

中图分类号:C93-03
学科分类号:120110

论文编号:1028709 15-S037

硕士学位论文

基于独立成分分析的 国际原油价格分析预测研究

研究生姓名	李慧萍
学科、专业	管理科学与工程
研 究 方 向	能 源 金 融
指 导 教 师	周 鹏 教授

南京航空航天大学

研究生院 经济与管理学院

二〇一五年三月

Nanjing University of Aeronautics and Astronautics

The Graduate School

College of economics and management

**Analysis and Prediction of
International Crude Oil Price based on
Independent Component Analysis**

A Thesis in

Management Science and Engineering

by

Li Huiping

Advised by

Prof. Zhou Peng

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Management

Mar, 2015

承诺书

本人声明所呈交的硕士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。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，也不包含为获得南京航空航天大学或其他教育机构的学位或证书而使用过的材料。

本人授权南京航空航天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承诺书)

作者签名：_____

日 期：_____

摘 要

作为全球最重要的基础化工原料和战略资源，原油对世界各国的经济发展、国防安全等具有强大的支撑作用，国际原油价格的波动在世界经济政治形势中占据举足轻重的地位。随着中国对外石油依存度的日趋攀升，世界原油价格波动对我国社会经济发展的冲击愈加显著。深入理解国际原油市场，挖掘国际原油价格的隐藏波动机制，合理预测未来油价走势对于维护我国经济发展、企业运作和人们生活显得尤为重要。

鉴于国际原油市场价格形成机制具有复杂性和高度非线性特点，本文首次尝试选取独立成分分析方法(ICA)作为国际原油价格数据的预处理技术，探究国际原油市场价格波动的影响因素；结合支持向量回归模型(SVR)，提出了基于独立成分分析方法的国际原油价格非线性集成分析预测模型 ICA-SVR²：选取国际基准油价 WTI 原油现货价格作为预测目标。作为原油价格数据的预处理技术，独立成分分析实现了国际油价序列的特征提取，寻找出国际原油价格时间序列中的共同隐藏驱动因素，从原始 4 列国际油价信号提取出长期趋势、重大事件项和市场短期波动三个独立成分；接着针对三项独立成分，利用网格搜索法和交叉验证法相结合的方式进行参数寻优，构建出不同的 SVR 模型分别进行预测，获得各序列预测值；最终将三项独立成分的预测值进行 SVR 集成得到 WTI 油价最终预测值，实现了 WTI 油价走势的精准预测。通过对 WTI 原油价格的模拟预测表明：基于独立成分分析的国际原油价格分析预测模型不仅可以挖掘出国际原油市场的隐藏驱动因素，有效提升预测结果的可解释性；同时实现了国际原油价格的精准预测，证明了 ICA-SVR² 模型具有解释能力强、鲁棒性高、学习与泛化能力强的特点。

关键词：国际原油价格；油价影响因素；独立成分分析；支持向量回归；集成分析预测

Abstract

As the most important chemical raw materials and strategic resources, crude oil plays a crucial supporting role in national economic development as well as defense security. The fluctuations of international crude oil price act a pivotal role in international and domestic political and economic situation. As the increasing foreign oil dependence in China, the fluctuations of crude oil price play a more and more important impact in the social and economic development of our country. For the national development, enterprise operation and daily life of people, it is very important to understand the international crude oil market in a deep way and mining the intrinsic elements of oil price fluctuation and forecasting the future trend.

Due to the complexity and high nonlinearity of international oil price mechanism, this paper attempts to apply independent component analysis to crude oil prices data as a preprocessing tool, to explore the influence factors of fluctuations in crude oil prices. Through integrating independent component analysis and support vector regression, this paper proposes a nonlinear integration analysis prediction model(ICA-SVR²) to predict the West Texas intermediate spot prices: independent component analysis realizes the feature extraction of crude oil prices series and digs the common driving factors out, original four time oil price series are decomposed to three parts: long-term trend, extreme things and short-term market volatility; specific support vector regression models are established for the detail components, Finally, this paper reconstructs the results of three prediction value to get the final ensemble forecasting result of oil price. The illustrative example of WTI crude oil spot price indicates that analyzing and predicting model based on independent component analysis can not only explore the underlying driving factors of crude oil prices fluctuations and enhance the interpretability of predicted results, the ensemble paradigm also obtain high prediction accuracy, which illustrates the strong robustness, high learning and generalization ability, as well as explanation ability of proposed ICA-SVR² model.

Keywords: International crude oil price; Influence factors; Independent Component Analysis; Support Vector Regression; Integrated analysis and forecasting

目 录

摘 要	I
Abstract.....	II
第一章 绪论	1
1.1 选题背景及意义.....	1
1.1.1 选题背景.....	1
1.1.2 选题意义.....	2
1.2 论文的结构安排.....	2
1.2.1 研究内容、方法与思路.....	2
1.2.2 论文章节安排.....	4
1.3 论文的主要创新点.....	5
1.4 本章小结	5
第二章 国际原油价格分析预测基本现状.....	6
2.1 国际原油价格影响因素分析.....	6
2.1.1 供需因素.....	7
2.1.2 经济金融因素.....	11
2.1.3 突发事件.....	13
2.2 国际油价预测分析研究现状.....	15
2.2.1 石油价格分析预测理论方法.....	16
2.2.2 石油价格分析预测计量模型.....	17
2.2.3 数据驱动型模型.....	19
2.3 本章小结	21
第三章 独立成分分析方法与支持向量回归模型.....	23
3.1 独立成分分析(ICA).....	23
3.1.1 独立成分分析定义.....	23
3.1.2 独立成分估计原理.....	25
3.1.3 Fast-ICA 算法实现.....	27
3.1.4 独立成分分析方法的适用性.....	29
3.2 支持向量回归模型(SVR).....	30
3.2.1 SVR 模型原理.....	30
3.2.2 核函数的选取.....	32
3.2.3 SVR 模型参数的寻优.....	33
3.2.4 SVR 模型的适用性.....	34
3.3 本章小结	35
第四章 国际原油价格集成分析预测模型.....	36
4.1 ICA-SVR ² 模型框架.....	36
4.2 ICA-SVR ² 集成分析预测流程.....	37
4.3 集成分析预测中的关键问题.....	38

4.3.1 独立成分个数的确定.....	38
4.3.2 SVR 核函数的选择.....	39
4.3.3 SVR 参数寻优.....	40
4.4 本章小结	40
第五章 实证分析—WTI 原油价格分析预测	42
5.1 实证分析设置.....	42
5.1.1 数据来源和预测评价准则.....	42
5.1.2 实验设置和实现.....	42
5.2 油价序列的独立成分分析.....	43
5.2.1 数据描述.....	43
5.2.2 确定独立成分个数.....	43
5.2.3 独立成分的估计及其经济含义.....	44
5.3 独立成分的 SVR 预测.....	45
5.3.1 SVR 预测模型的参数寻优.....	45
5.3.2 三项独立成分的预测模型及预测结果.....	48
5.4 独立成分预测结果的 SVR 集成.....	49
5.4.1 SVR 集成模型的参数寻优.....	49
5.4.2 SVR 集成模型的预测结果及鲁棒性分析	51
5.5 预测结果对比分析.....	52
5.6 本章小结	54
第六章 结论与展望	55
6.1 研究结论	55
6.2 不足和展望	56
参考文献	57
致 谢	61
作者在攻读学位期间发表的学术论文目录.....	62

图表清单

图 1.1 本文的研究技术路线图	3
图 2.1 国际原油价格的影响因素关系结构图	6
图 2.2 世界经济增长率与世界石油消费增长率(单位: %)	7
图 2.3 世界探明原油储量(单位: 亿万吨).....	9
图 2.4 非商业净多头与 WTI 期货价格	11
图 2.5 WTI 期货价格与美元指数关系图.....	12
图 2.6 世界原油市场突发事件	14
图 2.7 国际油价分析预测模型	15
图 2.8 国际油价分析预测模型发展趋势.....	22
图 3.1 独立成分分析方法	24
图 3.2 支持向量结构示意图	30
图 4.1 ICA-SVR ² 集成预测框架.....	37
图 5.1 四列国际原油价格时间序列.....	43
图 5.2 国际油价时间序列中的三项独立成分	45
图 5.3 不同 ε 水平下 SVR 预测模型的均方根误差.....	45
图 5.4 参数 σ 对 SVR 预测模型性能的影响($\varepsilon = 0.01$).....	47
图 5.5 参数 C 对 SVR 预测模型性能的影响($\varepsilon = 0.01$).....	47
图 5.6 SVR 预测模型的最优参数组	48
图 5.7 不同 ε 水平下 SVR 重构模型预测的均方根误差.....	50
图 5.8 参数 σ 对 SVR 重构模型预测性能的影响($\varepsilon = 0.01$).....	50
图 5.9 参数 C 对 SVR 重构模型预测性能的影响($\varepsilon = 0.01$).....	51
图 5.10 ICA-SVR ² 对 WTI 原油价格的预测结果(单位: 美元/桶).....	52
图 5.11 三种方法的原油价格预测结果对比(单位: 美元/桶).....	53
表 5.1 协方差矩阵	44
表 5.2 协方差矩阵的特征值分解	44
表 5.3 分项 SVR 预测模型核函数和参数	48
表 5.4 分项 SVR 预测模型测试集的预测误差.....	48
表 5.5 不同样本比率下 ICA-SVR ² 对 WTI 原油价格测试集的预测结果	52
表 5.6 三种方法测试集的预测误差比较	53

第一章 绪论

1.1 选题背景及意义

1.1.1 选题背景

作为人类赖以生存和发展的物质基础，原油被称为“工业的血液”，是全球范围内最重要的化工原料和战略资源。原油在工业产业链中的基础地位和世界贸易中的关键地位，决定了其价格波动不仅关系到全球经济领域的发展，甚至影响到各国的国家安全乃至世界稳定。作为反映国际政治经济风云变幻的“晴雨表”，近几十年来，频繁波动的原油价格给全球经济的快速稳定发展带来了严重的阻碍，如何深入理解国际原油市场，正确把握国际原油价格走势日渐成为世界各国政府、石油产业界以及学术界加倍关注的热点问题。

近些年来，我国国民经济呈现出持续快速发展的良好势头，随着工业化和城市化进程的逐渐加快，我国石油消费量增长幅度日趋显著，作为世界石油第二大消费国，不断攀升的对外原油依存度，将我国国民经济的发展与国际原油市场紧密联系在一起^[1]。1993年，我国初次由石油净出口国变为石油净进口国；2008年，中国取代日本成为世界第二大原油进口国；2009年伊始，我国原油对外依存度首次逾越50%的警戒线，跃升至51.3%的水平；2011年，国际能源机构(IEA)在报告中称，到2020年中国对外原油依存度将会上升至65%，届时，中国将超过欧盟成为全球最大的原油进口国。作为世界第二大原油消费国和进口国，我国尚未建立成熟的石油市场和成型的石油定价体制，国内油价的设定很大程度上依赖于国际石油市场；但反过来而言，中国石油市场对于世界石油市场并未发挥相应程度的影响，我国在国际石油市场中极度缺乏石油价格话语权。在这种状况下，国际原油价格频繁而又剧烈的波动无疑给我国国民经济持续健康的发展带来极大的不确定性，很大程度上威胁到国家的安全；同样，面对动荡频繁的国际原油市场，国内大型石油企业及商业机构如何制定合理的经营决策显得极为关键。具体而言，一方面，世界原油价格的持续增长不断拉动商品价格，引发通货膨胀，不利于我国国民经济的稳定增长；同时，作为工业生产中最重要生产原料，国际原油价格的持续上涨致使石油化工等相关行业的经营成本呈现大幅上涨的态势^[2]，巨大的成本压力沿着产业链逐步侵蚀下游行业的利润，最终严重影响到我国国民经济的整体健康发展。

综上所述，世界原油价格的频繁剧烈波动，很大程度上给我国国民经济的持续稳定发展带来重重困难，严重影响到宏观经济的合理调控；同时，不断攀升的石油对外依存度使我国面临着巨大的国际油价波动风险，造成我国石油市场的不稳定，增加石油下游产业的成本并引发更多外汇储备的流失，带来了严峻的能源安全问题。如何在合理认识国际原油市场油价影响因素

的基础上，正确把握世界油价走势已经成为保障我国国民经济高速稳定发展和能源安全的重要课题。

1.1.2 选题意义

国际油价的波动从方方面面影响着全球经济政治的稳定持续发展。目前，相关国际组织、政府部门、石油产业界以及包括高等学校在内的学术界，均致力于合理分析国际原油市场，进而正确把握国际油价的变动方向和波动幅度。国际原油市场价格的分析预测不仅是学术界研究的热点问题，合理把握国际油价对于国家、企业和投资者而言存在至关重要的现实意义。

目前，学术界力求通过模型化市场结构，建立各种数学推理对油价波动的原因和机理给出合理解释，进而预测油价未来的波动方向。本文通过引入独立成分分析作为国际油价时间序列的预处理技术，实现该分解方法与非线性支持向量回归模型的有效集成，不仅丰富了国际原油价格的分析预测方法，提高了传统预测模型的经济解释能力和预测精度，更在一定程度上扩展了“先分解后集成”思想在复杂系统预测问题中的应用，具有重要的理论意义。

深入理解国际原油市场进而准确把握国际油价走势具有一定的现实意义：不仅有助于国家制定合理的能源政策，保障国家经济和能源安全；更有助于指导企业的生产与投资行为，帮助人们合理规划生活。本文在全面分析影响国际油价的各种因素的基础上，合理预测国际原油价格走势，力求准确地洞悉国际原油价格的变动方向和波动幅度，这对于保障我国的经济能源安全，规避企业风险有着颇为重要的现实意义。

1.2 论文的结构安排

1.2.1 研究内容、方法与思路

根据国际原油价格分析预测研究的发展方向，结合本研究课题的需要，本文首先对国际原油市场价格的基本理论进行介绍，总结国际油价波动的基本影响因素，梳理对比各种油价预测方法的优缺点，着重探究国际油价分析预测集成模型的构建：利用独立成分分析(ICA)对国际原油市场价格时间序列进行预处理，结合分解得到的独立成分深入探究国际原油市场价格的影响因素，然后运用支持向量回归模型(SVR)对各个独立成分分别进行时间序列预测，最终构建 SVR 模型重构各独立成分的预测值，得到目标油价序列的最终预测值。选用国际基准原油 WTI 价格作为实证算例，实证分析 ICA-SVR² 集成预测模型的可行性和有效性，挖掘国际原油价格波动的隐藏驱动因素，实现 WTI 原油价格走势的精准预测。

结合设定的研究内容及方法，本文拟定结构化的技术路线见图 1.1。

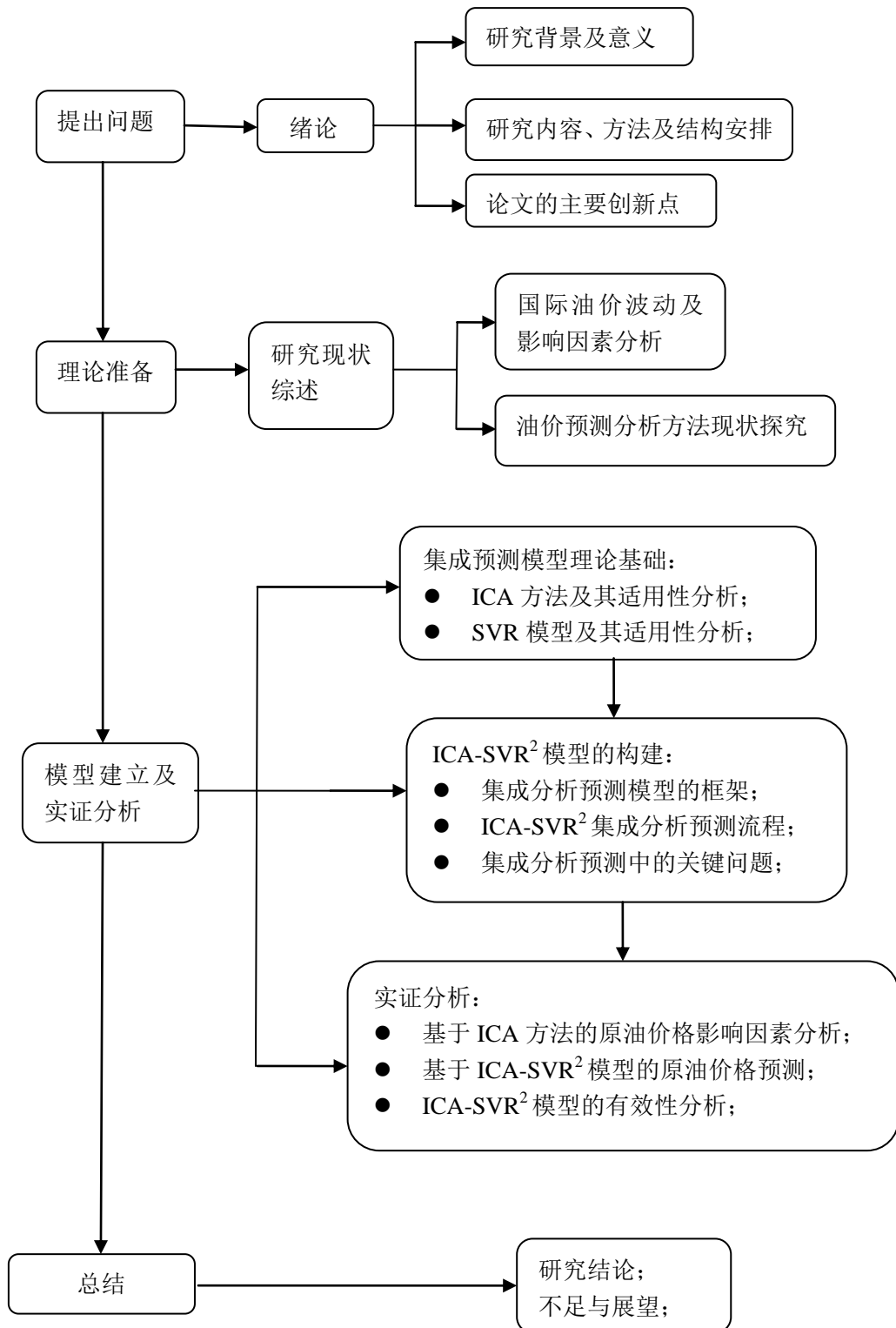


图 1.1 本文的研究技术路线图

1.2.2 论文章节安排

在完成研究内容、方法和技术路线的设定后，进一步确定本文的基本结构：

第一部分 绪论：重点介绍国际原油价格分析与预测研究的背景及意义，进而阐述本文的整体内容、思路和构架组织。

第二部分 国际原油价格分析预测的基本现状：回顾总结国际原油价格波动的基本影响因素，梳理国际原油价格的分析及预测的相关研究成果，并重点关注国际油价预测方法的发展趋势，指明本文“先分解后集成”油价预测模型的研究思路。

第三部分 独立成分分析方法与支持向量回归模型：详细介绍独立成分的定义、原理及其常用的 Fast-ICA 算法；系统阐述了以统计学习理论方法为基础的支持向量回归方法的建模原理、核函数的选择以及参数寻优方法；探究独立成分分析方法和支持向量回归方法在分析预测国际油价中的适用性，为构建 ICA-SVR² 国际油价集成分析预测模型提供坚实的理论基础。

第四部分 集成预测模型的构建：构建基于独立成分分析的国际原油价格预测模型。提出 ICA-SVR² 集成模型的框架，整体把握该集成模型的预测流程，最后具体阐述了模型实施过程中包括独立成分个数的确定、SVR 最优参数的设定等关键问题。

