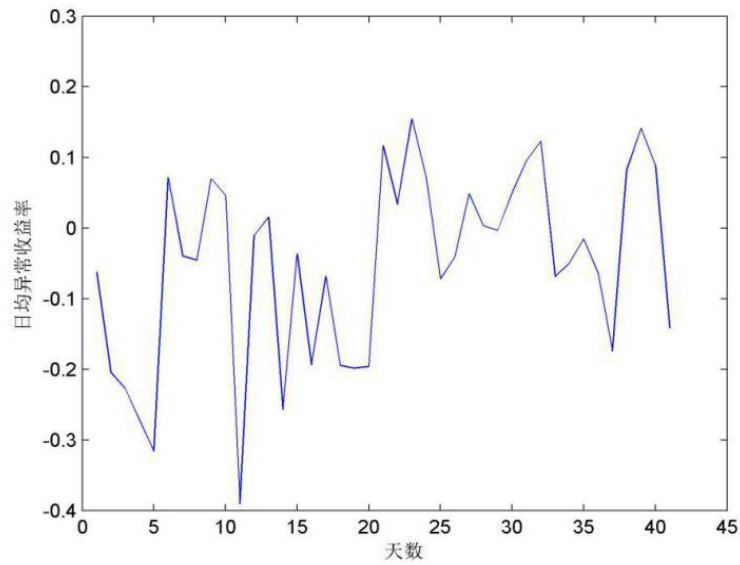


上图表示对于人事变动风险，事件发生前 20 天到事件发生后 20 天之间 40 天的检验 p 值，每天分别计算累计异常收益率并按照上一章中提出的假设检验方

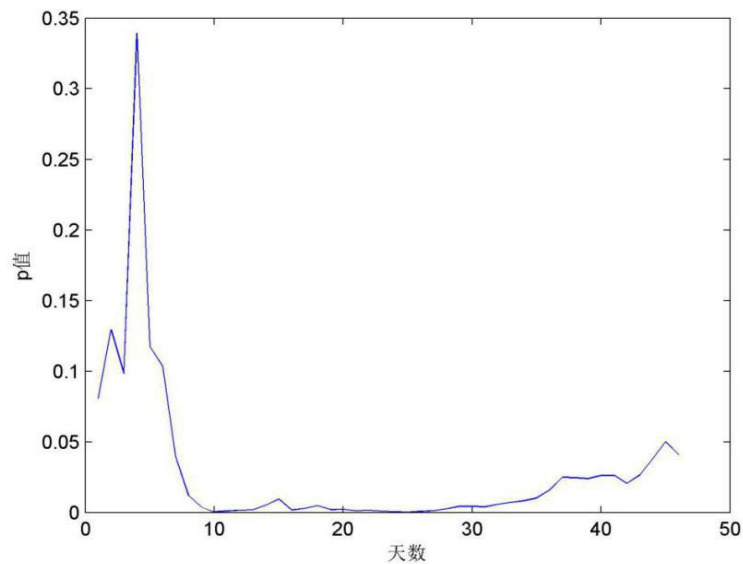
法进行假设检验，我们得到的衡量显著性水平的指标 p 值不断下降，在 30-35 天左右， p 值下降到 0.1 左右，也即我们可以以 90% 的把握拒绝原假设，说明人事变动风险对于股票价格波动在假设检验角度上是较为显著的。（这里我们参考支持向量机方法中选取阈值的方法，同样可以设定合适的阈值，对于假设检验问题本身就会有一定的概率犯错误的概率，而且对于一些确实对该风险不敏感的股票我们无法通过这种方式给出一个一致的判别，通过设定阈值我们能够使得 p 值减小，得到更加显著的检验结果）



上图表示 519 支股票发生资产交易后累计异常收益率的分布情况，上图表明，累计异常收益率的正负数量大致相当，并没有明显的趋势，我们依然不能给出一致判断给出事件对于股价影响是利好或者利空，同时我们发现异常收益率的值大部分落在 $[-50\%, 50\%]$ 之间。



上图表示有效数据为 519 支股票发生资产交易时对应日均异常收益率得到的折线图，同样对应事件发生前 20 天到事件发生后 20 天的情况，对比事件发生前一周与事件发生后 20 天之内日均异常收益率波动，发现波动明显增加，可以说明风险事件发生导致股价波动剧烈。