第五部分 实证分析：选取国际基准原油 WTI 现货价格数据作为预测目标，验证本文构建的 ICA-SVR² 集成预测模型的有效性：首先运用独立成分分析方法分解国际油价序列，确定独立成分，具体分析各项独立成分的经济含义；根据不同独立成份的特定波动规律，采用网格搜索法和交叉验证法确定模型最优参数，构建相应的最优支持向量回归模型逐一进行预测；最后对三项独立成分的预测值进行 SVR 重构，取得 WTI 原油价格的最终预测值。改变训练集和测试集的比率，验证 ICA-SVR² 模型的鲁棒性，最后将本文所构建的预测模型与传统预测方法进行分析对比，验证构建模型的可行性和有效性。

第六部分 结论与展望：阐述本文的研究创新点及其研究成果，结合本文研究过程中的不足之处，对未来的研究工作进一步展望。

本文的结构安排存在紧密的内在相关，第一章对国际原油市场做出背景介绍，确定本文的技术路线和结构安排。第二章对国际原油市场价格的影响因素和国内外油价分析预测的研究现状给出具体阐述，探究油价分析预测研究的发展方向，指明了本文构建“先分解后集成”油价预测模型的研究思路。第三章详细介绍了独立成分分析方法和支持向量回归模型，探究了独立成分分析技术作为金融时间序列预处理方法的有效性，阐述了支持向量回归模型在解决复杂系统建模问题中的适用性，为下文构建集成预测模型提供坚实的理论基础。第四章则构建了基于独立成分分析方法的油价集成预测方法，介绍方法的执行框架和流程，并重点阐明如何解决油价集成预测过程中的关键问题。第五章选用 WTI 原油价格作为预测目标，对本文构建的集成预测模型的有效性进行实证分析，结果表明该模型可以有效的解释国际原油市场价格的影响因素，

并最终取得了较好的油价预测效果。在前文研究的基础上，第六章对本研究的成果和不足之处进行了总结和展望。

1.3 论文的主要创新点

作为应用性研究，本文的主要创新点并不在于对特定方法的改进，而在于如何在新的研究思路的指导下，结合新的研究方法分析预测国际原油市场价格，构建创新性的集成预测模型，最终通过实证分析证明所构建模型的有效性。具体而言，本研究的创新点主要包括：

1) 在不同微观层面上分析研究国际油价的波动特征，提取出国际原油价格的基本影响因素，对国际油价的形成机制进行深入分析把握，有效地克服了传统预测模型缺乏经济含义解释的不足之处。

独立成分分析技术可以从多项国际原油市场价格时间序列中提取出国际油价的共同影响因素，将原油价格分解为市场短期波动项、重大事件影响项和趋势项三个部分，找到了国际油价的隐藏驱动机制。与以往国际油价分析预测的文献相比，本文通过对国际油价的预处理，揭示了国际原油市场价格波动的影响因素及各项因素所引起油价波动的特征，得出了更具有经济解释能力的结论。

2) 构建 ICA-SVR² 集成模型来分析预测国际原油价格：首先整体把握集成模型的分析步骤、流程，接着具体分析模型实施过程中的关键问题。该集成模型实现了独立成分分析方法和支持向量回归模型的有效结合，针对独立成分分析分解得到的三列独立成分的数据特点，选用网格搜索法和交叉验证法相结合的方式确定 SVR 模型最优参数——展开预先估计计算，从而获取三项成分的未来估计值；然后对这些估计值建立 SVR 重构模型，以得到原油价格的最终未来预测值。本文构建的集成模型对于原油价格的分析预测具有较强的针对性，增强了预测结果的解释能力，提高了以往预测方法的预测精度。

1.4 本章小结

本研究致力于通过综合集成定性、定量分析和实证研究，合理分析国际原油市场价格波动进而正确把握国际油价的波动趋势。本文的研究意义不仅仅在于研究过程中所获得的实证结论，更为重要的是为准确把握国际原油市场做出自己的尝试。本章结合研究课题的需要和国际原油价格分析预测问题的研究发展方向，确定了本文所研究的主要内容及相应的研究方法，最终确立本研究的技术路线，合理安排整篇论文的总体结构。

第二章 国际原油价格分析预测基本现状

2.1 国际原油价格影响因素分析

作为一种全球性的特殊商品，国际油价的形成不仅依赖于最基本的石油供需关系，也受到包括经济金融因素、突发事件等因素的综合冲击^[3]。国际原油价格的影响因素可以划分为三大类：

1) 供需关系是国际油价走势的基本决定性因素：供给因素主要包括世界石油储量(长期)、世界石油产量(中期)和短期石油库存；世界石油需求与世界石油消费量之间存在相辅相成的关系，世界经济增长和替代性能源均对世界石油需求产生不同程度的影响；

2) 突发事件引发世界原油市场价格中短期的波动：影响油价的突发事件包括自然灾害、地缘政治冲突和一些重大的经济政治事件等。异常天气和地缘政治冲突等突发事件会引发油价出现短期而剧烈的波动，而重大的经济政治事件则在中长期范围内严重影响世界油价的走向，甚至在一定程度上引发国际原油市场的定价机制的改变；

3) 经济金融因素引发国际油价频繁而微小的波动：随着国际原油市场金融属性的日益增强，包括投机炒作和美元汇率等经济金融因素，一方面直接对国际石油供求关系产生短期内的冲击，同时通过间接性改变投资者对一定时间范围内的国际石油供求关系的预期，进而引发油价波动。本文对三类国际原油价格的基本影响因素之间的关系进行简化，构建出国际油价的基本影响因素的关系结构图2.1。

国际原油价格的影响因素

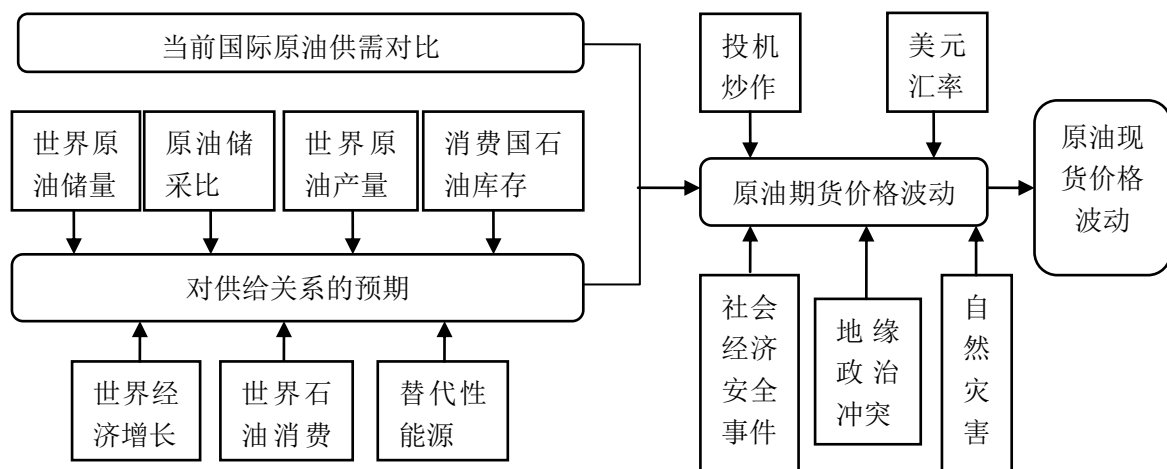


图 2.1 国际原油价格的影响因素关系结构图

2.1.1 供需因素

原油作为全球配置的重要商品，其价格的波动很大程度上服从于供求规律。作为供求关系国际原油价格波动的基本影响因素之一，通常情况下，当石油的供给量无法较好的服从石油需求量时，世界油价会呈现出相应的上涨趋势；反面而言，在国际石油总需求量低于总供给量的情形下，石油价格会显现出下降态势；当石油供给量与需求量的相对关系出现严重失衡时，油价就会暴涨暴跌。

2.1.1.1 需求面分析

石油需求是影响价格的基本因素之一，快速增长的世界石油需求量势必造成石油价格的不断上涨。在保持其它影响因素不变的情况下，国际油价与石油需求量之间呈现正相关关系：石油需求量的增加会拉动油价上涨；而石油需求出现疲软则会导致一定程度的油价下跌。需求因素中，世界石油消费量、国家经济增长、替代性能源等具体因素会改变世界原油供需平衡关系，国际原油投机市场敏锐捕捉到世界石油需求量改变的信号，一系列投资行为引发国际石油市场价格的波动。本文主要分析世界经济发展水平以及各类替代能源的发展对世界石油需求的影响。

a) 世界经济增长与世界石油消费量

作为国民经济的重要能源和基础原材料，世界石油消费量与世界经济发展水平有着密切关系。通常来说，随着世界经济的平稳增长，各行各业纷纷选择扩大生产，原油需求量和消费量均随之上涨，进而拉动原油价格的不断攀升；反之，当世界经济不景气时，各行业纷纷选择相对缩小生产规模，原油市场需求量明显减少，致使国际油价的迅速回落。图 2.2 反映了全球石油消费变化与世界经济增长率之间的正相关关系。

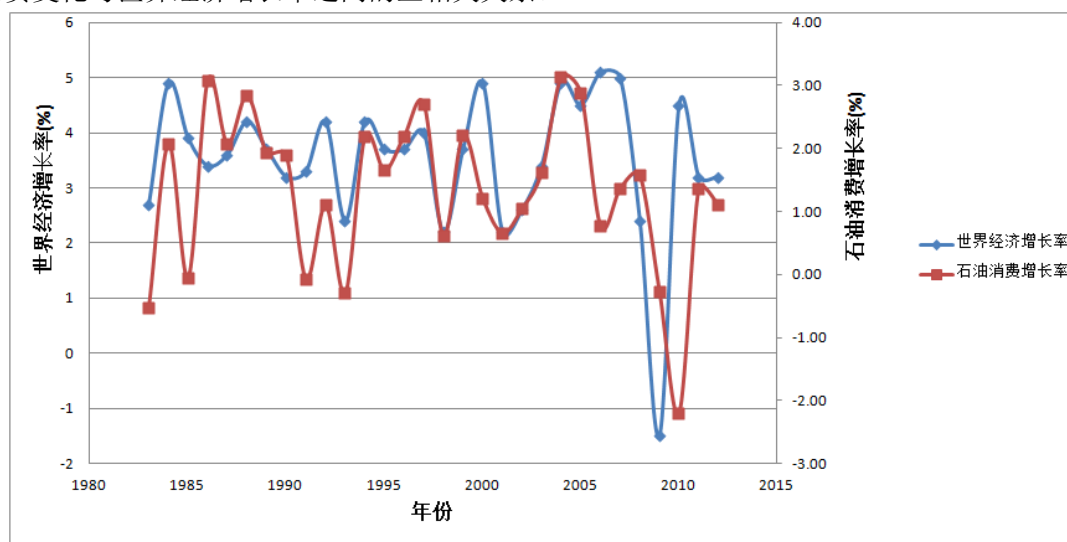


图 2.2 世界经济增长率与全球石油消费增长率(单位：%)

由图 2.2 可知，20 世纪 80 年代以后，国际高油价的冲击引发了人们对于能源效率的关注，

随着世界经济的不断增长，科技的发展以及不断的技术创新提升了整个社会的用油效率，再加上世界经济结构的转型，石油消费量的增长率呈现出远远落后于世界经济增长的趋势。2004 年以后，由于国际油价持续增长，全球石油消费增长率又呈现递减趋势，至 2006 年，全球石油消费增长率降低到 0.78%。从总体而言，全球石油消费量和世界经济发展水平虽然在增长速度上存在一定的差异，但仍呈现出比较明显的同步上升趋势。作为拉动石油需求量最直接的驱动因素，世界经济发展水平通过影响石油消费量而引发了国际油价波动。

b) 替代性能源与世界石油需求

面对不断高涨的国际油价，世界各国加快了新能源开发利用的进程，势必引发世界能源消费结构的逐渐多元化，替代能源的价格及其发展状况将对国际原油价格产生一定程度的抑制作用^[4]。进入 21 世纪，能源逐渐划分为两大类：1)常规能源中的煤炭、天然气、水力水气以及核裂变等；2)非常规能源：包括核能、太阳能风能等。虽然尚未存在石油的一种完全替代型能源，但鉴于不同能源拥有相近的用途，石油面临着煤、天然气、水能、核能等其他能源的替代。

在两种产品 A、B 互为替代品的情况下，A 商品的价格与其替代品 B 商品的需求量变动方向一致，可用需求交叉弹性来表示 A、B 两种商品之间的替代程度：所谓需求交叉弹性系数，即指 A 商品需求量变动的百分比与 B 商品价格变动的百分比之商，公式如下：

$$E_{AB} = \frac{\Delta Q_A / Q_A}{\Delta P_B / P_B} \quad (2-1)$$

其中， ΔQ_A 、 Q_A ---A 商品的需求量变化及初始需求量；

ΔP_B 、 P_B ---B 产品的价格变化及初始价格；

E_{AB} --需求交叉弹性。

需求交叉弹性越大，即指 A、B 商品两者之间拥有越高的替代程度。随着第一次“石油危机”的爆发，石油消费量的减少，17%来自于其他能源类别对石油消费的替代；而第二次“石油危机”的爆发，则引发其他能源对石油的替代量的大幅度上升：已经占所减少的石油消费量的 35%。一般说来，就长期而言，煤炭是石油的主要替代能源。而从短期内看，天然气作为石油的主要替代能源类别，随着其产业的逐渐国际化，未来石油价格很大程度上会收到未来天然气价格的影响。随着时间的推移，石油与其他能源燃料之间的竞争愈发激烈，尤其是发展迅猛的核能：随着核能应用领域的拓宽以及逐渐成熟的技术，作为石油替代性能源的核能势必对石油价格造成愈加显著的影响。

2.1.1.2 供给面分析

一般而言，石油价格与石油供应量之间呈现出负相关关系。在石油供给增加，供大于求的情况下，石油价格会出现下跌；相反，供给趋紧或是预期趋紧的情况往往会引发石油价格上涨。

石油供给面的影响因素可划分为三个方面：a)长期影响因素即世界石油储量，定义长期的供给总量；b)中期的影响因素主要来源于世界石油产量波动；c)短期影响因素，则来自于石油库存的不断调整：针对国际油价的短期范围内的波动，石油库存的适时调整会对油价的短期波动产生进一步的推波助澜。具体来说，石油供给面影响因素主要包括世界石油储量、OPEC 国家石油储量、世界石油产量、OPEC 国家石油产量、非 OPEC 国家石油产量、OPEC 商业库存、美国原油商业库存、储采比^[5]等。

a) 世界原油储量

原油作为一种不可再生资源，其世界范围内的探明储量的变化对原油价格波动产生直接影响。图 2.3 充分呈现了世界石油探明储量近 20 年来持续增加的态势。

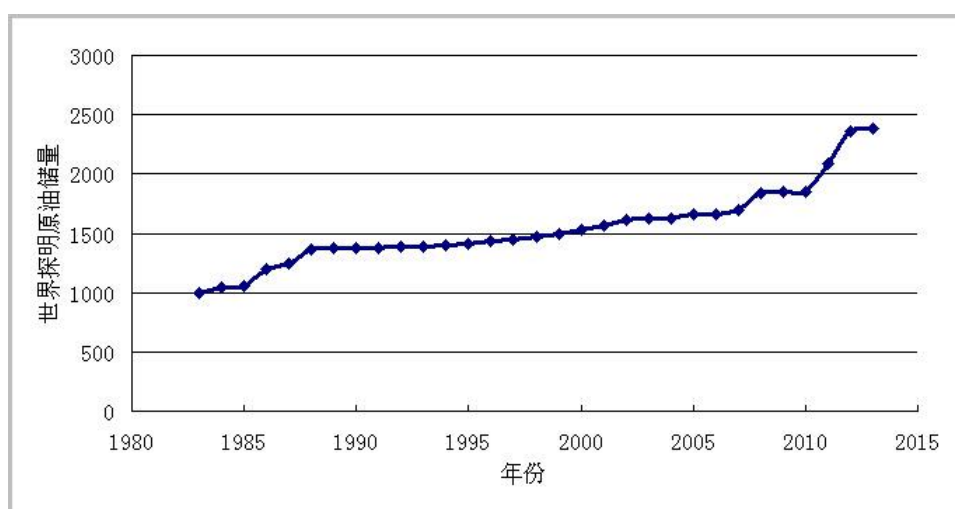


图 2.3 世界探明原油储量(单位：亿万吨)

从图 2.3 可以看出，世界探明原油储量呈稳定上升趋势，这在一定程度上减少了人们对未来石油消耗殆尽的担心，促进了石油消费量的增加。世界石油储量对石油市场发展的阻碍作用逐渐削弱，但石油资源量是否可以满足未来日益增长的石油需求，仍是人类关注的热点问题，因此，可探明石油储量对长期的油价波动仍存在不可忽视的影响。

所谓石油储采比，指的是年末剩余储量除以当年开采量，依照当年生产水平所计算出的石油可供开采的年数。据《2014 年世界能源统计报告》称，至 2013 年底，全球范围内原油探明储量为 2383 亿吨，储采比增至 53.3 年。作为反映原油的稀缺程度的重要指标，原油储采比在一定程度上代表了世界石油产业的可持续发展水平。当石油开采量的增加速度较之石油储量增长速度出现一定程度的缺口时，势必会引发世界原油价格走势的波动。梅晓峰在油价与储采比之间建立回归模型，研究结果证明储采比每增加 1%，真实油价会随之降低 77 美分/桶(1996 年美元币值)。世界石油储采比对国际原油价格产生较为显著的滞后影响^[6]。

b) 世界原油产量

世界原油生产和消费在地域上呈现出极端不平衡的现状，世界石油产量在很大程度上受到全球石油供给结构的影响。总所周知，世界石油产量的波动主要来源于掌握世界丰富石油资源的 OPEC 国家的产量变化。然而，遵循世界上所有事物的发展规律，矛盾或对立的双方，往往在一定条件的趋势下，是可以相互转换的。

OPEC 占有全球大部分乃至 70% 以上的石油储量，尤其是中东国家，不仅占据世界 60% 的原油储量份额，而且集成了世界上最为丰富的高产大油田，降低了石油的生产成本。但是，在原油价格上涨到一定程度的情况下，部分非 OPEC 国家的石油生产亦得到了发展机会，一定程度上对 OPEC 国家的作用产生了制约。相关研究表明：非 OPEC 国家的产能与油价之间并不存在直观的线性关系，总体而言，非 OPEC 国家的石油产能逐步上升，而原油价格则呈现出先降后升的周期性变化，整体上两者呈现略显反向的趋势变化^[7]。近些年来，部分非 OPEC 国家不断增加石油供给量，例如俄罗斯、挪威和墨西哥等国家正逐步改变世界原油供给格局：OPEC 国家和非 OPEC 国家在目标定位和生产策略方面均存在差异，两者的共同相互作用对世界石油供给情况产生重要影响，就长期而言，未来国际石油市场价格的形成将来源于 OPEC 国家和非 OPEC 国家的相互博弈。

c) 世界石油库存

在世界石油市场的大系统中，石油库存在调节总供需水平方面发挥着重要作用，其数量的增减标志着国际石油市场的供需状况：当库存降低时，说明石油的需求高于供给，反之亦然。石油库存从长期和短期两个方面错综复杂作用于国际原油市场，对国际油价的波动产生直接的影响。长远范围内，库存作为供给和需求之间的缓冲剂，可以一定程度上保障油价的稳定：低油价情形下石油库存的增加，会推动油价上涨，而面对高涨的国际油价，抛出石油库存将引发国际原油价格不同程度的下跌；但从短期内考虑，石油库存会对国际原油市场价格的波动起到推波助澜的作用。因为相关机构在原油价格看跌时会选择抛出石油库存，势必加剧油价短期范围内的下跌态势，反之，在油价看涨时购进原油库存则会导致油价的进一步上升。目前，石油库存对国际原油价格产生愈加显著的影响，甚至成为了世界原油价格波动的重要信号。

石油库存由商业库存和国家战略石油储备构成，国家石油战略储备的建立则是为了平抑石油危机和短期石油供应冲击所造成的油价异常波动，保持全球石油生产、加工、供应系统的正常运作。商业库存通常掌握在各大石油公司手中，其会根据特定的市场情况和预期对自己的库存水平进行灵活调整。相关研究表明，原油商业库存对世界油价的影响程度明显高于国家战略储备库存^[7]，该实证研究重点分析了美国原油商业库存对世界原油价格的影响，结果表明，随着美国商业库存变化的增加，WTI 期货价格、WTI 期货价格波动均呈现出增长态势^[8]。

石油供需关系与国际油价的波动之间呈现出一定的相关性。但在实际情况中，国际油价的波动来自于供求关系和非供求类因素的共同作用，油价的升降并非完全服从供需关系。

2.1.2 经济金融因素

进入20世纪90年代,随着世界石油市场金融属性的逐渐攀升,国际原油所呈现的金融属性将逐步取代其商品属性,这意味着国际原油价格不再单纯的仅仅体现实物商品价格,其在很大程度上涵盖了世界石油现货市场以及石油金融市场的双重属性。国际原油的期货价格对现货市场价格的影响日益显著,由国际原油期货市场向原油现货市场传导的价格形成机制逐渐成熟。随着国际原油市场金融属性的日渐显著,期货市场中大量投机炒作行为引发国际石油现货市场价格的频繁波动;而作为世界原油交易的统一货币,美元不同程度的汇率变化势必引发国际原油价格相应幅度的波动。

2.1.2.1 投机炒作

随着经济全球化和国际石油期货市场的日趋完善,石油已经完全融入到世界金融体系中。投机资金在当前国际原油市场中占据举足轻重的地位,约70%的国际石油期货交易属于投机炒作^[8]。面临供需偏紧的世界石油市场和频发的地缘政治冲突等多种不确定因素,国际原油期货市场吸引众多投机者的关注目光。国际金融资本始终在寻找最佳投资机会以在国际石油市场存在高套利机会时进入。大量投机资金的进入将会助涨石油的短期需求,引发原油价格的上升。而面对已经很脆弱的石油供需关系,大量国际投机资本涌入到原油期货市场进行投机性交易,势必会进一步引发国际油价的持续上涨。作为投资基金交易动向的重要指标,原油非商业净多仓量与国际油价之间呈现存在显著的相关性,图2.4直观反映了2001年月至2011年间原油非商业持仓净多头与WTI期货价格之间的关系。

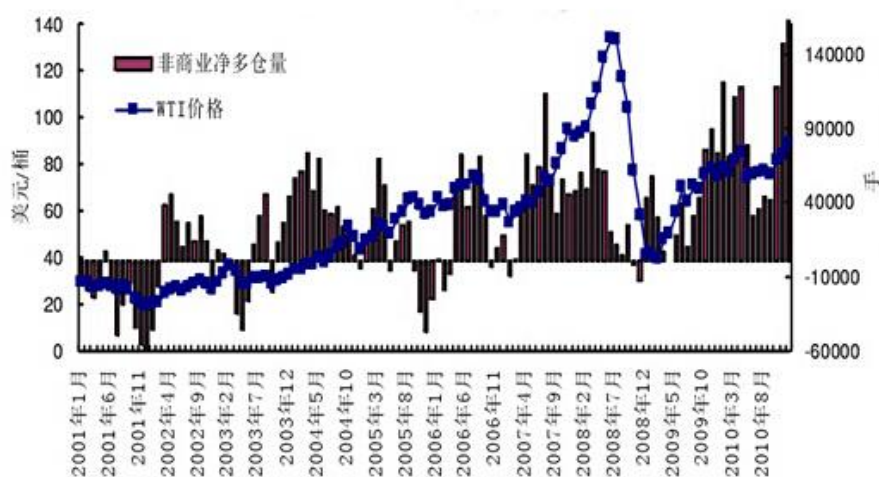


图 2.4 非商业净多头量与 WTI 期货价格

从图 2.4 可以看出,近些年来,投机资金在国际原油期货市场十分活跃,伴随着原油非商

业净多仓量飙升至十几万手，油价也频频上涨，国际油价与原油非商业净多仓量之间密切相关。2003 年随着“美伊战争”的爆发，世界原油市场形式出现了重大变化，原油期货市场的投机活动极度活跃，众多投机商选择大量购进卖出石油期货合约，引发国际国际油价形势看涨看跌的频繁变化，就短期内而言，这类期货投机活动必然对石油期货市场产生助涨助跌的作用，很大程度上放大了国际原油期货市场的价格波动，进而左右着现货市场的走势。但是，石油期货的投机炒作仅仅集中反映了国际石油市场的心理预期，换言之，作为短时期内引发国际原油价格频繁波动的原因之一，投机炒作并非推动国际油价走势波动的根本原因。

2.1.2.2 美元汇率

长期以来，国际市场上的石油贸易通常以世界货币美元来计价和结算，美元汇率的变化必定引发世界原油价格的波动。美元币值的变化程度一定程度上决定了国际油价的升降幅度。本文选取 2001~2014 年的美元指数和 WTI 原油价格数据对两者之间的关系进行探究。



图 2.5 WTI 期货价格与美元指数关系图

不难看出，自 2002~2008 年，美元呈现出不断贬值状态，其他国家在国际石油市场贸易中不仅仅要承担油价的波动，而且还要面临实际支付时美元汇率变动所带来的风险。美元贬值标志着以美元计价的名义油价的提高，为了规避美元贬值风险，投资者纷纷转向石油商品市场，进而增加了额外的石油“需求”，导致原油价格节节攀升；2014 年 7 月份开始，受美国经济向好的推动，美元指数持续走高，这对以美元计价的原油期货带来连续打压，引发国际油价下跌。总体来说，美元指数和油价之间具有显著的负相关关系。通常情况下，美元升值对在一定程度上引发石油价格的下降，而美元贬值则会推高国际油价。

2.1.3 突发事件

作为一种重要的战略物质，石油的价格不仅受市场供需因素的影响，一些国际地缘政治和异常天气、重大事件等突发因素会对石油价格造成短期内的冲击^[9]。突发事件对国际石油市场的影响存在两种表现形式：a) 直接影响石油的实际供给与需求，包括实际生产或者运输等；b) 改变人们对世界石油市场的心理预期，进而引发原油市场价格的波动。引发世界油价波动的突发事件可归为三类：一是地缘政治冲突，包括中东局势的变化以及中东以外的主要生产国家内部的政变等因素，如伊拉克战争、利比亚战事、石油禁运等；二是重大社会经济安全事件，代表性经济安全事件为 2008 年发生的全球金融危机；三是自然灾害，主要包括气象灾害、地质灾害等，如 2004 年 9 月发生的“伊凡”飓风。

进入 21 世纪后，突发事件对国际油价的影响愈加频繁且影响程度日趋增长。影响国际油价的突发事件通常会引发国际油价运行方向的变动，或者助涨已有的油价变动趋势。图 2.6 标注了自 2000 年以来引发国际原油市场价格波动的重要突发事件：2001 年美国发生的“911”恐怖袭击事件，2003 年爆发的“伊拉克战争”以及 2006 年突发的“尼日利亚石油供应袭击”等国际政治军事事件，均通过引发人们心理情绪的波动而造成国际石油市场的恐慌，致使实时国际油价产生剧烈波动；2005 年 8 月“卡特里娜”飓风席卷美国，严重破坏了包括美国南部地区以及墨西哥湾地区在内的炼油公司的石油设施，从而引发了石油市场供给的暴跌，导致 8 月 30 日 WTI 原油价格出现 69.91 美元/桶的历史新高；2008 年上半年，国际油价呈现出急剧暴涨的态势，7 月 WTI 原油价格更是跃至 145 美元/桶的历史至高点。同年 7 月中旬后，在“国际金融危机”的影响下，世界经济出现显著衰退引发世界石油需求出现 55% 的大幅下跌；2009 年伊始，各国纷纷采取“金融救援和财政刺激”计划，实现了全球经济衰退的有效遏制，石油市场需求不断回升，国际原油价格稳步上调；而 2011 年“利比亚战争”的爆发则直接影响到国际原油市场的供给，同时引发了人们对未来国际石油供给的恐慌，导致油价的一路飙升。2014 年 7 月伊始，美国 and OPEC 国家石油产量的持续增长引发了国际原油市场供应的相对过剩，美国原油商业库存的持续攀高更对油价下跌起到推波助澜的作用，国际原油价格出现了近 4 个月的暴跌，至 2014 年 10 月底，WTI 价格由 6 月 30 日的 106.07 美元/桶跌至 81 美元/桶。

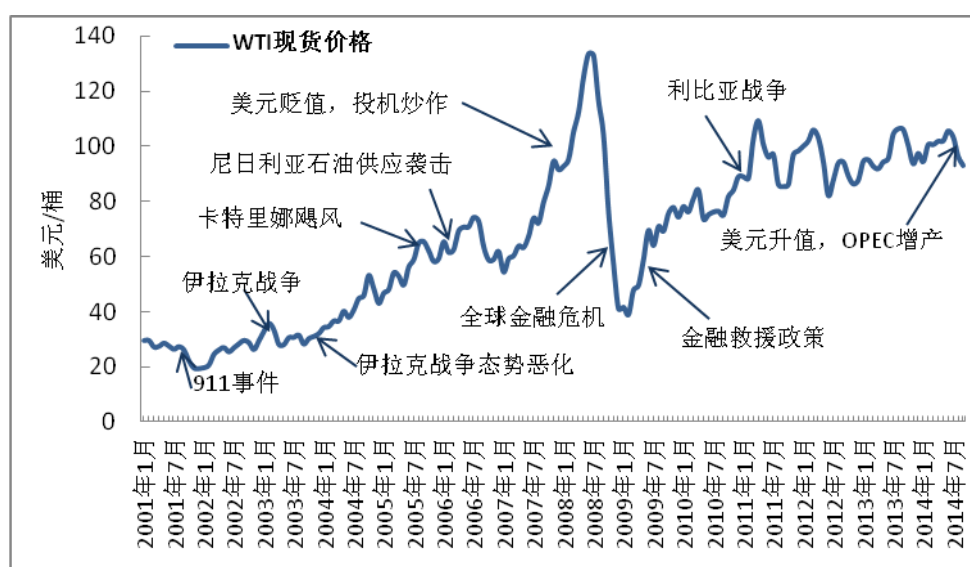


图 2.6 世界原油市场的突发事件

由图 2.6 可以看出，突发事件往往会引发中期或是相对短暂的国际原油价格的剧烈波动。不同类型的突发事件引发国际油价不同持续时间的波动：a) 脉冲式事件:此类事件会导致油价突然的上涨或者下跌，但在事件结束后油价会迅速恢复到之前的水平。这类脉冲式突发事件往往仅对原油价格产生短期的影响，2005 年 8 月“卡特里娜”飓风席卷美国就是典型的脉冲式事件；b) 台阶式事件:国际原油价格在台阶式事件的冲击下发生突变，且长时间内无法恢复到事件以前的水平。台阶式事件使得国际油价在波动过程中形成结构性断点，一定程度上改变了国际原油市场供需力量对比或者定价机制，其对油价产生长期范围内的影响，2003 年的“伊拉克战争”以及 2008 年的“金融危机”是代表性台阶式事件，该类事件导致油价向高位攀升并对油价产生 4~5 年的持续性影响^[9]。

众所周知，准确衡量某种突发事件对国际油价的影响程度是非常困难的，而进一步定量预测这种影响程度更是难上加难。历史环境的不同致使在不同时期下，类似事件对国际原油价格产生不尽相同的影响。但是通过对历史上发生的突发事件进行分析，总结出各类事件所拥有的油价影响模式和影响程度，可以在一定程度上为类似事件再发生时把握油价的波动提供参考信息。

综上所述，国际石油供求关系是决定油价上升或者下降的根本原因，而国际地缘政治，投机资金炒作，美元汇率及其他一些无法度量的微妙的社会、政治等因素的交互作用共同影响并左右着国际油价的波动。同时，国际油价的影响因素之间并不存在严格的界限，石油供需关系、突发事件和经济金融等因素紧密联系，可以相互转化：

1) 石油供需关系是国际油价变化的基本决定因素。需求因素中，世界石油消费量、OECD 国家石油消费量、世界经济水平增长、替代性能源等具体因素会改变世界原油供需平衡关系，向国际原油期货市场传达世界石油需求信号，进一步改变国际油价。类似的，世界石油储量、

全球石油产量、OPEC 石油产量、储采比、供需差额、美国原油商业库存等供给因素会向国际石油期货市场传达石油供给信号，引发国际油价波动。

2) 国际石油期货市场交易的迅猛发展，致使投机因素与美元指数等经济金融因素冲击国际石油的供求关系，使得国际油价走势短期内可以在一定程度上摆脱商品基本供求关系的束缚，致使世界油价产生短期且频繁的波动。

3) 不可预期的突发事件，包括政治因素、战争、突发天气等会影响世界石油供需关系，改变石油投资基金方对供求关系的心理预期，影响世界原油市场，引发世界油价短时期内的剧烈波动；而金融危机这一类重大脉冲式事件则会改变国际油价的定价规则，形成结构式断点，会对国际油价产生相当长一段时间内的影响。

2.2 国际油价预测分析研究现状

国际原油价格波动分析和预测的相关问题，既是研究热点，也是研究难点。各学者从多元化的角度和视野出发，采取各种研究方法和手段对国际油价进行深层次的分析预测研究。目前，国内外原油价格波动分析预测理论及方法具体可以分为三大类：一类是定性分析，强调从不同的理论假设出发，分析国际原油价格的定价机制，并定性预测油价走势，即为石油价格理论模型，包括不可再生资源理论、成本理论以及市场结构模型；第二类是石油价格计量模型，通过选取特定的影响因素来解释国际石油市场价格波动并预测未来油价走势；第三类是以传统时间序列模型以及快速发展的石油价格智能模型为主流的数据驱动型方法。图 2.7 总结概括了当前国际油价分析预测主流模型。

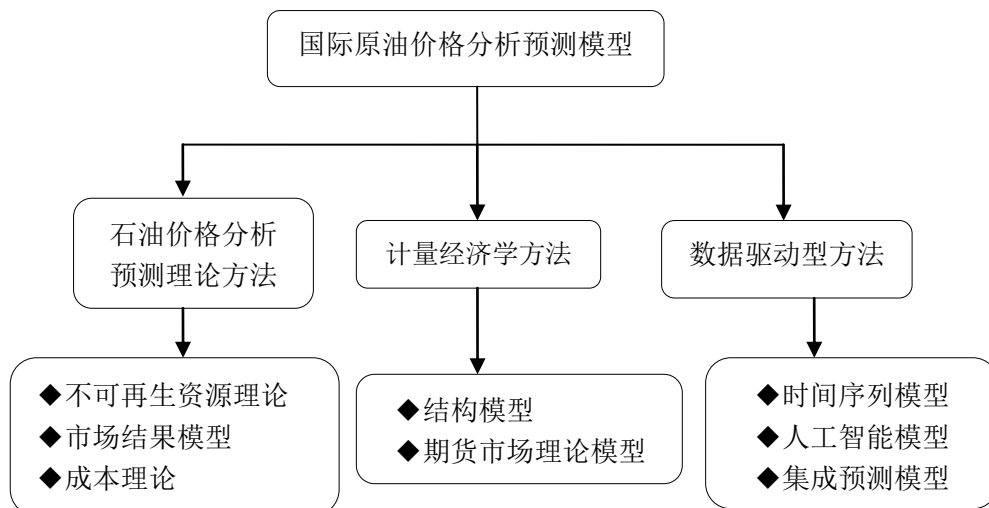


图 2.7 国际油价分析预测模型

2.2.1 石油价格分析预测理论方法

2.2.1.1 不可再生资源理论

Hotelling 于 1931 年^[10]提出了著名的可耗竭资源模型，该模型被视为最早的油价波动分析预测方法。可耗竭资源模型认为资源是一种埋藏在地下的资产，资源所有者试图寻求最优的定价方法以最大化净现值，作为该模型的基本结论，Hotelling 认为资源所有者在面对开采资源资本所获得的收益增长率小于利率的形势时，会选择继续把资源埋藏在地下；而在大于利率的情形下，则选择开采资源，以期获得更高的净现值，亦即资源价格增长率将以等同于市场利率的增长率持续上涨。Hotelling 模型建立在确定的资源量以及不变的开采成本等确定性的假设条件下，而在实际石油产业中，存在各种各样的不确定性，例如新的石油资源的不断发现，各种生产水平提升所带来的不断下降的开采成本，这些因素都增加了石油市场的不确定性。在西方主要国家当期利率的上升时，该模型无法合理对油价并未随着部分而连续上升的实际情况给出合理解释。

沿着 Hotelling 的研究思路，一些学者通过引入石油储量的不确定性假设进一步完善不可再生资源理论模型，力图更为合理的解释国际原油价格的波动原因，进而更为精准的把握石油价格的未来走势。Deshmuka^[11](1980)通过引入资源存量的不确定性，修正了 Hotelling 模型中“资源初始存量是已知且固定的”这一先决条件，指出不同的特定时期拥有不确定的未来储量，在一定程度上解释了国际原油市场中原油价格并未按照西方国家正增长的当期利率的速度上升的现象。类似的，Arrow^[12](1982)指出，原油是随机分布的自然资源，通过进一步引入资源储量的不确定性，对原 Hotelling 模型进行修正，以提高模型的准确性。不可再生资源模型作为研究国际油价的开山之作，为分析国际原油市场建立了坚实的理论基础，但其存在明显的局限性，在定量预测油价中的应用并不常见。

2.2.1.2 市场结构模型

市场结构模型，强调从国际原油市场结构出发，通过探究世界 OPEC 成员国和非 OPEC 成员国以及其他各方国际石油市场参与者之间的动态博弈结构，力图正确把握国际石油市场价格的大致走向。Salant^[13](1976)构建了著名的纳什—古诺模型对世界石油市场的结构进行分析。市场结构模型包括竞争型市场和垄断型市场两大类。两者的共同特点在于均把 OPEC 认定为卡特尔组织。竞争型市场结构模型重点关注国际石油市场的竞争性，强调 OPEC 国家无法控制油价，该模型将市场参与主体当成价格接受者^[14]；反之，以垄断市场结构为基础解释世界油价波动的文献^[15]把 OPEC 及其成员国当作价格制定者，把世界石油市场当做由 OPEC 组织主导定价的垄断市场，强调国际原油价格走势依赖于欧佩克国家的定价原则。针对 OPEC 定价原则的研究，主要包括财富最大化型和生产能力目标型两种。在 Hotelling 模型的基础上，财富最大化型^[16]

重点关注垄断部分，选取一些重要国家作为产量调节者，研究其在不同时期如何选择最优石油产出量以最大化石油收益的净现值；而生产能力利用目标型^[17]则强调垄断者无法获取完全信息，无法确定最优的价格路径，OPEC 国家仅仅可以通过产能利用目标来探索隐藏的最优价格路径。实际的产能利用率较高的情况下，标志着生产处于供求紧张状态，OPEC 国家会选择抬高原油市场价格，反之则降低价格。

市场结构模型处理的是一种无限期的问题，通过设定不同的世界石油市场结构，一定程度解释了各市场参与主体间的动态博弈关系。较之 Hotelling 模型，市场结构模型在分析石油市场的长期趋势具有一定的优势，但其无法合理的解释两期叠代下的变动情况，并且市场结构模型无法对石油价格的未来趋势给出定量分析，对国家以及相关企业合理制定石油战略决策缺乏充分的现实指导意义。

2.2.1.3 成本理论

成本理论从成本角度进行分析，认为生产者通过选择最优开采率实现利润净现值的最大化^[18]。鉴于石油的可耗竭性，其总成本不仅包括开采成本，而且应当将使用者成本(现在一代人的使用造成的后代不能使用的机会成本)考虑在内，因此石油的最优价格应等于边际开采成本与边际使用者成本之和。在完全竞争销售市场下，当边际使用者成本上升使得预期未来资源价格上涨时，又或者是当预期未来技术进步降低边际开采成本时，资源所有者将会减少当期的开采量。而在较高的资本的回报率下，即贴现率出现下降时，当期的资源开采量会随之上升。