上图表示在资产交易数据下，对应事件发生前 25 天到事件发生后 20 天，每天对累计异常收益率进行显著性检验得到 p 值变化趋势，上图表明 p 值在事件发生后随天数的增加不断降低到 0.05 以下，较好的说明该风险事件对于股价波动的影响是显著的。

5.2 支持向量机方法

我们重新思考最初的目标，希望在已知事件发生前所有数据情况下，预测该风险事件发生后价格的波动方向，对于价格波动方向的考量，我们选用前文中事件研究法中用到的异常收益率来衡量，在这里选用异常收益率来衡量股价的涨跌是较为合理的，其中异常的思想类似于数学中的差分，更大程度上考量了该风险在目前状态下对于股价的影响，我们希望对不同股票得到考虑到各股票目前状态时的判断结果。

首先我们确定能够使用的协变量，根据之前的分析，我们能够使用风险事件发生前的一切信息，综合分析各种文献及衡量，我们选择事件发生前一段时间的六个代表性技术指标作为协变量；对于响应变量，我们一直选取累计异常收益率正负作为股价涨跌标准的技术指标，此时的异常收益率应当为事件发生后一段时间算得的异常收益率。因此，我们将历史上每次风险事件发生前对应的六个指标与事件窗后的累计异常收益率正负作为训练集中的一个向量，对于我们收集的该风险事件的所有数据中，我们得到一组向量作为训练数据集并将其输入支持向量机，使用 R 语言已有的支持向量机程序包，我们可以得到对于累计异常收益率正负的判断。此时能够达到的目标即为，每给出一组风险事件发生前的股票数据，计算对应六个指标并输入支持向量机，我们能得到事件窗后对应股票的累计异常收益率正负，也即该风险对于该股票价格的影响。这里我们对于累计异常收益率靠近 0 的那些股票，我们可以考虑设定一个阈值对这些数据进行合理取舍，而此阈值取决于客户对于风险的喜好，如果客户厌恶风险，那么我们可以将阈值取大，使用阈值以外的即更加远离 0 的累计异常收益率对应的股票数据进行分析计算，我们得到的预测将会更加精确，但此种情况下我们放弃了对于阈值以内很多股票的盈利机会；反之如果客户偏好风险，那么我们可以将阈值取小，使用阈值以外的累计异常收益率，我们得到的预测没有之前的预测准确，但我们为客户抓住更多盈利机会。