面对低迷的原油价格，成本作为主要解释变量可以合理的描述原油市场，但鉴于完全竞争仅仅是石油销售市场的一种高度抽象，在原油高涨的情况下，它很难对过高的油价给出合理的解释，成本理论的预测能力显然是要打些折扣。

2.2.2 石油价格分析预测计量模型

为了更好的分析影响世界原油价格走势的各种因素，正确把握国际石油价格的未来走势，人们从石油市场影响因素、期货市场数据入手，构建相关的计量预测模型，力图对石油价格做出精确的定量分析。常见的国际油价分析预测计量模型包括基于影响因素的结构模型和期货市场理论模型。

2.2.2.1 结构模型

结构模型强调从影响油价的诸多因素中筛选出与油价波动具有密切关系的因素作为自变量，选取原油价格作为因变量，建立二者之间的函数关系进而实现油价未来走势的合理预测。常用的自变量分为三大类：第一类，石油的供给面和需求面，包括世界石油需求，世界原油消费和产量以及各种石油库存水平等；第二类是包括金融危机、政治因素、战争、突发天气等在

内的不确定事件；而第三类则强调世界原油市场的金融属性所引发的油价波动。

相关学者最初通过对原油供需基本面的研究来解释国际原油市场价格的波动，并预测油价的未来走势。Mehrzaad^[19](2004)建立了动态计量模型，以探究 OECD 石油库存，非 OECD 石油需求及 OPEC 国家石油供应三种影响因素对石油现货价格的影响程度。Kaufman^[20](2007)选取 1984-2002 年的季度数据，对 OPEC 政策及执行情况给与重点关注，进而构建了原油价格预测的向量误差修正模型对 OPEC 对国际原油价格的影响给出了定量分析，并对样本外数据给出了动态预测，以期正确把握原油价格的未来走向。各种石油库存水平也是结构模型所选取的常用变量。Ye^[21](2002)选取了从 1992 年 1 月到 2001 年 2 月工业原油相对库存水平和一个代表原油库存相对正常库存水平的偏离值作为主要变量，成功构建了月度原油价格预测模型，该模型在样本内外均表现出较好的预测精度；在 2006 年，他们进一步将相对库存水平细化为两个变量，考虑了当库存低于正常水平和高于正常水平时对油价所产生的非对称影响，分析了从 1991 年海湾战争到 2003 年 10 月的国际石油市场并进行了短期油价预测^[22]。

不可预期的突发事件包括金融危机、战争等重大政治经济事件，也包含着各种地缘政治冲突、突发天气等在内的不规则事件。不可预期的突发事件会影响世界石油供给与需求，改变投资基金方的行为，影响世界原油期货市场进而引发油价波动。张珣^[23](2008)采用 EMD 方法系统的分析了历史上较为重要的海湾战争和伊拉克战争，测定了战争对原油价格的影响模式和影响程度，结果显现了战争进行期间，原油价格对局势的发展非常敏感，波动增大；后通过 EMD 对 2008 年金融危机对原油价格的影响进行测定，分析结果表明，金融危机前期，国际原油价格从平稳波动转为高速增长；在 2008 年 8 月油价达到顶峰后，油价从高增长状态转入急速下降状态；在 2009 年后，回到平稳波动状态^[24]。金融危机这一类极端重大事件不仅仅只是造成国际原油市场短期内的波动，而往往从长期上改变国际原油市场供需力量对比或者定价机制。而对于突发天气、一些地缘政治冲突这类不规则事件往往发生频繁且持续时间比较短，它们对短期原油市场价格波动的影响较大。

作为一种重要的金融衍生工具，国际石油期货的快速发展使得国际油价走势在短期内可以摆脱石油商品基本供需关系的束缚，表现出独有的价格运动规律。不同学者致力于对国际石油市场的金融属性所引发的油价波动进行研究，探究国际油价与投机活动、美元汇率之间的关系。

随着国际原油期货市场的日趋成熟，学术界对投机活动与原油价格之间关系的影响给与了重点关注。Cifarelli 等^[25](2010)采用 ICAPM 模型和 GARCH-M 模型证明了石油市场上投机活动的增加会引起石油价格的频繁波动。Sanders^[26](2004)用 Granger 因果检验法对投机者活动与石油价格之间的关系进行探究，实证结果表明，交易员的净头寸对石油市场价格的波动产生显著的影响。Zhang^[27](2013)研究了投机活动对 WTI 原油期货价格的影响，文章使用了 2007 年到 2010

年的国际石油期货 COT 报告数据,实证结果展示了在样本期间投机交易与石油价格数据之间重要的线性关系以及较弱的非线性关系。其次,美元作为国际原油交易货币,其汇率变化与国际油价波动之间的关系成为众多学者研究的重点。Chaudhuri^[28](1998)采用协整理论和 Granger 因果检验证明了国际原油市场真实价格的不稳定引发了美元汇率的波动。2008 年,张跃军等人通过使用传统的协整理论、VAR 模型和 ARCH 模型以及新型计量经济学模型来检验国际油价与美元汇率之间的溢出效应,实证分析的结果证明了油价与美元汇率之间所存在的因果关系^[29]。

2.2.2.2 期货市场理论模型

国际石油期货市场的迅速发展,使得学术界将研究目光投向期货市场的“价格发现”功能,以期利用世界原油期货价格的指示能力来预测未来的原油现货价格波动。基于期货市场理论的这类模型通常选取期货价格作为自变量,以现货价格为因变量,检验期货市场中形成的价格是否能有效的为现货市场提供参考价格。基于期货市场理论的相关研究并未达成一致结论,既有支持期货市场有效性的,也有质疑其有效性的。Quan^[30](1992)选取原油现货价格作为因变量,引入利率变量,探究原油期货价格和利率对原油期货价格的解释能力。研究表明,利率的解释能力并不显著,原油期货和现货价格之间存在着显著的相关性,但很难辨别两者之间的因果关系。Gulen^[31](1998)指出,对未来短期的油价预测而言,一个月的原油期货价格存在比较明显的指示作用。Salah^[32](2004)进一步分析了不同的原油期货合约对现货价格的不同的预测效果,结果显示 3, 6, 9 月的原油期货价格无法准确的预测未来的现货价格,而 1 月和 12 月的原油期货价格拥有显著的现货价格预测能力。

利用期货市场的价格仍发现功能来预测国际原油现货价格走势,其优点在于所使用数据的易搜集性和操作方便性,短期的期货价格一定程度上能够揭示未来原油现货价格的走势,然而,鉴于国际原油市场的复杂性,单一的使用期货价格对未来现货价格进行预测往往无法给出精确的预测值。

从计量经济学的角度所建立起的石油价格计量预测模型,一定程度上实现了对石油价格的短期精确定量分析;然而,鉴于世界石油市场极易受到各种外生变量的影响,在中长期的预测过程中,不具代表性的初始数据对预测的扰动将会造成最终结果出现很大的误差,从而得到一些缺乏说服力的结果,而如何更好更有效的选择计量方程中的变量,更是成为众多学者不断研究的难题。

2.2.3 数据驱动型模型

2.2.3.1 时间序列理论模型

时间序列理论模型的基本思想是时间序列的过去值已经全面包含了时间序列的结构特征,

因而可以通过分析研究预测目标的过去值找出其变化特征和发展规律，进而外推出预测目标的将来值。国际原油价格本期价格与前期价格之间紧密相关，这使得众多学者采用经济时间序列方法致力于国际原油价格波动及预测。常用的方法包括自回归(Auto Regressive Integrated Moving Average, 简称 ARIMA)系列模型和自回归条件异方差(Auto Regressive Conditional heteroskedasticity model, 简称 ARCH)系列模型。

Box 和 Jenkins 于 1970 年提出 Box-Jenkins 模型，即为 ARIMA(p,d,q)模型。其中，AR 是自回归模型，p 为自回归项，MA 是移动平均模型，q 是移动平均项数；而 d 则指时间序列进行平稳处理所需差分的次数。ARIMA 模型在一定程度上可以反映原油价格变动的特点，国内外众多学者运用此模型把握油价的未来走势。Cabedo^[33](2003)采用自回归移动平均模型 ARMA(Auto Regressive Moving Average)模型，结合 VaR 的历史模型方法预测研究了国际原油价格。ARCH 模型由恩格尔教授于 1982 年提出，可用于描述时间序列的残差序列所存在的异方差性。ARCH 对残差序列中存在的方差进行自相关性研究，适用于对国际油价波动程度的分析预测。Morana^[34](2001)利用 ARCH 衍生模型 GARCH(Generalized Auto Regressive Conditional heteroskedasticity model)半参数估计方法对 Brent 原油价格进行了短期预测，实证研究结果表明该方法在石油价格区间预测方面的适用性。陈益芬^[35](2010)运用 ARCH 类模型，详细分析了 2007 年 6 月 8 日至 2009 年 10 月 8 日间 WTI 原油价格日数据的波动性，实证结果表明样本区间内 WTI 原油价格收益率呈现出显著的 GARCH 效应，国际原油价格波动主要来源于原油期货市场价格以及其他短期因素的影响，并显现出较长时间的持续性。

时间序列模型拥有系统的理论基础和丰富的应用经验，长期广泛的应用显示在油价波动的趋势较明显或者价格波动较小的情况下，时间序列模型可以取得较好的短期油价预测效果。其中，ARIMA 对数据资料的要求较低，仅需收集变量本身的过去值，简化了多种外部因素对预测目标的综合作用，得到了广泛的实际应用。ARCH 类模型一定程度上实现了对油价的波动程度的估计预测，有助于合理的石油市场投资战略的制定，但使用 ARCH 类模型对油价本身进行预测暂未取得明确的成果。但传统的事件序列模型建立在线性假设的基础上，无法较好的拟合国际原油市场的非线性特点。

2.2.3.2 石油价格智能模型

传统的计量经济学模型和时间序列模型建立在数据是线性的假设之上，无法有效的拟合非线性的原油价格序列，通常不能取得精确的原油价格预测值。随着人工智能技术的快速发展，人工神经网络(Artificial Neural Network, 简称 ANN)、遗传规划、目标区域理论、人工神经网络法、小波分析、状态空间模型、支持向量机(Support Vector Machine, 简称 SVM)等非线性智能模型被应用于原油价格预测，有效的拟合了非线性国际油价序列。Bruce^[36](1991)运用信任网络

模型进行原油价格预测,进一步深入的分析了原油价格波动;Kaboudan^[37](2001)以 ANN 模型为基础建立了原油价格预测模型;Mimirani^[38](2005)通过实证研究对比了人工神经网络模型和 VAR 模型对原油价格预测的预测效果;此外,Radchenko^[39](2005)构建了长期原油市场价格的状态空间模型;Xie^[40](2006)利用支持向量机对原油价格进行预测;Fan^[41](2008)构建了模式匹配技术模型来预测原油价格;新的人工智能分析技术的引进,指明了世界油价预测的新方向,非线性方法的引入可以更好的处理原油价格的非线性特征,取得较之传统的计量经济学模型和时间序列模型更为精确的预测值。但是单一的人工智能模型同样存在一些问题,比如,ANN 在运行的过程中经常会出现局部最小值以及过度拟合现象。单一模型不足以刻画复杂的国际原油价格市场价格时间序列的动态变化。

2.2.3.3 油价集成预测模型

油价集成预测模型将两个或两个以上不同的预测方法应用于原油价格的预测,在充分发挥每个模型各自预测优势的基础上,较好的克服各自的缺陷,进而实现预测效果的有效提升。近些年来,集成模型在国际油价预测应用中得到快速发展。汪寿阳^[42](2005)针对复杂系统提出的 TEI@I 方法论,开辟了集成模型在油价方面的应用。通过系统地融合文本挖掘技术、计量经济模型以及智能技术构建出 TEI@I 方法论,并将这一方法应用到复杂国际原油市场中,最终取得了较好的油价预测结果。从组合预测的角度出发,李彬^[43](2006)构建了一种基于模糊神经网络的原油市场价格非线性集成预测的新方法,实现了 RBF 神经网络模型、基于马尔可夫链的半参数模型以及基于小波分解的模型的有效组合,并将该组合预测模型应用于 Brent 原油价格的预测,实证研究结果表明,该组合预测模型具有预测精度高、学习与泛化能力较强、适应性良好的特点。吴虹^[44](2010)通过实现 ARIMA 和 SVM 的有效组合,构建了取得高精度预测结果的石油价格预测模型。在“先分解后集成”思想的指导下,近年来,基于分解的原油价格预测模型成为学术界的研究热点,小波分解以及经验模态分解算法等分解方法的引入大大提升了油价预测精度。梁强等^[45](2005)构建了以小波分析为基础的原油价格长期预测模型,构建了小波分析与人工神经网络的集成长期趋势预测模型,与传统的 ARIMA, GRACH 方法相比,该组合模型的预测精度呈现出显著的优势。Yu^[46]基于经验模态分解算法(EMD)分解国际油价序列;并选用人工神经网络模型分别预测各项分解序列,最终构建出高精度的集成油价预测模型。集成预测思想在国际原油价格分析预测方面取得较大突破,并取得了较好的预测效果。

2.3 本章小结

本章将国际原油价格影响因素分为供需关系、突发事件和经济金融因素三个方面,总结梳理已有的国际原油价格分析预测的相关文献,图 2.8 对原油价格预测方法的发展现状和趋势进行归纳:

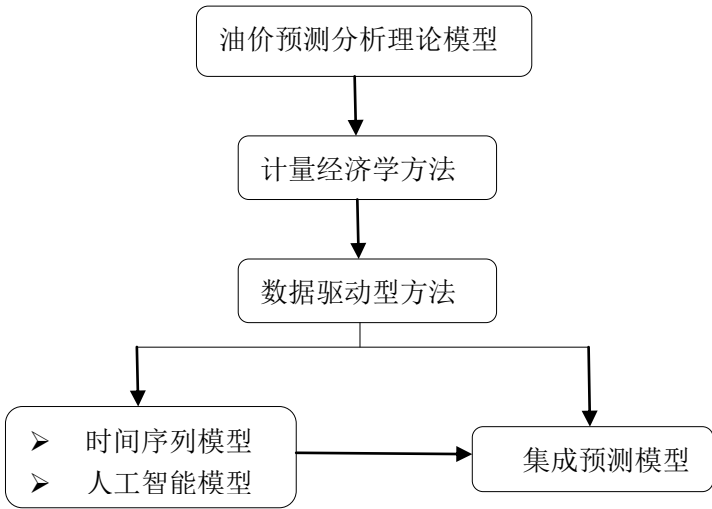


图 2.8 国际油价分析预测模型发展趋势

1) 国际油价预测方法存在着从理论模型向线性计量模型，进一步向非线性人工智能模型的发展趋势；而集成预测模型在预测效果上所拥有的明显优势，使其逐渐成为当前国际原油价格分析与预测研究的主流。

2) 计量经济方法和人工智能模型等数据驱动模型可以取得较好的预测效果，但是其无法揭示国际油价波动的影响因素，预测效果缺乏合理的经济解释，这个关键问题未得到充分解决。

3) 小波分解、经验模态分解方法在国际油价分析中的成功应用，证明了分解类方法在油价预测中的可操作性；通过合理的分解方法将油价时间序列进行分解，进而对分解后的单一序列进行分别预测，最后对预测结果进行重构，这种被称为“先分解后集成”的思想正逐步被引入国际油价波动分析及预测研究中。

第三章 独立成分分析方法与支持向量回归模型

通过第三章对现有国际油价分析预测研究基础的分析对比,本文在“先分解后集成”这一思想的指导下,试图建立国际原油价格的集成预测模型:首次尝试将独立成分分析作为国际油价数据的预处理技术,试图对分解确定的独立成分给出合理的经济解释;而对于预测方法以及集成方法的确定,本文选取非线性支持向量回归模型对分解出的独立成分进行预测,最终对各项预测值重构给出最终预测值。本章重点介绍了独立成分分析和支持向量回归模型的基本原理和算法实现,对两种方法在解释国际原油价格波动机制以及实现油价准确预测中的可行性给出具体阐述,为构建国际原油价格集成分析预测模型提供坚实的理论基础。

3.1 独立成分分析(ICA)

设想一种场景,在同一个房间里的不同位置安置着两个麦克风,两个人在同一时刻说话。两个麦克风可以同时记录下两个时间信号,我们如何仅仅通过这两个记录的混合信号分离出两个原始的语音信号?这就是有名的“鸡尾酒会”问题。在实际生活中,很大一部分被观察到的信号均由各个源信号以及包括噪声在内的其它信号混合而成,面对这类问题,如何根据混合信号提取出源信号显得尤为重要。

最初提出独立成分分析(Independent Component Analysis,简称 ICA),是为了分析“鸡尾酒会”问题:在源信号和传输通道参数未知的情况下,我们仅需做出两个人的声音信号是统计独立的这一相当合理的假设,ICA 就可以仅由观测到的信号恢复出源信号的各个独立分量^[47]。独立成分分析以高阶统计特性为基础,它的目的是把观察的数据进行某种分解,以分离得到尽可能相互独立的成分(记为独立成份, ICs)。ICA 在实际问题中的应用主要包括提取数据特征^[48]、实现盲源信号分离^[49]、研究生理学数据^[50]、处理语音信号^[51]、人脸识别^[52]、及图像处理^[53]等。

主成分分析(Principal components analysis, 简称 PCA)是基于信号二阶统计特性将多个指标通过正交投影转化为少数几个互不相关的指标的一种统计方法,以去除数据各分量之间的相关性为最终目的,因而在实际中,我们通常将其应用于压缩图像数据,作为主成分分析方法的基本限制,其仅仅可以分离出互不相关的信号源;而作为基于信号高阶统计特性的独立成分分析方法,其可以分解出互相独立的原始信号分量。鉴于这一显著优势,ICA 被广泛应用于各种学科领域。

3.1.1 独立成分分析定义

Jutten 和Herault^[54]最初提出独立成分分析这一概念时,将其简单的定义为从线性混合信号

里分离出基本源信号的工具。随着1994年严格的ICA理论框架^[55]的提出，ICA在理论领域和实用领域均得到广泛的关注。独立成分分析方法所解决的问题就是如何仅由观测到的数据来估计其是由哪些独立成分混合而成以及混合的方式，见图3.1：

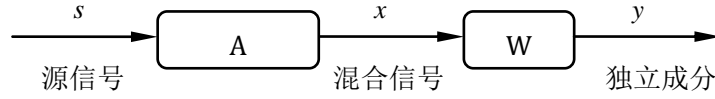


图 3.1 独立成分分析方法

ICA 的基本模型定义如下：

$$X = AS \quad (3-1)$$

其中， $X = (x_1, x_2, \dots, x_n)^T$ 为观测到的 $n \times m$ 维混合信号， $S = (s_1, s_2, \dots, s_n)$ 为相互独立的 $n \times m$ 维未知源信号，即为独立成分，为无法直接观察到的隐藏变量；矩阵 A 为 $(n \times n)$ 维未知的混合矩阵。ICA 的目的就是确定分离矩阵 $W(A^{-1})$ ，得到尽量独立的 y ：

$$\begin{bmatrix} y_{11}, y_{12}, \dots, y_{1m} \\ y_{21}, y_{22}, \dots, y_{2m} \\ \dots \\ y_{n1}, y_{n2}, \dots, y_{nm} \end{bmatrix} = \begin{bmatrix} w_{11}, w_{12}, \dots, w_{1n} \\ w_{21}, w_{22}, \dots, w_{2n} \\ \dots \\ w_{n1}, w_{n2}, \dots, w_{nn} \end{bmatrix} \times \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1m} \\ x_{21}, x_{22}, \dots, x_{2m} \\ \dots \\ x_{n1}, x_{n2}, \dots, x_{nm} \end{bmatrix} \quad (3-2)$$

矩阵形式如下：

$$y = WX \quad (3-3)$$

综上所述，独立成分分析方法的核心思想就是在统计独立的假设前提下，根据观察到的信号 X ，出发寻找分离矩阵 W ，然后对 X 进行线性变换，得到输出向量 y 。为了找到这样一个矩阵 W ，我们需做出以下假设：

- 1) 独立成分被假定是统计独立的；
- 2) 最多只有一个 ICs 满足高斯分布；
- 3) 假定未知的混合矩阵是方阵。

以上所定义的 ICA 模型存在一些必然的含混因素或者不确定性：a) 我们无法非常精确地估计出 s ；b) 我们无法确定 ICs 的顺序；c) 我们无法确定 ICs 的方差。不过幸运的是，这些含混和不确定因素在绝大多数中的应用是无关紧要的。在以上对 ICA 模型定义的过程中，出于简化问题的目的，我们假设独立成分和观测混合信号的数目相等。但在实际应用中，两者之间的数目可以不等，一般情况下独立成分的个数小于观测到混合量的个数。

3.1.2 独立成分估计原理

独立成分分析的目标是找到合适的线性变换以求解得到一组相互独立的独立成分，基于不同的信号之间独立性的度量标准，目前独立成分分析方法存在三种主要的估计原理：最大化非高斯性、最小化互信息、最大似然函数估计方法。

3.1.2.1 最大化非高斯性的 ICA 估计方法

中心极限定理是概率论中的一个经典结论，该定理表明在一定条件下，独立随机变量的和的分布趋于高斯分布，独立随机变量之和形成的分布比各个独立随机变量更加接近于高斯分布。由中心极限定理我们可以看出，非高斯性就意味着独立，因此，可以通过度量非高斯性来寻求独立成分。考虑独立成分分析问题：

$$y = W^T x = \sum_i w_i x_i \quad (3-4)$$

其中， w 是待定的向量。如果 w 为 A^- 中的一行， y 就等于一个独立成分。那么如何利用中心极限定理来寻求 w ？定义 $z = A^T W$ ，则式(3)变换如下：

$$y = W^T x = W^T A s = z^T s \quad (3-5)$$

从上式可知， y 作为 s_i 的一个线性组合，权重由 z_i 给出。根据中心极限定理可知，较之所有的 s_i ， y 更接近高斯分布，因而我们可将 w 看作是最大化非高斯 $W^T x$ 的一个向量，而对应于 z 的这样一个向量， $y = W^T x = z^T s$ 则等于其中的一个独立成分。

总结而言，独立成分的确定可通过最大化 $W^T x$ 的非高斯性来实现。常用的非高斯性度量指标有峭度(kurtosis)和负熵两种。峭度的计算非常简单，但在实际计算中，其受外层的影响较大，鲁棒性较弱。而作为非高斯性度量标准的负熵，拥有较强的鲁棒性，但计算过程相对复杂。因此，在实际实用中，通常采用负熵的近似来作为衡量非高斯性的标准。

(1) 非高斯性度量—峭度(Kurtosis)

作为度量非高斯性的一种经典方法，峭度(kurtosis)又被称为随机变量的四阶累积量。 y 的峭度被定义为：

$$kurt(y) = E(y^4) - 3[E(y^2)]^2 \quad (3-6)$$

在实际中，我们假设 y 是单位方差，则可将式子简化为：

$$kurt(y) = E(y^4) - 3 \quad (3-7)$$

高斯随机变量的峭度等于零，相反，对于绝大部分非高斯随机变量而言，其峭度不为零。超高斯随机变量拥有正的峭度，低高斯随机变量的峭度为负值。非高斯性越强的随机变量，其峭度的绝对值越大。在独立成分分析及其相关领域中，可采用峭度的绝对值或是峭度的平方度

量非高斯性, 取值为0的是高斯变量, 取值大于零的则是非高斯变量。

(2) 非高斯性度量—负熵(Negentropy)

基于负熵的非高斯性度量, 其理论基础为信息论中熵的概念。作为反映随机变量信息度的指标, 变量越随机, 熵值越大--在拥有相等方差的众多随机变量当中, 高斯变量具有最大的熵。这一特点标志着我们可以使用熵来度量随机变量的非高斯性, 随机变量的分布越集中于某一特定值, 其熵越小。随机变量的熵的定义为:

$$H(y) = -\int p_y(\eta) \log p_y(\eta) d\eta \quad (3-8)$$

为构建合理的非高斯性度量, 使其为非负的量值, 且对高斯变量其取值为零。可对熵进行标准化修改, 定义为负熵 J , 如下:

$$J(y) = H(y_{gauss}) - H(y) \quad (3-9)$$

这里 y_{gauss} 是与 y 具有相同协方差矩阵的高斯随机向量。负熵的特性在于其为非负值, 当且仅当 y 是高斯分布时其值为0。公式(5)、(6)表明, 通过定义计算负熵, 必须获取随机变量概率分布的先知信息, 而在实际中我们往往无法准确获知这一信息, 负熵的计算显得尤为困难。因此, 我们通常选用负熵的近似来度量随机变量的非高斯性。估计负熵近似值的一种常用经典方法如下:

$$J(y) \approx [E\{G(y)\} - E\{G(v)\}]^2 \quad (3-10)$$

v 满足标准正态分布。函数 G 是任意的实际非二次函数, 其一般选择的两种方式如下

$$G_1(y) = \frac{1}{a_1} \log \cosh a_1 y \quad (3-11)$$

$$G_2(y) = -\exp(-y^2/2) \quad (3-12)$$

其中的常数 a_1 取在 $[1, 2]$ 内比较合适, 通常取其为1。

这样近似得到的负熵, 实现了峭度和负熵的两种经典非高斯性度量方法的完美折衷。其概念简单、计算量小, 且具有包括鲁棒性在内的众多良好的统计特性。因此, ICA 方法通常使用负熵的近似这种更加客观的函数作为随机变量的非高斯性的度量。

3.1.2.2 极小化互信息估计方法

作为度量随机变量独立性的另一个重要的指标, 互信息总是非负的, 仅仅在随机变量是统计独立的情况下其才取值为零。在进行独立成分分析时, 我们可将最小化互信息作为寻求独立成分的准则。从熵的概念出发, 可将 m 个随机变量的互信息定义如下:

$$I(y_1, y_2, \dots, y_m) = \sum_{i=1}^m H(y_i) - H(y) \quad (3-13)$$

我们可以将互信息解释为熵的代码长度, 如果限制 y_i 为不相关且拥有单位方差的随机变

量，且存在可逆的线性变换 $\mathbf{y}=\mathbf{W}\mathbf{x}$ ，则互信息与负熵之间存在如式所示的关系：

$$I(y_1, y_2, \dots, y_m) = C - \sum_{i=1}^m J(y_i) \quad (3-14)$$

式(3-14)中， C 为不依赖于 \mathbf{W} 的常数。我们可以看出，在估计成分互不相关的情况下，以极小化互信息来进行 ICA 模型估计这一基本思想等价于最大化所估计独立成分的非高斯性之和。利用最小化互信息来表述 ICA 问题时，可以严格的解释非高斯性启发式原思路。但是在实际应用中，互信息无法与负熵一样实现一个接一个地估计出独立成分。

3.1.2.3 最大似然估计方法

极大似然估计是估计 ICA 中非常普遍的使用方法，它与最小化互信息方法之间存在紧密的联系。最大似然估计法的基本思想在于通过寻求模型中的参数力图实现观测样本出现概率的最大化。在无噪声的独立成分分析方法中，首先对似然函数给出定义，进而采用最大似然估计法来确定最终的 ICA 模型。在实际运用中，可以采用似然度的对数形式以简化计算过程。假设我们有 T 个 \mathbf{X} 的观测，表示为 $\mathbf{x}(t)(t=1, 2, \dots, T)$ ，则对数似然度可定义为 \mathbf{W} 的函数，如下式所示：

$$\log L(\mathbf{W}) = \sum_{t=1}^T \sum_{i=1}^n \log f_i(w_i^T \mathbf{x}(t)) + T \log |\det \mathbf{W}| \quad (3-15)$$

式中，定义 \mathbf{W} 为 \mathbf{A} 的逆矩阵， f_i 代表的是假设已知的密度函数。

极大似然函数估计的原理是指，若求出一个 \mathbf{W} 可以最大化对 \mathbf{X} 的平均值，则该 \mathbf{W} 就是所需的解。因此，在 ICA 模型中，我们可以将下式定义为自变量为 \mathbf{W} 的目标函数：

$$\frac{1}{T} \log L(\mathbf{W}) = E \left\{ \sum_{i=1}^n \log f_i(w_i^T \mathbf{x}) \right\} + \log \det |\mathbf{W}| \quad (3-16)$$

若存在 \mathbf{W} 使得 $L(\mathbf{W})$ ，就可以确定 ICA 的独立成分。

3.1.3 Fast-ICA 算法实现

目前学术界常用的 ICA 算法包括两种代表性算法：第一类算法是基于某些最大或最小化的相关准则函数，优点在于其适用于任何分布的独立成分，缺点在于其复杂繁琐的计算量；第二类算法则以随机梯度方法为基础，作为一种自适应算法，该类算法存在显著的优点，其能保证收敛到一个相应的解，而其主要劣势在于较慢的收敛速度以及对学习速率参数的正确选择的高度依赖。

近年来使用广泛的快速 ICA 算法(Fast-ICA)以非高斯性最大化原理为基础，采用固定点(Fixed-point)迭代理论以确定非高斯性最大值。Fast-ICA 算法适用于不同类型的数据，基于牛顿迭代算法批量处理测量变量 \mathbf{X} 的大量采样点，每次从观测混合信号中分离出一个独立分量，该算法实现了独立成分分析快速在分析高维数据方面的有效应用。故本文采用 Fast-ICA 算法实现独立成分的逐一提取。

基于负熵的 Fast-ICA 算法是以最大化负熵作为搜寻方向，逐步提取出 x 的独立成分 y 。为了简化问题，在进行数据分析之前，首先要对数据进行中心化，使得其均值为 0，然后对中心化后的数据进行白化处理，得到 $Z=VX$ ，其中 $V=2(C_X)^{-1/2}$ ， $C=E(XX^T)$ 。

设 w 为单位向量，即 $\|w\|=1$ ，为了寻求 y ，需最大化负熵近似值：

$$J(y) \approx [E\{G(w^T z)\} - E\{G(v)\}]^2 \quad (3-17)$$

注意到， $J(y)$ 近似值的最大值通常在 $E\{G(w^T z)\}$ 的极值点处取得最大。根据拉格朗日条件，在 $\|w\|=1$ 的约束条件下的极值需要满足以下公式：

$$E\{z \cdot g(w^T z)\} + \beta w = 0 \quad (3-18)$$

式中， g 为 G 的导数。采用牛顿迭代法来求解此方程。首先，用 F 表示式中左侧的部分， $JF(w)$ 为 F 的 Jacobi 矩阵 (即拉格朗日乘子式的二阶导数)，知：

$$JF(w) = E\{zz^T g'(w^T z)\} + \beta I \quad (3-19)$$

因为 z 已经是原始数据球面化得到的，可以给出的 $E\{zz^T g'(w^T z)\}$ 合理近似 $E\{zz^T\} E\{g'(w^T z)\} = E\{g'(w^T z)\} I$ ，以简化矩阵求逆过程。于是我们可以得到 w 的近似牛顿迭代值，如下：

$$w \leftarrow w - [E\{zg(w^T z)\} + \beta w / [E\{g'(w^T z)\} + \beta]]$$

为了实现进一步简化，可以在上式两边同乘以 $E\{g'(w^T z)\} + \beta$ ，经过代数化简，可以得到：

$$w \leftarrow [E\{zg(w^T z)\} - E\{g'(w^T z)\} w]$$

公式就是对应的快速不动点算法中不动点迭代的基本公式。我们可以将 Fast-ICA 算法寻找一个极大化非高斯方向的步骤总结如下：

- 1) 中心化原始数据 x ，使其均值为 0；
- 2) 对中心化后的数据进行白化处理，得到 z
- 3) 随意选取一个初始化向量 w ， w 需具有单位范数，即满足 $\|w\|=1$ ；
- 4) 更新 $w \leftarrow [E\{zg(w^T z)\} - E\{g'(w^T z)\} w]$ ；
- 5) 标准化 w ， $w \leftarrow w / \|w\|$ ；
- 6) 如果 w 尚未收敛，则返回步骤(4)；

在已经求出了 m 个独立成分 w_1, w_2, \dots, w_m ，求第 $m+1$ 个独立成分的情况下，首先随意选定一个初始向量 w_m ：

第一步，正交化处理：

令 $w_{m+1} = w_{m+1} - \sum_{j=1}^m (w_{m+1}^T w_j) w_j$ ，则此时的 w_{m+1} 与 w_1, w_2, \dots, w_m 正交

第二步，单位化处理：

令 $w_{m+1} = w_{m+1} / \|w_{m+1}\|$ ，接着使用步骤4与5对 w_{m+1} 进行迭代。需要注意的是，在每一次迭代之前，我们都要首先对 w_{m+1} 和 w_1, w_2, \dots, w_m 进行正交化与单位化处理。

与其他独立成分分析算法相比，Fast-ICA算法具有许多特性优势，如下：

- 收敛速度快，且有一定的精度保证：与普通的ICA算法的线性相比，Fast-ICA有效提高了收敛速度，是传统ICA算法收敛速度的 2至3倍。
- 可直接找到任何非高斯分布的独立成分：无需进行概率密度分布函数的估计，仅仅通过使用一个非线性函数 g ，Fast-ICA 算法即可找到非高斯的独立成分。

3.1.4 独立成分分析方法的适用性

近年来随着对 ICA 研究关注的增加，众多学者尝试将 ICA 应用到金融时间序列数据中，以寻找隐藏在复杂现象背后的深层次的波动机制。许多情形下，我们可以搜集到并行的金融时间序列数据，例如流通交易率或每日的股票成交量等，这些数据里可能包含并行的潜在影响因素。ICA 实现了隐藏驱动机制的有效挖掘，寻找金融数据中潜在的内在影响因素。对独立成分分析在分析金融问题的应用研究，经历了从单纯挖掘影响因素到与其他模型集成进一步实现预测的飞跃。在这里，本文对 ICA 在金融时间序列的预测问题中的应用做简单综述。

作为寻找时间序列背后真正驱动因素的有效方法，学者最初尝试采用 ICA 来揭示金融问题的驱动机制。Oja^[56](1998)采用 ICA 对同一连锁店中若干分店的现金流量进行分析，分离出影响商店现金流量的一些共同的基本因素，同时给出不同影响因素的权重，由此可以确定这些因素对任意一个特定商店的影响，即分析出不同的管理行为在不同个体店的具体环境下带来的具体后果。

独立成分分析变换所产生的成分信号相对观测到的混合信号更加结构化和规则化，这就促使学者尝试通过将 ICA 与其他模型结合，进一步实现时间序列的预测。Malaroiu^[57](2000)构建 ICA-AR 算法并运用于 10 组外汇兑换率的时间序列上。结果显示，较之经典 AR 模型，该集成算法的预测结果更为精确。由于金融时间序列具有非线性、高度不规则性的特点，学者引入非线性技术，将线性独立成分分析与一些非线性预测方法混合，Lu^[58](2009)通过集成 ICA 和 SVR 模型，构建了一个两阶段的预测模型处理金融数据，首先抽取出不含噪声的独立成分，后将得到的独立成分作为 SVR 模型的输入变量，给出高精度的预测结果。

独立成分分析在金融时间序列方面的成功应用，显示了其在揭示金融序列的内在推动机制方面的可操作性，很好的克服传统时间序列模型缺乏合理的经济解释这一缺陷。在相关文献研究的基础上，本文以“先分解后集成”作为指导思想，首次尝试选用独立成分分析作为国际油价数据的预处理技术，以探究国际原油市场价格波动的影响因素，并通过与其他模型的有效集

成实现国际油价的准确预测。

3.2 支持向量回归模型 (SVR)

Vapnik^[59]于 1995 年提出新兴机器学习技术, 支持向量机(support vector machine, 简称 SVM), 该方法以统计学习理论为基础, 实现了最大间隔超平面、凸二次规划、Mercer 核、稀疏解和松弛变量等相关技术的成功结合。

支持向量机主要应用于解决高度非线性分类和回归等问题, 实现了问题由低维空间向高维空间的有效推广, 进而解决了维数过高的问题。在低维空间中, 遇到线形不可分问题的情况时, 以内积函数(核函数)思想为理论基础, 实现样本从原低维空间至高维特征空间的映射, 将原问题转化成一个凸二次优化问题, 然后在所选择的高维特征空间中进行分类或是回归训练。

3.2.1 SVR 模型原理

支持向量回归机(Support Vector Regression SVR)处理预测问题的基本思想在于如何合理的构造非线性的映射, 将原样本数据由原空间映射到高维空间甚至无穷维的特征空间(Hilbert 空间), 最终在得到的高维空间中应用线性学习机的方法解决样本空间中的高度非线性的实际回归问题^[60]。

SVR 应用于非线性回归问题时, 其学习目标是实现一个凸函数的最小化, 其结构示意图见下图 3.2。

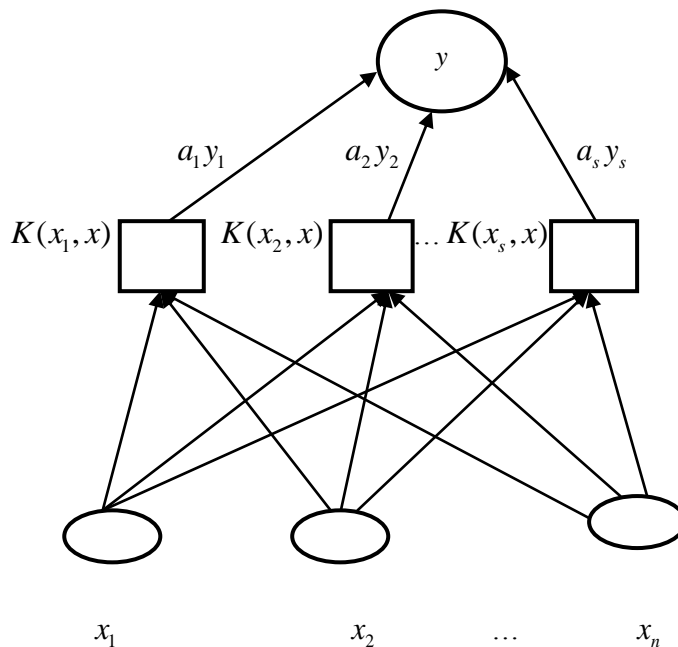


图 3.2 支持向量结构示意图

其中, 给定 N 个训练样本数据 $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_N, y_N)$, 其中

$x_i \in R_n, y_i \in R, x_i$ 为 n 维输入向量, 中间层 $K(x_i, x)$ 是 S 个支持向量所做的非线性变换, 即为内积, y_i 为 x_i 所对应的输出值, $y = \sum_{i=1}^N \alpha_i K(x_i, x) + b$ 是目标输出函数。支持向量回归模型即是对给定的训练集 $T = \{(x_1, y_1), \dots, (x_n, y_n)\}$, 在线性函数集合 $f(x) = (w \cdot x) + b, w \in R^n, b \in R$ 中确定回归函数, 依照正则化和结构风险准则估计出 w 和 b 。根据最小二乘法(Least Square)求解, 其中,

$$\min L(w, b) = \sum_{i=1}^N (y_i - w^T x_i - b)^2, \quad \xi_i = w^T x_i + b - y_i \text{ 记为回归误差。}$$

综上所述, 根据结构最小化准则, 建立回归问题的数学模型如下:

$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \zeta_i) \\ \text{s.t.} \quad & (W^T x_i + b) - y_i \leq \varepsilon + \xi_i \\ & y_i - (W^T x_i + b) \leq \varepsilon + \zeta_i \\ & \xi_i, \zeta_i \geq 0, i = 1, 2, \dots, N \end{aligned} \quad (3-20)$$

目标函数中 $\|w\|^2$ 指置信范围, 它代表了函数的推广能力; ξ_i, ζ_i 为松弛变量, 分别代表了在不敏感损失 ε 的约束下, 所容许的允许误差的上限和下限, $\sum_{i=1}^N (\xi_i + \zeta_i)$ 则表示了函数的学习能力, 被称为经验风险; ε 表示不敏感损失系数, 指的是存在一个不敏感区域 ε , 在这个特定区域内, 目标函数的点误差为 0。 C 参数 ($C \geq 0$) 即为惩罚因子, C 的合理设置可以实现经验风险与置信范围的有效这种, 最小化这两项指标的加权之和, 通过控制对误差超出 ε 的样本的惩罚程度以最小化模型的结构风险。

为求解上述回归问题, 引入拉格朗日函数:

$$\begin{aligned} L = & \frac{1}{2} W^T W + C \sum_{i=1}^N (\xi_i + \zeta_i) + \sum_{i=1}^N \alpha_i ((W^T x_i + b) - y_i - \varepsilon - \xi_i) \\ & + \sum_{i=1}^N \alpha_i (y_i - (W^T x_i + b) - \varepsilon - \zeta_i) - \sum_{i=1}^N \beta_i \xi_i - \sum_{i=1}^N \beta_i \zeta_i \end{aligned} \quad (3-21)$$

根据 KKT 条件:

$$\begin{aligned} \frac{\partial L}{\partial w} &= w + \sum_{i=1}^N (\alpha_i - \alpha_i) x_i = 0 \Rightarrow w = \sum_{i=1}^N (\alpha_i - \alpha_i) x_i \\ \frac{\partial L}{\partial b} &= \sum_{i=1}^N (\alpha_i - \alpha_i) = 0, \alpha_i \cdot \alpha_i = 0, \xi_i \cdot \zeta_i = 0 \\ \frac{\partial L}{\partial \xi_i} &= C - \alpha_i - \beta_i = 0 \Rightarrow 0 \leq \alpha_i, \alpha_i \leq C \\ \frac{\partial L}{\partial \zeta_i} &= C - \alpha_i - \beta_i = 0 (C - \alpha_i) \xi_i = 0 \\ \alpha_i, \alpha_i, \beta_i, \beta_i, \xi_i, \zeta_i &\geq 0 \\ \alpha_i ((w^T x_i + b) - y_i - \varepsilon - \xi_i) &= 0, \beta_i \xi_i = 0, \\ \alpha_i (y_i - (w^T x_i + b) - \varepsilon - \zeta_i) &= 0, \beta_i \zeta_i = 0. \end{aligned} \quad (3-22)$$

SVR 通常引入拉格朗日乘子以得到原问题的对偶问题:

$$\begin{aligned} \max W = & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_j)(\alpha_j - \alpha_i) \langle x_i, y_i \rangle + \sum_{i=1}^N \left[\alpha_i (y_i - \varepsilon) - \alpha_j (y_i + \varepsilon) \right] \\ \text{s.t. } & \sum_{i=1}^N (\alpha_i - \alpha_j) = 0, 0 \leq \alpha_i, \alpha_j \leq C \end{aligned} \quad (3-23)$$

且有 $w = \sum_{i=1}^N (\alpha_i - \alpha_j) x_i$ 。最终构造出线性回归函数:

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_j) K(x_i, x) + b \quad (3-24)$$

其中, α_i , α_j 是非负的拉格朗日乘子, 对上述凸二次规划问题求解, 可以得出 α_i , α_j 。根据 KKT 条件, 可求得 α_i , α_j 和 b , 其中只有少数 α_i , α_j 为非零, 这些参数对应的样本, 即为处于不灵敏边界上或是处于边界外的样本, 被称为支持向量。我们可以看出, 原问题的对偶形式的函数为凸函数, 其并不存在局部最优解, 保证了最终结果为全局最优解。

3.2.2 核函数的选取

作为决定 SVR 模型性能优劣的关键因素, 核函数的种类决定了支持向量机算法的类型^[62]。如何合理构造出与实际问题相关的核函数, 一直是支持向量机研究的热点。一般来说, 我们首先通过非线性变换, 把原始非线性回归问题转化为某个高维空间甚至无限维空间中的线性回归问题, 然后在确定的高维空间(亦或无穷维空间)中确定最优回归曲线。这种由低维向高维的映射关系可以表示如下:

$$x \rightarrow \Phi(x) = (a_1 \Phi_1(x), a_2 \Phi_2(x), \dots, a_n \Phi_n(x)), a_n \in R, \Phi_n \in R \quad (3-25)$$

例如, 将二维数据 (x_1, x_2) 通过二阶多项式映射到六维空间:

$$(x_1, x_2) \rightarrow \Phi(x_1, x_2) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, \sqrt{2}x_1x_2, x_2^2) \quad (3-26)$$

类似的, 支持向量回归机模型首先通过把训练点 (x_i, y_i) , $i=1, 2, \dots, N$ 映射到一个高维空间中, 确定训练点 $(\Phi(x_i), y_i)$, $i=1, 2, \dots, N$, 然后把映射后的训练集在高维空间中进行线性回归。在寻求对偶问题的最优函数的过程中, 仅仅涉及到训练集样本间的内积运算, 故可以使用核函数 $K(x_i, x) = \langle \varphi(x_i), \varphi(x) \rangle$ 代替对偶函数中的内积运算, 非线性回归函数变换如下:

$$f(x) = w^T x + b = \sum_{i=1}^N (\alpha_i - \alpha_j) K(x_i, x) + b \quad (3-27)$$

任何半正定的对称函数均可作为一个核函数, 核函数满足以下 Merce 条件:

$$\alpha_1, \alpha_2, \dots, \alpha_N \in R \rightarrow \sum_{i=1}^N \alpha_i \alpha_j K(x_i, x_j) \geq 0 \quad (3-28)$$

对任意的 $\int g^2(x) dx \leq \infty$, 均有 $\iint K(x, z) g(x) g(z) dx dz \geq 0$ 。

核函数是 SVR 算法的一个重要参数,选择不同的核函数就会产生不同的算法,目前常用的核函数主要分为以下四类:

1)线形核函数

$$K(x_i, x_j) = x_i \cdot x_j \quad (3-29)$$

2)多项式核函数

$$K(x_i, x_j) = [(x_i \cdot x_j) + 1]^q \quad (3-30)$$

式中, $q = 1, 2, \dots, n$;

3)RBF 径向基核函数

$$K(x_i, x_j) = \exp\left\{-\frac{\|x_i - x_j\|^2}{\sigma^2}\right\} \quad (3-31)$$

其中, σ 表示核函数的宽度;

4)Sigmoid 核函数

$$K(x_i, x_j) = \tanh(v(x_i, x_j) + c) \quad (3-32)$$

不同类型核函数之间的性能存在显著差异。其中,线性核函数无法实现非线性输入值的有效处理;在应用 Sigmoid 函数的过程中,某些参数可能会出现错误值;多项式核函数的高维核参数量明显多于径向基核函数,在训练 SVR 模型的过程中,与 RBF 核函数相比,多项式核函数会消耗更长的训练时间。

3.2.3 SVR 模型参数的寻优

在选取了合理的核函数后,需要对 SVR 算法中的参数进行调整。合理调整模型参数就是依照一定的搜索准则,在所设定的参数空间内寻求最优参数组合的过程。在选定 RBF 核函数的情况下,超参数 C , ε 以及核参数 σ 的设置合理与否,直接决定了支持向量模型的泛化性能的优劣程度,最终影响支持向量回归模型的学习预测能力^[61]。

惩罚参数 C 决定了权重向量的大小。在已知的数据空间中, C 通过控制置信区间的比例大小和学习机器的经验风险,保证学习机器具有更优秀的泛化能力。针对不同的问题,最优的惩罚参数 C 取值不同。而在确定的数据集中,惩罚参数 C 取值越小,意味着对样本数据中经验误差的样本惩罚就越小,在一定程度上导致训练误差和经验风险的变大;而 C 取值越大就表示模型拥有较高的学习精度,但同时其泛化能力会降低,SVR 模型的推广能力就会变差。

作为 SVR 模型的另一重要参数,不敏感损失函数 ε 控制着回归函数对训练集的不敏感区域的宽度,利用不敏感损失函数 ε 可以保证全局最小解的存在,优化可靠泛化界。而不敏感损失函数 ε 越小,SVR 模型的回归精度越高,但是同时产生的过多的支持向量数量会提升模型的复

杂度；反之，过大的不敏感损失函数 ε 会降低模型精度，支持向量数的相对减少，导致模型过于简单，从而降低模型的回归精度。作为 RBF 核函数中的重要参数， σ 较小可能会导致 SVR 模型出现过拟合现象，得到趋于单位阵的核矩阵；相反，较大的 σ 会导致核函数趋于常值，而使 SVR 模型的学习无效。

当前对于 SVM 算法中的参数优化研究主要包括两个方面：部分学者寻找更快速的方法计算推广误差的界或是更易于计算的推广误差的界；部分学者致力于提高参数空间搜索的效率。SVM 寻优方法主要包含经验法以及 VC 维方法两种。经验法具有相对简单性以及良好的使用效果两个显著优势，有助于人们对 SVM 寻优过程的理解和实现。学术界比较常用的经验参数优化算法包含留一法(Leave one out)和步步为营法(Bootstrap)两种，不同优化算法拥有独特的特点：

1)留一法

选取原始训练样本集中 $N-1$ 个数据，第 N 个数据用做验证集，这样就可以得到 N 个模型。最优参数选择依赖于这 N 个模型的最终训练结果。该方法即为留一法。理论上讲，留一法是期望风险的无偏估计。首先，不同的回合中，全部的样本均被选用为训练模型，这种方法所确定的 SVM 模型与原始样本的分布最为接近，进而最终取得可靠性较高的评估结果。此外，在实验的过程中，由于不存在干扰实验数据的随机，实验过程可以实现复制。但是，由于需要建立与原始数据样本数量相同的模型数量，留一法耗费较高的计算成本很高；且在实际操作中，面对过多的原始数据，留一法大大增加工作量且消耗大量时间，且不易寻得最优参数。

2)步步为营法

作为估计统计参数的采样方法，步步为营法并未像留一法一样除去样本，而是采用替代的方法，确保训练样本数量 N 值不变。由步步为营法所确定的推广误差的估计拥有很小的方差。较之于直接估算 SVM 模型的推广误差，Bootstrap 方法则通过估算推广以及训练过程误差的差，最终确定拥有极小方差的推广误差的估计值。

在确定支持向量机中的最优参数的过程中，网格搜索法是一种常用的经验法，其基本思想在于首先确定各个参数的合理搜索范围，接着基于经验法等对参数搜索范围进行等分，然后对每间隔等分中的任意确定出的某组参数进行验证，最终选择使验证集的预测性能最优的那组参数作为 SVM 模型所采用的最优参数。从本质上考虑，作为一种穷举法，网格搜索法可以通过在某个区间内按步长进行穷举，取代了传统的整个参数空间内的穷举，降低了计算时间和计算成本。在实际操作过程中，研究者如何依据经验确定最合理的参数搜索范围以及如何正确设置搜索步长显得极为关键。

3.2.4 SVR 模型的适用性

SVR 算法具有结构简单、学习速度较快、全局最优以及良好的泛化性能等优点，它在解决

小样本、非线性的高维数问题中呈现出卓越的性能，成功克服了传统方法中所出现的局部极小等问题：

1)与传统基于经验风险最小化原则的神经网络方法相比，SVR 是针对结构风险最小化理论而提出的，对于有限样本的训练 SVR 取得了较小的误差，这保证了对于独立的测试样本，该方法仍可以保持小的误差，它拥有非常卓越的学习能力和推广能力，；

2)本质上而言，SVR 算法解决的是凸二次规划问题，可以保证最终获得的局部最优解为全局最优解，克服了人工神经网络中无法取得全局最优解的困难，提高了模型的泛化能力；

3)通过引入核函数，SVR 算法的最终目标函数仅仅取决于支持向量的数目，并非由样本空间的维数决定，克服了传统算法的“维数灾难”难题；少数支持向量决定最终结果这一特征有助于剔除大量冗余样本，进一步抓住关键样本，极易将 SVR 算法推广到非线性问题中。

支持向量回归在国际油价分析预测中的应用，是学术界研究热点之一。其中 Fernandez^[62](2003)等应用支持向量机预测了国际石油价格，实证结果显示了 SVM 的预测精度优于传统的 ARIMA 和 ANN 模型。类似的，祝金荣^[63](2007)构建了基于支持向量回归机方法的应用国际石油期货价格短期预测模型，综合应用 RBF 核函数和序贯最小优化算法以最终确定最优参数组。本文通过实证分析有效证明，与 RBF 神经网络和 ARMA 模型相比，支持向量机取得更高的预测精度，在分析国际油价波动趋势方面具有很好的应用前景。刘立霞等(2008)^[64]学者将最小二乘支持向量回归模型应用于国际油价的集成预测，模型最优参数的设定依赖于网格搜索法，通过对比该方法与 RBF 神经网络模型的实证预测结果，证明了所提出的支持向量机预测模型拥有较高的拟合优度和预测精度。

3.3 本章小结

本章详细介绍了独立成分分析方法以及支持向量回归的相关理论。作为实现盲源分离的一种重要方法，独立成分分析仅仅依赖于统计独立这一假设，就可以从观测到的混合信号中分解出若干独立成分，其在金融数据中的成功应用，显示了其在揭示国际原油价格波动影响因素方面拥有显著的可行性：在寻求独立成分的过程中，结合时间序列信号的具体特征合理选取独立成分个数，并对分解出的独立成分给出具体经济解释，帮助国家正确理解国际原油市场。支持向量回归模型以结构风险最小化为原则，从统计学习理论出发，适用于复杂系统的建模等问题，具有极好的学习能力和泛化能力，可以较好的处理国际原油市场波动的非线性特点。本章对独立成分分析和支持向量回归模型及其适用性给出具体阐述，为接下来构建国际原油价格集成预测模型提供坚实的理论基础。

第四章 国际原油价格集成分析预测模型

第三章详细介绍了独立成分分析方法和支持向量回归方法的基础理论，并对独立成分分析在寻找金融数据潜在因素方面的可行性，以及支持向量回归模型在解决复杂非线性系统问题中的适用性进行了具体分析。基于“先分解后集成”的思想，一个自然的想法是，将独立成分分析这种分解方法与支持向量回归模型相结合，构建集成的国际原油价格预测模型。因此，本章针对国际原油市场提出了 ICA-SVR² 集成分析预测模型。

ICA-SVR² 是一种多尺度分析预测模型，首先采用 ICA 分解国际原油价格时间序列数据确定独立成分，找到国际油价的潜在影响因素；接着逐一对各独立成分建立 SVR 模型并预测，最后再利用 SVR 集成各独立成分的预测值，最终确定油价的重构值。鉴于分解得到的独立成分能够更好的反映原时间序列的波动特征，且相对复杂的油价原始序列值拥有更为稳定的内在结构。从理论分析来看，这种多尺度分解模型的预测效果，应该优于单尺度模型，且可以有效提升预测值的经济解释能力。本章首先提出 ICA-SVR² 集成模型的框架，整体把握集成模型的分析预测流程，对模型实施过程中包括独立成分个数的确定、SVR 核函数和最优参数的选择设定等关键问题给出详细分析。

4.1 ICA-SVR² 模型框架

ICA 可以将非平稳的国际油价序列分解为若干个拥有具体经济含义的独立成分，其有助于更清楚的把握油价的复杂的波动机制。而支持向量回归方法能很好的实现非线性问题的回归预测。本文在综合独立成分分析方法与 SVR 模型优点的基础上，建立了 ICA-SVR² 集成国际油价分析预测模型，具体步骤如下：

步骤 1: 国际原油价格时间序列的独立成分分析。利用 ICA 模型将国际原油价格时间序列信号分解，结合分解得到的独立成分的具体波动态势分析其经济含义，深入理解把握国际原油价格特性。

步骤 2: 逐一预测独立成分：根据各个独立成分的数据特征，分别寻求最优的 SVR 参数进行逐一预测：合理设置核参数 σ 和超参数 C 、 ϵ 的搜索范围，分别确定可以最好的拟合独立成分的 SVR 模型，对各项独立成分进行预测。

步骤 3: 重构独立成分的预测值：本文采用 SVR 训练学习各个分项预测值之间的组合关系，确定最优 SVR 模型，得出最终预测值。该部分的基本思想是：选取同一时刻各分项独立成分的预测值作为支持向量模型的输入值，选用相应时刻国际原油实际价格作为输出，对足够的样本进行不断训练学习，确定各项独立成分预测值和油价实际值之间的函数映射关系；而面对最终

训练好的最优模型，将各项独立成分的预测值输入到模型中后，SVR 模型的输出结果即为国际油价的最终预测结果。

结合 ICA-SVR² 模型实施的具体步骤，图 4.1 给出了该模型的集成预测框架。

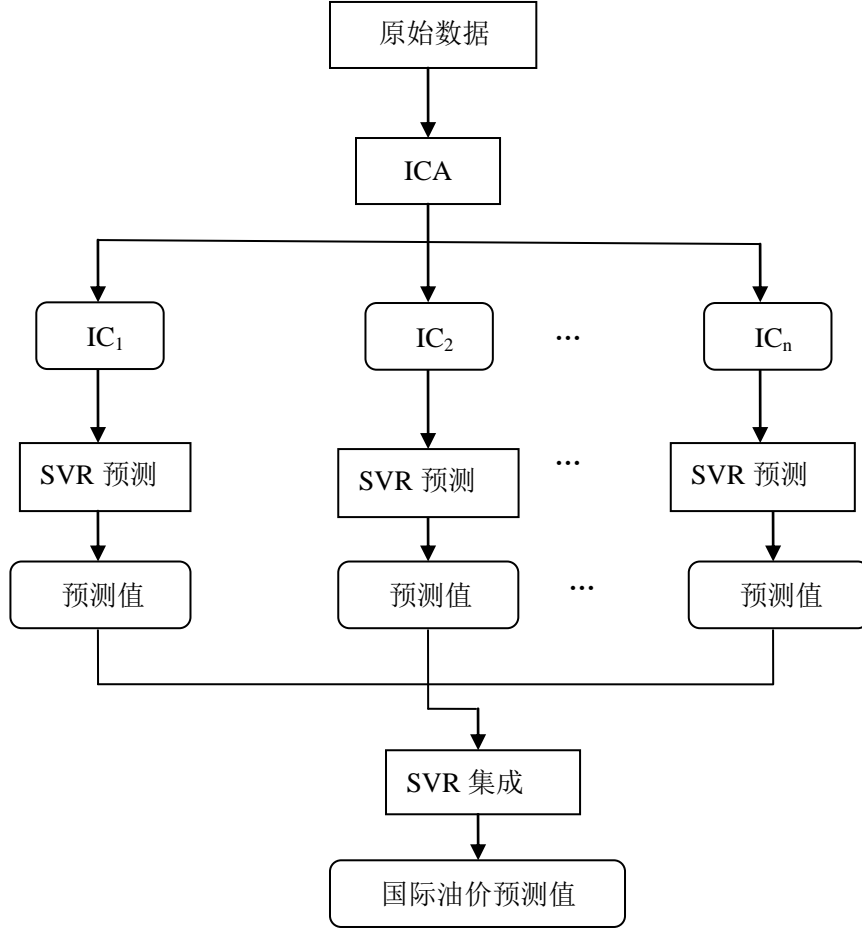


图 4.1 ICA-SVR² 集成分析预测框架

4.2 ICA-SVR² 集成分析预测流程

设有原始序列 $\{x_t\}_{t=1}^{t=n}$ ，要预测的数据从 $t=n+1$ 时刻到 $t=n+l$ ，共 l 期数据，ICA-SVR² 集成预测模型流程包括独立成分的确定、SVR 预测、SVR 集成三个步骤。