具体操作中，为了避免输入指标不同数量级之间造成淹没，原始输入数据首先进行归一化处理，变换到 $[-1, 1]$ 范围内，提高预测结果准确性。用到的归一化公式：

$$x_i' = \frac{x_i - E(x_i)}{\sqrt{Var(x_i)}}$$

根据前文所述，预测模型可以表示为

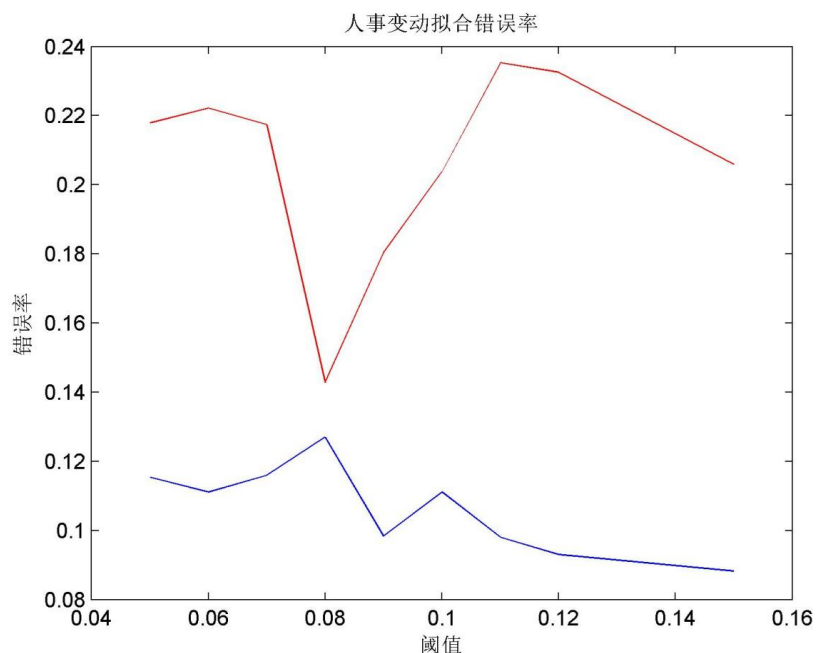
$$y_i = f(I_i^1, I_i^2, \dots, I_i^6)$$

其中， $I^i, i=1,2,\dots,6$ ，代表 6 个不同的输入技术指标， $y_i = 1, -1$ 表示预测的变化方向。我们把全部数据分为两段，前段作为训练集，后段作为测试集，为了使结果更好一点，分别计算了训练集为 0.5、0.6、0.7、0.8、0.9 的数据的结果，并分别对其用 tune.svm 函数进行参数调优。在核函数的选择中，高斯径向基核函数无论在预测结果的准确性还是训练时间上都要优于多项式核函数，因此这里也选择径向基函数作为支撑向量机分类的核函数。实际预测结果如下：

关于人事变动的预测错误率

	linear	radial
5%	0.2179487	0.1153846
6%	0.2222222	0.1111111
7%	0.2173913	0.115942
8%	0.1428571	0.1269841
9%	0.1803279	0.09836066
10%	0.2037037	0.1111111
11%	0.2352941	0.09803922
12%	0.2325581	0.09302326
15%	0.2058824	0.08823529

为了使结果看得更清晰，我们将该表作图如下，由该图可以看到，当选择线性函数（linear）作分类时，随着阈值的增大，即风险偏好降低，风险厌恶提高，错误率先降低后增大，而且错误率比较高；当选择径向基函数（radial）作分类时，随着阈值的增大，发生的错误率降低，正确率提高；与我们的认知是一样的，而且错误率最小可以达到 0.08823529，所以使用径向支持向量机方法得到的结果是非常好的。

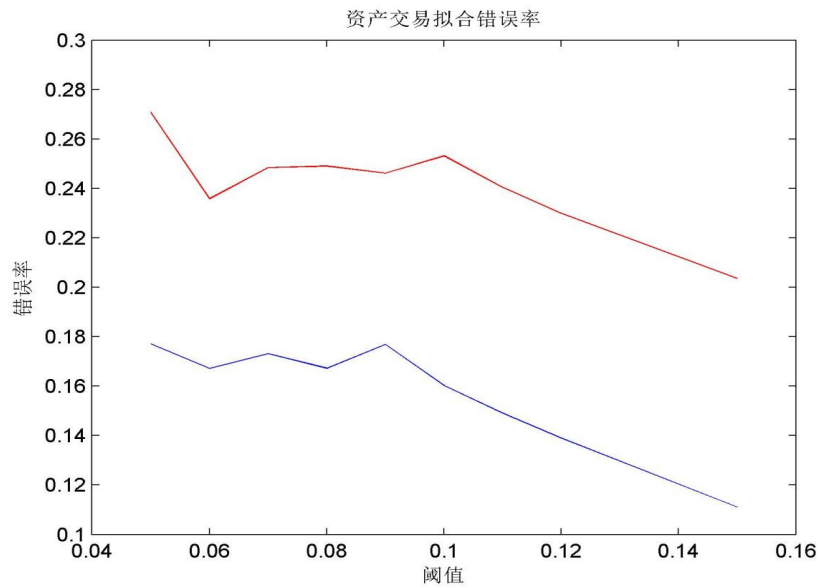


关于资产交易的预测错误率

	linear	radial
5%	0.2708333	0.1770833
6%	0.2358209	0.1671642
7%	0.248366	0.1732026
8%	0.2491103	0.1672598
9%	0.2461538	0.1769231
10%	0.2531646	0.1603376

11%	0.2403846	0.1490385
12%	0.2299465	0.1390374
15%	0.2037037	0.1111111

为了使结果看得更清晰，我们将该表作图如下：



由该图可以看到，当选择线性函数（linear）作分类时，随着阈值的增大，即风险偏好降低，风险厌恶提高，错误率趋势是降低的，但是错误率比较高；当选择径向基函数（radial）作分类时，随着阈值的增大，发生的错误率降低，正确率提高；与我们的认知是一样的，而且错误率最小可以达到 0.1111111，所以使用径向支持向量机方法得到的结果是非常好的。

为了使结果更具有说服力，分别对人事变动和资产组合的检验集股票做了涨跌预测，结果如下表所示：

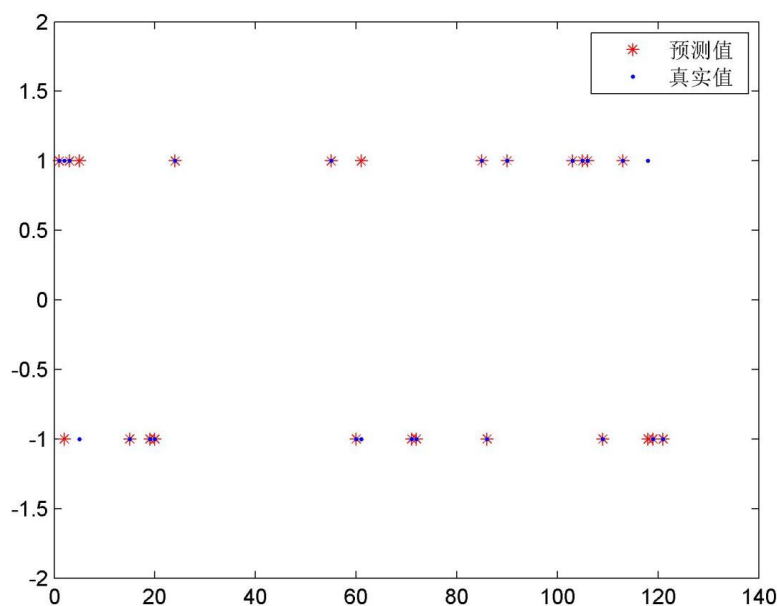
关于人事变动的股票涨跌检验集预测错误率

	有效数据	0.5	0.6	0.7	0.8	0.9
5%	78	0.205128	0.21875	0.166666	0.25	0.375
6%	78	0.277777	0.379310	0.454545	0.266666	0.5
7%	69	0.257142	0.25	0.523809	0.571428	0.571428
8%	63	0.4375	0.192307	0.368421	0.307692	0.142857
9%	61	0.290322	0.2	0.263157	0.384615	0.285714
10%	54	0.407407	0.363636	0.352941	0.454545	0.333333
11%	51	0.307692	0.190476	0.1875	0.272727	0.166666
12%	43	0.318181	0.333333	0.307692	0.333333	0.4
15%	34	0.352941	0.428571	0.363636	0.428571	0.5

该表第一行表示，将累积异常收益率大于阈值 5% 的股票都看作上涨的，记为+1；累积异常收益率小于阈值 5% 的股票看作下跌的，记为-1；然后将这部分数据（有效数据）取出，记为 A，取 A 数据的 0.5, 0.6, 0.7, 0.8, 0.9 倍分别作为训练集，其余数据作为检验集。用支持向量机对训练集分类后，再对检验集进行检验，得到错误率就是第一行中的结果，其余行类似。

该表结果表明，当阈值为 5% 时，其有效数据为 78 支股票，训练集取为该数据的 0.7 倍时，错误率比较低，为 0.166667；当阈值增大时，错误率波动比较大。在阈值为 8% 时，错误率为 0.142857，在阈值为 11% 时，错误率为 0.166667。随着阈值增大，错误率相对来说，比较高，造成这种情况的原因是数据本身较少，取了训练集后，用于检验的数据就更少了，出现这种波动是可以理解的。总之，在阈值为 5% 时，错误率达到 0.166667，这是一个相对不错的结果。

为了使结果看得更清晰，我们将检验集股票涨跌真实情况和预测情况作一个对比，其中股票价格上涨记为+1，股票价格下跌记为-1，作图如下：



该图为阈值为 5%时，错误率为 0.1666667 的股票涨跌对比情况。由图可以看到，图中预测值和真实值的重叠度较高，说明预测的结果是不错的。所以直观地可以看到，预测的结果是比较好的。

由于人事变动的数据比较少，并且为了防止某个风险事件的偶然性，我们又重新找了一个风险事件——资产交易，且针对该分险事件所找的数据也比较多。资产交易共找了 558 支股票。该分险事件的结果如下：