步骤 1 国际油价的独立成分分析

对原始国际油价时间序列进行独立成分分析，假定得到 m 个独立成分，记为：

$$\{IC_{1t}\}_{t=1}^{t=n}, \{IC_{2t}\}_{t=1}^{t=n}, \dots, \{IC_{mt}\}_{t=1}^{t=n}$$

结合分解得到的独立成分，深入探究国际油价的影响因素。

步骤 2 独立成分的 SVR 预测

本文建立的 SVR 预测模型进行的是单步预测,以获得一步提前预测值。对每个单独的独立成分,分别选取特定的核函数,寻求最优参数对 (C, ε, σ) ,建立 SVR 模型,进行预测,得到预测值 $\{IC_{it}\}_{t=n+k}^{t=n+k}$, $i = (1, 2, \dots, m)$, $k = (1, 2, \dots, l)$ 。

在 SVR 建模过程,选用波动序列的滞后 p 期值作为输入变量,目标变量为往后一期的真实值。在外推预测(测试集)时,首先预测得到样本外第一期的估计值,接着将该期的实际值代入模型,预测样本外第二期的估计值,直到得到样本外第 l 期估计值。用函数 f 表示 SVR 模型,则单步预测模型可表达为

$$IC_{i(t+1)} = f(IC_{i(t-lag_order+1)}, IC_{i(t-lag_order+2)}, \dots, IC_{it}) \quad (4-1)$$

$$IC_{i(t+2)} = f(IC_{i(t-lag_order+2)}, IC_{i(t-lag_order+3)}, \dots, IC_{i(t+1)}) \quad (4-2)$$

$$IC_{i(t+l)} = f(IC_{i(t-lag_order+l)}, IC_{i(t-lag_order+l+1)}, \dots, IC_{i(t+l-2)}, IC_{i(t+l-1)}) \quad (4-3)$$

其中, $i = (1, 2, \dots, m)$, lag_order 为模型的滞后阶数 p 。

步骤 3 独立成分预测结果的 SVR 集成

步骤 2 得到了各项独立成分的预测值,步骤 3 则以这些预测值作为输入,通过构建 SVR 集成模型得到最终的对目标量的预测值。集成方法的选择,对于“先分解后集成”方法性能的好坏,有决定性的影响。本文在详细分析支持向量回归在复杂非线性系统建模问题中的有效性的基础上,最终采用 SVR 作为集成方法。由于独立成分的预测值可通过 SVR 重构出目标序列,用函数 g 表示 SVR 集成模型,最终的预测值 $x_{i(n+k)}$ 可由下式得到:

$$x_{i(n+k)} = g(IC_{1(t+k)}, IC_{2(t+k)}, \dots, IC_{m(t+k)}) \quad (4-4)$$

式中, $i = (1, 2, \dots, m)$, $k = (1, 2, \dots, l)$ 。

4.3 集成分析预测中的关键问题

本文构建的 ICA-SVR² 集成模型性能的优劣依赖于独立成分的正确选择,以及支持向量回归模型中核函数和超参数的合理设置:利用独立成分分析对国际原油价格原始序列进行处理,独立成分的个数决定了分解得到的独立成分能否很好的反映国际原油价格波动特征;而作为集成预测模型中的关键步骤,集成方法的好坏直接影响最终的预测效果,合理选择设置支持向量回归模型中的核函数及相关参数极为重要。

4.3.1 独立成分个数的确定

在实际应用 ICA 算法估计独立成分的过程中,首要问题是如何在没有关于独立成分数目的先验信息确定合适的独立成分个数:确定独立成分个数的过程中,需要注意的是,在独立成分

个数过小的情形下，无法充分分解出国际原油价格不同的影响因素，体现不出独立成分分析的优越性；但是若独立成分个数过多，则需要更多的模型对各项独立成分分量进行预测，不仅大大增加了计算量，也加大了预测难度。目前学术界尚未存在统一的独立成分个数确定方法，本文参考主成分分析中主成分个数的寻找方法来确定独立成分个数。

在主成分分析方法中，设样本集 $X = \{x_1, x_2, \dots, x_m\}$ 包含 m 个 n 维样本，即 $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ ，我们希望找到一个 R^n 的低维子空间 $R^p (p < n)$ ，使得 y_1, y_2, \dots, y_m 是 x_1, x_2, \dots, x_m 在子空间 R^p 上的投影，且用 y_1, y_2, \dots, y_m 来描述 x_1, x_2, \dots, x_m 时信息损失达到最小。不相关的主成分的确定方法主要是通过对方差矩阵进行特征分解，以得出数据的主成分(即特征矢量)与它们的权值(每个主成分的方差贡献率)。

令 $x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T$ ，协方差矩阵为：

$$\Sigma = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^T \quad (4-5)$$

其中 $\bar{x} = 1/m(\sum_{i=1}^m x_i)$ ，则协方差矩阵 Σ 为非负定矩阵，必定存在正交矩阵 U ，使得，

$$U \Sigma U^T = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \quad (4-6)$$

其中，特征值 $\lambda_i \geq 0$ ， $i = 1, 2, \dots, n$ ， $U^T = (U_1, U_2, \dots, U_n)$ ， U_i 表示 U^T 的第 i 列，则 U_j 是对应于 λ_j 的协方差矩阵的标准正交化特征向量。令 $X = (X^{(1)}, X^{(2)}, \dots, X^{(n)})^T$ ，则

$$Y = UX = (U_1, U_2, \dots, U_n)^T X \quad (4-7)$$

Y 的各个分量 $Y^{(j)} = U_j^T X$ 即为不相关的主成分。 $Y^{(j)}$ 是 X 的第 j 个主成分， $Y^{(j)}$ 的贡献率为 $\lambda_j / \sum_{i=1}^m \lambda_i$ ，而 $\sum_{i=1}^j \left(\lambda_i / \sum_{i=1}^m \lambda_i \right)$ 为前 j 项主成分的累计贡献率。

在实际问题中，主成分的确定主要存在三种通用的选择标准：1)累积贡献率法：如果前 k 个特征值的累计贡献率达到 85%，则通常取相应的 k 个主成分，因为这 k 个主成分的方差已占全部的总方差的 85% 以上，反映了至少 85% 的原有变量的信息；2)陡坡风险，寻找“肘点”，在该点之前，特征值呈现快速下降趋势，而之后下降逐渐平缓；3)Kaiser 法：Kaiser 建议选取特征值大于 1 的主成分。

本文在确定独立成分个数时，参考主成分个数的确定方法，计算原始混合数据协方差矩阵及其特征值，根据特征值计算累积贡献率并结合 Kaiser 原则来决定独立成分的个数。

4.3.2 SVR 核函数的选择

目前比较常用的核函数类型包括线性核、多项式核、RBF 核和 Sigmoid 核等，实践中多使用 RBF 径向基核函数：

$$K(x_i, x_j) = \exp \left\{ -\frac{\|x_i - x_j\|^2}{\sigma^2} \right\} \quad (4-8)$$

式中， σ 为核函数的宽度；

与其他核函数相比, RBF 函数拥有以下几方面优势: 首先, 作为处理以非线性方式存在的线性问题的有力工具, RBF 核函数能够有效的实现输入空间至特征空间的非线性方式映射, 在特定参数选择下核函数只是 RBF 核的一个特例; 其次, 较之多项式核函数以及 Sigmoid 核函数, RBF 核函数在数值计算中会更少遇到困难。故在不清楚函数性质的情况下, RBF 函数是最好的选择。因此, 本文选择 RBF 核函数作为国际原油价格集成预测方法中的非线性变换方式。

4.3.3 SVR 参数寻优

当选定 RBF 核函数建立 SVR 模型后, 就存在超参数 C, ε 以及核参数 σ 需要进行合理优化, 学术界并没有公认的 SVR 参数寻优的最好方法。本研究利用网格搜索法与交叉验证法相结合的方式训练学习规则, 寻找出 SVR 模型的最优参数组, 最终确立国际油价预测模型。

首先合理设置待搜索参数组的空间范围, 将参数组在其设定的空间范围内进行划分, 得到一定数量的网格, 接着遍历所确定网格中所有的点来确定最优参数组, 取代了传统的整个参数空间内的穷举, 降低了计算时间和计算成本。因此本文选用交叉验证法和网格搜索法相结合的方式开展重复训练和测试工作, 在一定的范围划分网格并遍历网格内的所有点进行取值, 对于取定的参数对, 通过交叉验证法进行验证集准确率率的计算, 最优参数组即为最大化验证集准确率的那组参数。

交叉验证法的基本思想在于如何将原始训练数据合理分配为训练集和验证机两组, 确定部分原始数据作为训练集, 原始数据中的剩余部分则被选用为验证集。交叉验证法首先针对训练集进行 SVR 模型的学习, 接着采用验证集进行训练出的 SVR 模型的效果验证, 选用验证集处的预测准确率作为评价支持向量回归模型的最终性能指标, 选取验证集预测准确率最高的相应参数组作为最终确定的最优支持向量回归模型。学者通常采用 K -fold CV 方法, 首先将原始数据等分为 K 组($K>1$), 分别选用每个子集数据, 逐一作为验证集, 而剩余的 $K-1$ 组子集数据则相应的作为训练集, 最终会得到 K 个模型, 对 K 个模型所取得的验证集的预测准确率率的进行平均化处理, 最终的 K 个平均数作为性能指标, 选用准确率平均数最大的那个模型作为最终的 SVR 模型。较之传统的留一法, K -CV 方法不仅大大提升训练达成率, 实现过学习以及欠学习结果的有效避免, 最终获得说服力较高的结果。

本研究将网格搜索法与交叉验证法相结合, 对 SVR 模型进行训练学习, 寻找出 SVR 模型的最优参数组: 首先对网格搜索法的参数搜索范围和步长进行合理设置; 然后利用交叉验证法, 将训练样本等分为 K 份, $K-1$ 组作为验证集, 剩余的第 K 组作为验证集; 针对每一个参数对, 选用验证集的均方根误差作为评价准则, 寻求最优参数组, 训练出最终的油价序列预测模型。

4.4 本章小结

本文构建了基于独立成分分析和支持向量回归模型的 ICA-SVR²。该集成预测模型的流程包

包括独立成分确定、SVR 预测、SVR 集成三个部分：首先对历史油价序列进行独立成分分析，结合分解得到的独立成分的波动特征，具体分析每项独立成分的经济含义。然后对于不同的独立成分，建立不同的 SVR 模型进行预测。最后在预测值重构阶段，选定国际原油基准价格作为输出变量，同时刻各项独立成分值作为输入变量，利用足够的训练集和验证集数据重复训练进行参数寻优，最终确定拟合效果最为理想的预测模型。针对 ICA-SVR² 集成预测模型中的关键问题，本研究利用协方差矩阵特征值来确定独立成分个数，通过网格搜索法和交叉验证法来确定最优参数。为了验证本文构建模型的可行性和有效性，接下来的章节将系统评估集成模型在探究国际原油市场价格形成机制和油价走势预测方面的表现。

第五章 实证分析—WTI 原油价格分析预测

5.1 实证分析设置

5.1.1 数据来源和预测评价准则

作为实证算例，本文选取世界原油三大基准价格中的英国布伦特原油(Brent)现货，美国西德克萨斯轻质低硫原油(WTI)现货，以及在 NYMEX(纽约商品交易所)交易的 Brent 期货、WTI 期货 4 列国际油价序列进行独立成分分析，并选取世界原油市场最重要的基准油价 WTI 现货价格序列作为预测目标。实验中采用国际油价周度数据，数据样本为 2004 年 7 月 2 日~2014 年 6 月 27 日，共 4×522 组数据。本文所选用的数据均来自于美国能源署 EIA。

为了评价本文构建模型的预测性能，本文选用三个评价标准：

1) 均方根误差(RMSE):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (T_t - P_t)^2}{n}} \quad (5-1)$$

2) 平均绝对百分比误差:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{T_t - P_t}{T_t} \right| \quad (5-2)$$

3) 方向变化统计量:

$$DS = \frac{100}{n} \sum_{t=1}^n d_t, \text{其中 } d_t = \begin{cases} 1, (P_t - P_{t-1})(T_t - T_{t-1}) \geq 0 \\ 0, (P_t - P_{t-1})(T_t - T_{t-1}) < 0 \end{cases} \quad (5-3)$$

T_t 为 t 时刻的国际原油价格的实际值， P_t 为 t 时刻的国际原油价格的预测值， $t = 1, 2, \dots, n$ ， n 为预测值的总个数。

其中 RMSE 和 MAPE 用于衡量各个模型预测的准确性：RMSE 是预测值与实际值之间的误差指标，MAPE 则反映了预测值偏离于实际值的程度大小。对于原油价格预测而言，国际油价走势的方向性对于决策者来说也及为关键，本文采用整体的预测方向正确率 DS 来衡量各模型对油价方向性走势的预测能力。

5.1.2 实验设置和实现

在 SVR 建模过程中，本文选用波动序列的滞后一期值作为输入变量，采用移动窗口方法以获得一步提前预测值。本实验在 MATLAB 7.11.0 环境下进行，对于 4 组国际油价时间序列，独立成分的确定主要应用 Fast-ICA 算法；SVR 模型核心部分基于 Lin 等设计的 LibSVMs。包括

SVR 的应用和参数寻优等部分均在 Matlab 中自主开发。采用 M 语言对模型的训练步骤进行编写，通过 M 文件的建立执行，最终实现相关参数寻优过程以及最终预测效果的图像展示。

5.2 油价序列的独立成分分析

5.2.1 数据描述

本文选取 WTI 现货，Brent 现货，以及在 NYMEX 交易的 Brent 期货、WTI 期货 4 列国际原油周度价格序列进行独立成分分析，样本区间为 2004 年 7 月 2 日~2014 年 6 月 27 日，共 4×522 组数据。利用 2004 年 7 月 2 日到 2012 年 12 月 28 日的 4 列国际原油价格序列作为训练样本，共 4×444 组数据点，2013 年 1 月 6 日~2014 年 6 月 28 日作为测试样本，共 4×78 组数据。4 列国际原油价格原始时间序列见图 5.1 所示。

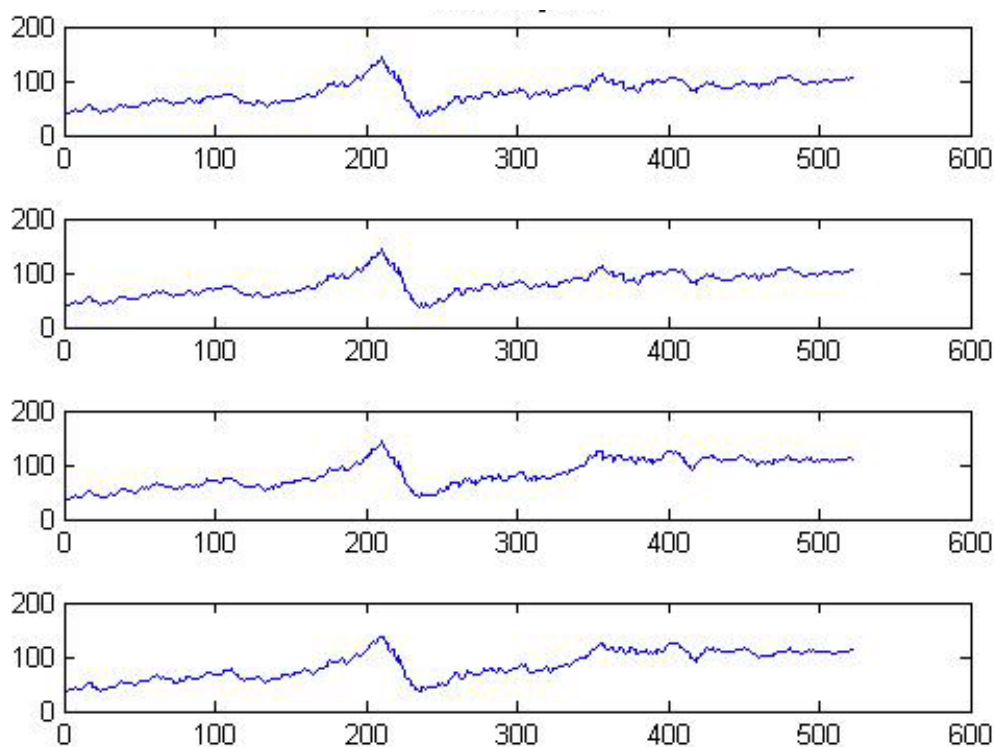


图 5.1 四列国际原油价格时间序列

图 5.1 从上至下的四项油价数据分别为：第一项是 WTI 现货价格，第二项是 WTI 期货价格，Brent 期货价格为第三项，第四项则是 Brent 现货价格。从图中可以看出，四项数据的波动态势比较接近，侧面反映不同的国际原油基准序列的形成过程中存在共同的驱动因素。

5.2.2 确定独立成分个数

在实际应用 Fast-ICA 算法估计国际油价独立成分的过程中，首要问题是独立成分个数的确定。本文基于累计贡献率原则和 Kaiser 法来确定独立成分个数，首先计算四列国际油价时间序

列数据的协方差矩阵，结果见表 5.1。

表 5.1 协方差矩阵

0.702864	0.00261	-0.55289	0.447538
-0.71119	-0.01626	-0.54264	0.446646
0.012361	-0.70594	0.45299	0.544326
-0.0066	0.708077	0.441201	0.551295

计算协方差矩阵的特征值，见表 5.2。通过对协方差矩阵的特征值表的分析，可以看出：

- 1) Kaiser 原则：特征值的下降趋势相当明显，最后一项特征值小于 1；
- 2) 累积贡献率法：前两项的累积贡献率达到 99.81%，远远超过额定阈值 85%。

表 5.2 协方差矩阵的特征值分解

2177.976	0	0	0
0	43.57151	0	0
0	0	4.156866	0
0	0	0	0.07025

在确定独立成分个数的过程中，需要注意的是，过小的独立成分个数过小，将导致不同频率国际原油价格信号所代表影响因素的遗漏，无法完全体现独立成分分析方法较之传统方法的优越性，故本文最终确定独立成分个数为 3。

5.2.3 独立成分的估计及其经济含义

独立成分个数确定以后，本文利用 Fast-ICA 算法对国际油价序列估计出三项独立成分(ICs)，见图 5.2，找到了隐藏在原始数据中的几类明显不同的基本影响因素。