关于资产交易的股票涨跌预测错误率

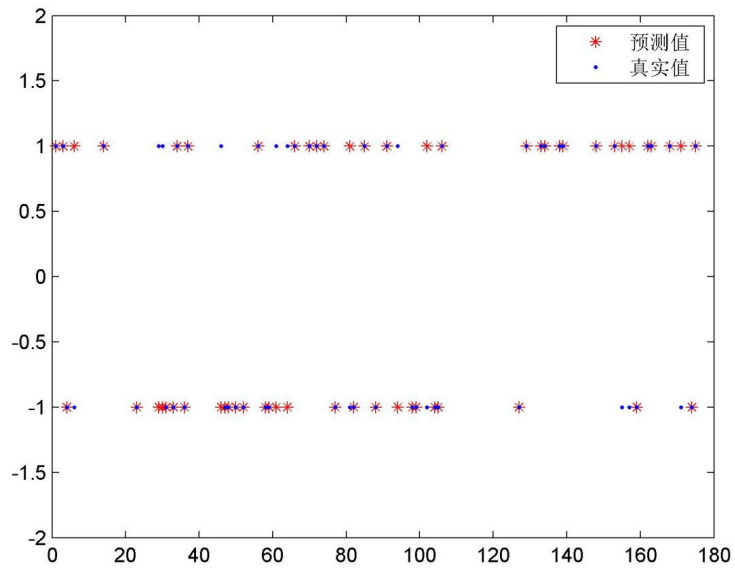
	有效数据	0.5	0.6	0.7	0.8	0.9
5%	362	0.251162	0.246073	0.262857	0.253164	0.241134
6%	335	0.231155	0.201117	0.221519	0.205673	0.204545
7%	306	0.324324	0.231707	0.229166	0.204724	0.220339
8%	281	0.337209	0.296774	0.283687	0.293650	0.304347

9%	260	0.246913	0.236111	0.309523	0.284482	0.276190
10%	237	0.360544	0.398496	0.327731	0.277777	0.285714
11%	208	0.330708	0.313043	0.352381	0.336956	0.325301
12%	187	0.324561	0.336538	0.28125	0.298850	0.316455
15%	128	0.1625	0.183098	0.206349	0.237288	0.25

该表的解释与上述人事变动表格的解释类似。第一行表示，将累积异常收益率大于阈值 5% 的股票都看作上涨的，记为+1；累积异常收益率小于阈值 5% 的股票看作下跌的，记为-1；然后将这部分数据（有效数据）取出，记为 A，取 A 数据的 0.5, 0.6, 0.7, 0.8, 0.9 倍分别作为训练集，其余数据作为检验集。用支持向量机对训练集分类后，再对检验集进行检验，得到错误率就是第一行中的结果，其余行类似。

该表结果表明，当阈值为 6% 时，其数据为 335 支股票，训练集取为该数据的 0.6 倍时，错误率比较低，为 0.2011173；当阈值增大时，错误率波动比较大。在阈值为 15% 时，错误率为 0.1625。尽管随着阈值增大，错误率相对来说，比较高，波动也比较大，但是相对人事变动来说就好一点了，整体看，错误率比人事变动的都要小一些；而且，在阈值为 15% 时，错误率达到所列出的错误率的最小值，这是符合正常的观念的（随着风险厌恶的降低，风险偏好的提高，出现错误的概率是下降的）。但是随着阈值的增大，错误率还是有比较大的波动（波动比人事变动的波动小），造成这种情况的原因可能还是数据较少，也可能是非系统误差比较大。总之，这个风险事件的结果相对人事变动来说好了很多，也侧面地说明了人事变动的结果不如这个结果好的原因是数据少；同样，这个风险事件的错误率最小 0.1625，并且在各个阈值中，错误率大多在 20%-30%，这个结果相对来说也是不错的。

同样地，为了使结果看得更清晰，我们将检验集股票涨跌真实情况和预测情况作一个对比，其中股票价格上涨记为+1，股票价格下跌记为-1，作图如下：



该图是阈值为 6% 时，错误率为 0.2011173 的股票涨跌对比情况。由图可以看到，预测值和真实值的重叠度较高，即对于使用 60% 数据作为训练集，40% 数据作为检验集情况下，模型对检验集的预测正确率达到 80% 左右，在预测股价变动方向上表现出良好的预测效果。

第六章 结论

综合运用事件研究法与支持向量机方法，我们得到了最初的目标：给出事件发生前股票相关数据以及发生的风险事件，能够较为准确的预测出事件发生一段时间之后累计异常收益率的正负以及大致范围，也即能够预测该风险事件在公司发生风险时刻对股价的利好或利空影响。

首先，选用在很多文献中使用的事件研究法，选取异常收益率作为指标，很好的衡量了在该风险事件的影响下股价的波动情况，去除了股票价格本身的趋势，同时选取累计异常收益率使得我们的分析结果更加稳健，我们成功检验出选取的风险事件（包括人事变动和资产交易）对于股价变动的影 响是显著的；同时，通过分析累计异常收益率的分布，我们发现风险事件对于不同公司或不同股票的影响程度甚至方向是不同的，我们无法简单的运用事件研究法的方式得到对于股价变动方向的准确预测，因此我们考虑使用在统计学及机器学习中广泛使用的支持向量及方法对股价变动方向进行预测。

利用支持向量机对股票价格进行涨跌分类，当将所有数据都用来分类时，检验所得到的涨跌预测值与真实的涨跌值拟合度非常高，对人事变动事件做检验，错误率最小可以达到 0.08823529，对资产交易做检验，错误率最小可以达到 0.11111111，说明用该方法对其进行分类是非常有效的。为了使结果更有说服力，我们又将数据分为训练集和检验集，对检验集做检验后得到的错误率增大了，但是也没有很大，在可以接受的范围内。总之，用支持向量机做股票的涨跌分类预测，可以对股票做出一个很好的涨跌预测。

参考文献

- [1] Ahern, R. K., (2009), “Sample selection and event study estimation”, *Journal of Empirical Finance*, 16, 466 - 482.
- [2] Banz, R. W., (1981). “The relationship between return and market value of common stocks”. *J. Financ. Econ.* 9, 3 - 18.
- [3] Boehmer, E., J. Musumeci, and A. B. Poulsen, (1991), “Event study methodology under conditions of event induced variance”, *Journal of Financial Economics* 30, 253 - 272.
- [4] Brown, S. J., Warner, J. B., (1985). “Using daily stock returns: the case of event studies”. *J. Financ. Econ.* 14, 3 - 31.
- [5] Carhart, M. M., (1997). “On persistence in mutual fund performance”. *J. Finance* 52, 57 - 82.
- [6] Campbell, J. Y., Lo, A. W., MacKinlay, A. C., (1997). “The Econometrics of Financial Markets”. Princeton University Press.
- [7] Corrado, C. J., (1989), “A Nonparametric Test for Abnormal Security-Price Performance in Event Studies”, *Journal Of Financial Economics*, 23, pp. 385-395
- [8] Cowan, A. R. (1992). “Nonparametric event study tests”. *Review of Quantitative Finance and Accounting* 2, 343 - 358.
- [9] Daniel, K., Grinblatt, M., Titman, S., Wermers, R., (1997). “Measuring mutual fund performance with characteristic-based benchmarks”. *J. Finance* 52, 1035 - 1058.
- [10] Dolley, J. (1933), “Characteristics and Procedure of Common Stock Split-Ups,” *Harvard Business Review*, 316-326.
- [11] Kim K. Financial time series forecasting using support vector

machines. *Neurocomputing*, 2003;55:307–319.

[12] Leung M T, Daouk H, Chen A S. Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting*, 2000;16(2): 173–190.

[13] Huang W, Nakamori Y, Wang S Y. Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 2005; 32(10): 2513–2522.

[14] Vapnik VN. An overview of statistical learning theory. *IEEE Transactionsof Neural Networks* 1999;10:988 – 99.

[15] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers. *Neural processing letters*, 1999; 9(3): 293–300.

[16] 申瑞娜, 曹昶, 樊重俊. 基于主成分分析的支持向量机模型对上海房价的预测[J]. *数学的实践与认识*, 2013, 43(23): 11–16.

[17] 刘悦. 基于支持向量机的股票价格预测: [硕士学位论文], 天津: 南开大学. 2013.

[18] 袁磊. 基于 ARIMA–LSSVM 混合模型的股指预测研究: [硕士学位论文], 黑龙江: 哈尔滨工业大学. 2015.

[19] 黄振. 基于支持向量机的沪深 300 指数预测研究: [硕士学位论文], 湖南: 湖南大学. 2012.