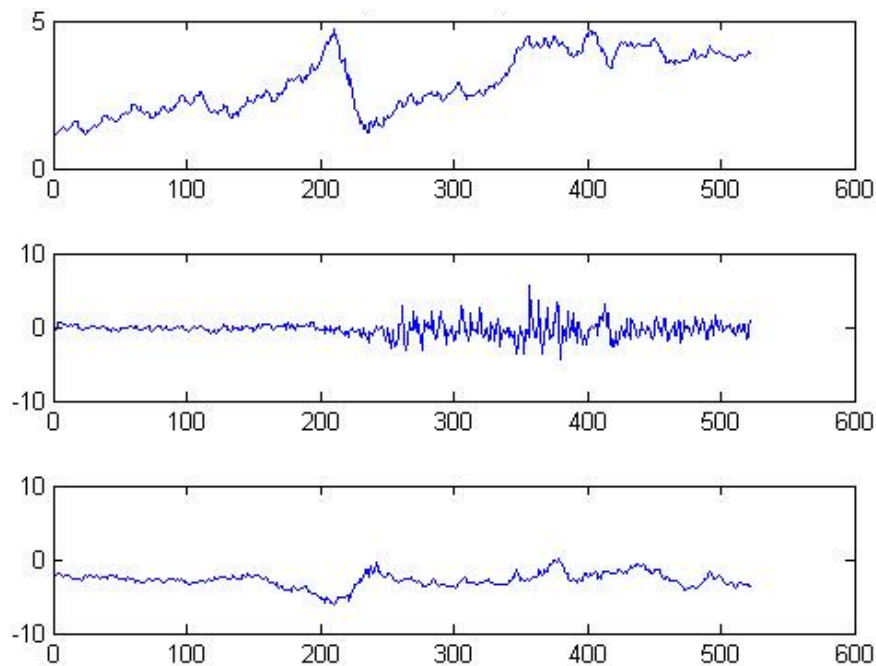


图 5.2 国际油价时间序列中的三项独立成分

三项独立成分拥有不同的波动特征，代表不同的影响因素：

a) IC_1 表现出低频且较为剧烈的波动特点。该项独立成分代表了经济政治等重大事件对油价的影响，是原油价格的重要组成部分。其中最突出的部分是 2008~2009 年金融危机期间，油价剧烈波动。国际油价出现结构断点，在金融危机结束之后，并未重现恢复到金融危机之前的水平。重大事件对油价产生频率较低的影响，持续期一般为 4~5 年；

b) IC_2 则表现出频繁而微小的波动特点，包括自然灾害、地缘政治冲突等各类突发事件所带来的影响，近些年来，伴随着快速发展的国际原油期货市场的，愈演愈烈的投机活动引发油价的频繁波动。考虑到该类因素对国际油价的影响时间较短，对油价的影响也相对较小，把这些因素划分到高频的市场短期波动项。市场短期波动项在世界原油价格短期预测中占据重要地位；

c) IC_3 反映了世界油价的趋势部分，其波动最为缓慢，对原油价格的长期趋势起到决定性作用。从原油价格的历史来看，国际油价在遭遇政治经济重点事件时会呈现出剧烈波动的态势，随着突发事件影响的逐渐削弱，世界原油价格会逐渐恢复到原油趋势价格的附近。趋势项为长期均衡油价提供参考。

由 Fast-ICA 算法分解得到的三项独立成分可以较好的反映 WTI 序列的波动特征，寻出隐藏在混合油价信号中的基本影响因素，克服了传统数据驱动预测方法无法给出合理解释的弱点，验证独立成分分析方法作为预处理技术在分解处理原始油价信号方面的合理性与有效性。

5.3 独立成分的 SVR 预测

5.3.1 SVR 预测模型的参数寻优

本文结合趋势项、重大事件影响项和市场波动项 3 项独立成分自身的波动规律，分别构建 SVR 模型进行预测。由于国际原油价格周度数据与前期数据密切相关，因此本文在预测独立成分时，选定滞后阶数 1，输入变量选择滞后一期的波动序列值。由于不清楚 SVR 模型中函数的具体性质，本文在构建 SVR 预测模型的时候均选取高斯 RBF 核函数。

对各项独立成分建立 SVR 预测模型时，以重大突发事件项为例，选取 2004.7.2~2012.12.28 期间 4×444 组数据作为训练样本确定最优 SVR 预测模型，2013.1.4~2014.6.27 期间 4×78 组数据作为测试样本。在采用网格搜索法来参数寻优的过程中，本文依照经验，首先设置超参数 ε 的搜索范围为 $\{1, 0.1, 0.01, 0.001\}$ ，超参数 C 的搜索范围为 $(2^{-1}, 2^8)$ ，步长为 1；核参数 σ 的搜索范围为 $(2^{-5}, 2^4)$ ，步长为 1，形成 $4 \times 10 \times 10$ 个 (C, ε, σ) 参数对。本文将训练样本等分为 5 份，4 份做训练集，一份做验证集，以验证集的均方根误差(RMSE)作为评价标准，最终取使得验证集 RMSE 最小的那组参数对作为最佳参数。

首先确定超参数 ε ，图 5.3 显示了不同 ε 水平下模型预测的具体性能。对比四幅图可以看出，模型的预测性能对 ε 不敏感，在 ε 取不同值的情况下最小训练误差均为 $\text{RMSE}=0.00339$ 。故本研究选取默认值 $\varepsilon = 0.01$ 。

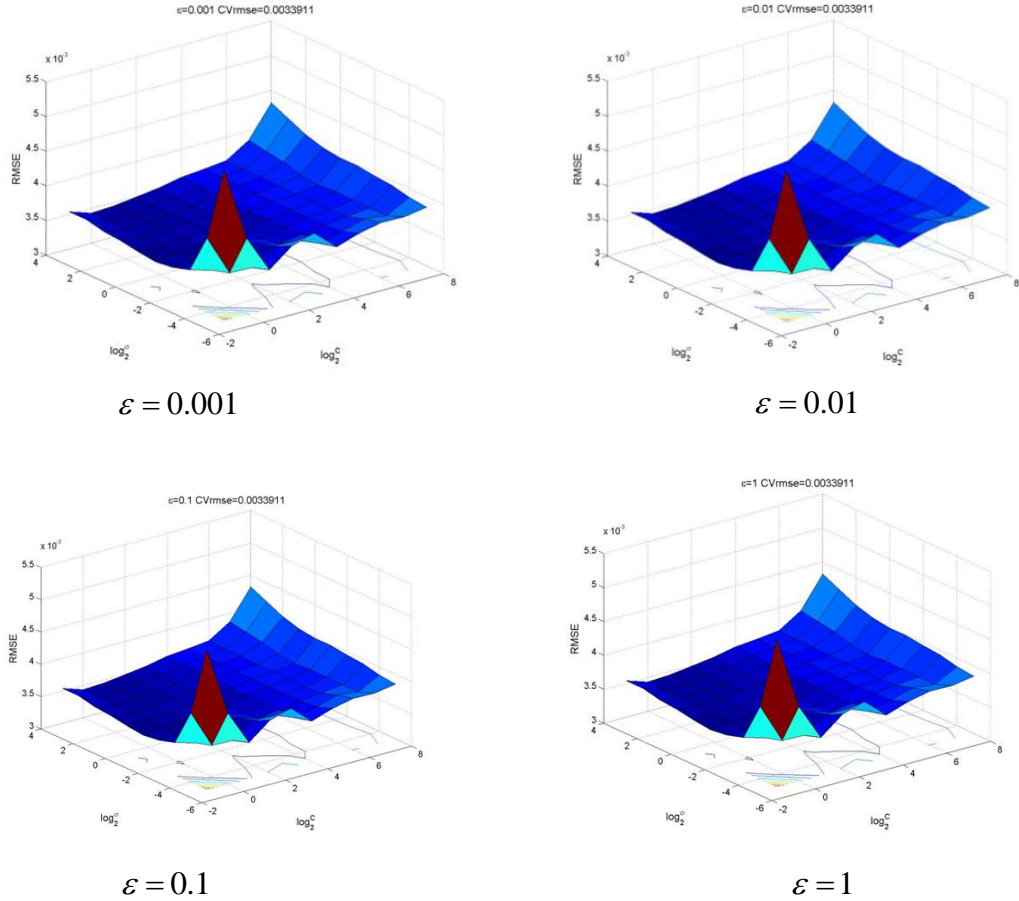


图 5.3 不同 ε 水平下 SVR 预测模型的均方根误差

当 $\varepsilon = 0.01$ 时，选择不同的 C 和 σ 建立的模型的预测效果分别见图 5.4 和图 5.5 所示。分析图 5.4 可知，在参数 C 的值一定的情形下，随着 σ 的逐渐增大，模型并未呈现出良好的收敛效果。由图 5.5 可知，当核参数 σ 不变时，随着 C 取值的增大，模型预测效果逐渐收敛，接着达到最小值，而后逐渐发散；其中 $C = 8$ $\sigma = 0.0313$ 时，模型验证集所得到的均方根误差取得最小值。

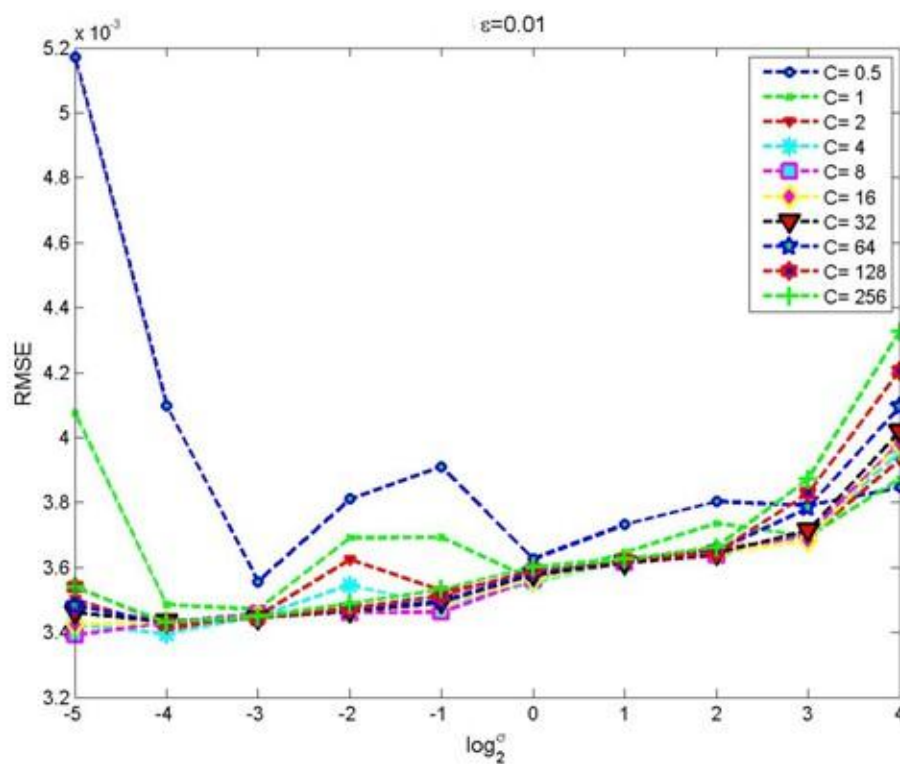


图 5.4 参数 σ 对 SVR 预测模型性能的影响($\varepsilon = 0.01$)

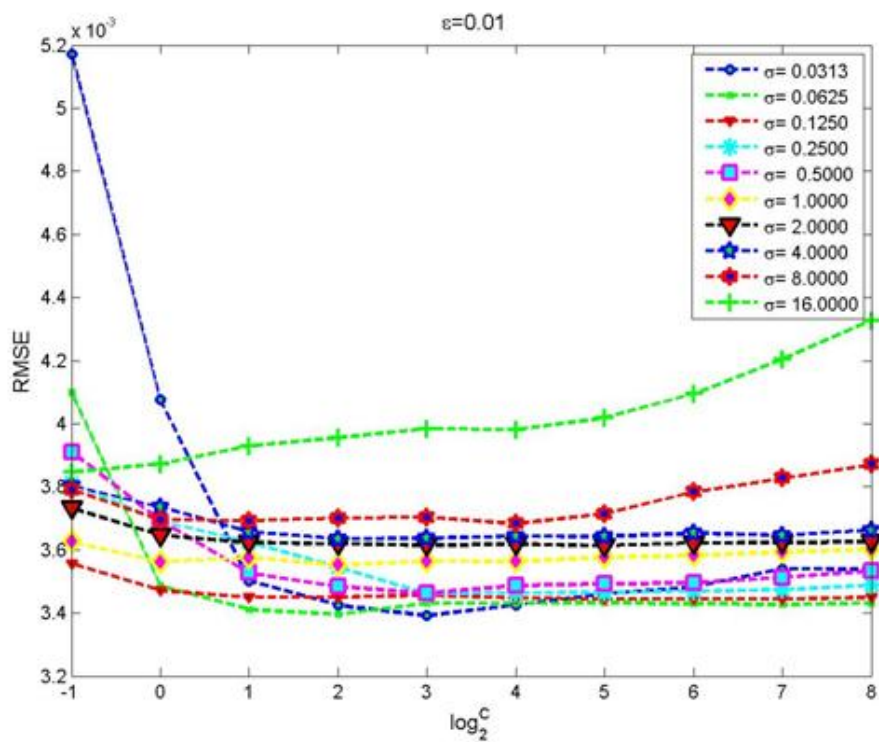


图 5.5 参数 C 对 SVR 预测模型性能的影响($\varepsilon = 0.01$)

由以上分析可知,当 $C=8, \sigma=0.0313, \varepsilon=0.01$ 时,模型对代表重大事件项的独立成分的预测效果最好。如图 5.6 所示,此时验证集的 RMSE=0.00339。

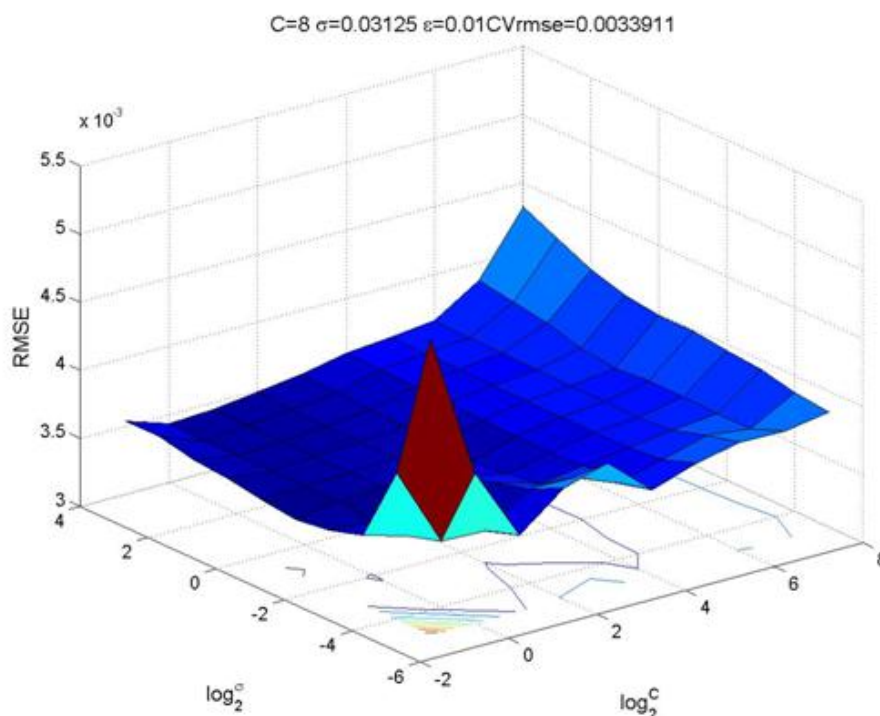


图 5.6 SVR 预测模型的最优参数组

5.3.2 三项独立成分的 SVR 预测模型及预测结果

表 5.3 总结了在构建重大事件项、市场短期波动项、趋势项的 SVR 预测模型过程中,分别选用的不同的核函数及参数项。

表 5.3 分项 SVR 预测模型核函数和参数

	核函数	C	ε	σ	支持向量个数
重大事件项	RBF	8	0.01	0.0313	31
市场短期波动项	RBF	2	0.01	2	162
趋势项	RBF	128	0.01	0.5	85

针对三项独立成分的特点,本文分别构建相应的支持向量回归模型,最终得到的三项独立成分测试集的预测误差见下表 5.4:

表 5.4 三项独立成分 SVR 预测模型测试集的预测误差

	RMSE	MAPE	DS(100%)
重大事件项	0.0108	0.0275	99.99%
市场短期波动项	0.1093	0.2083	87.12%
趋势项	0.0229	0.0131	99.99%

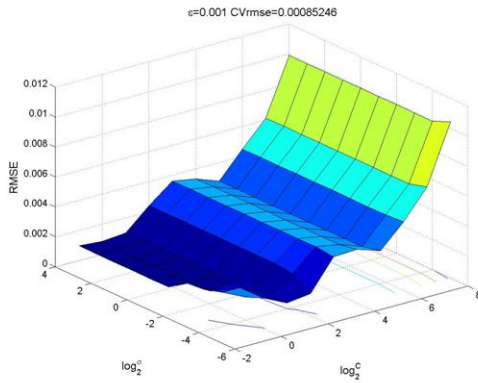
分析表 5.4 可知,针对于低频趋势项以及中低频重大突发事件项,SVR 模型呈现出相当出

色的预测能力，而在预测高频市场短期波动这一项时，SVR 模型表现出较为逊色的预测能力，造成这一结果的主要原因在于的频率较高的投机活动很大程度上造成油价的频繁波动，而伴随着其他天气、罢工等不可预期的因素的影响，油价更是呈现出了频繁无规律的波动特点，大大增加了预测难度。

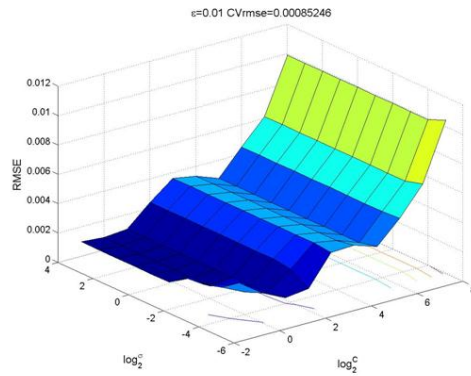
5.4 独立成分预测结果的 SVR 集成

5.4.1 SVR 集成模型的参数寻优

在对 3 项独立成分的预测值进行重构时，选取 2004.7.2~2012.12.28 期间 3 组独立成分的预测值和相应时刻的 WTI 原油价格共 4×444 组数据作为训练样本集，2013.1.4~2014.6.27 期间 4×78 组数据作为测试样本集。针对训练样本集，合理构造各项独立成分预测值与原油价格真实值之间的函数映射关系，构建出最优的集成 SVR 模型，后输入各独立成分的预测值，即可集成得到 WTI 原油价格的最终预测值。模型建立过程中，选用 RBF 核函数，将训练样本等分为 5 份，4 份做训练集，1 份做验证集，利用网格搜索法，设定超参数 ε 的搜索范围为 $\{1, 0.1, 0.01, 0.001\}$ ，超参数 C 的搜索范围为 $(2^{-1}, 2^8)$ ，搜索步长为 1；核参数 σ 的搜索范围为 $(2^{-5}, 2^4)$ ，搜索步长为 1，形成 $4 \times 10 \times 10$ 个参数对。对于每个参数对根据验证集的训练结果进行筛选，确定最终的模型参数。参数确定过程见图 5.7-5.9。



$\varepsilon = 0.001$



$\varepsilon = 0.01$

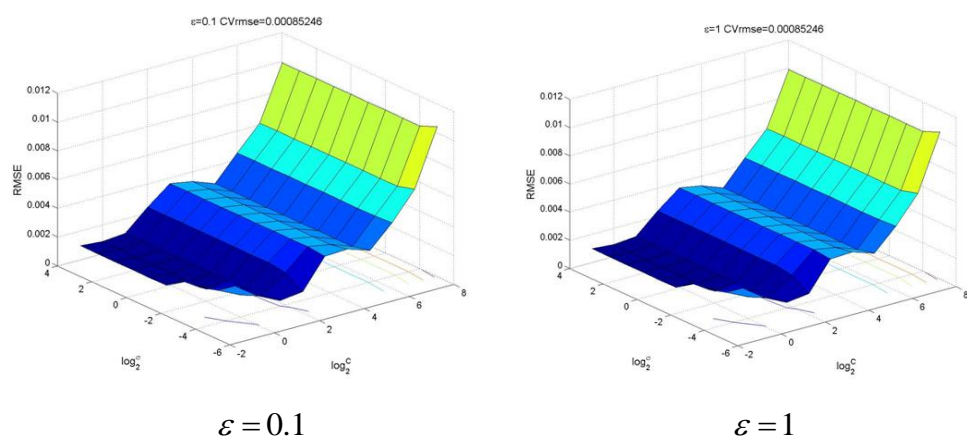


图 5.7 不同 ε 水平下 SVR 集成模型预测的均方根误差

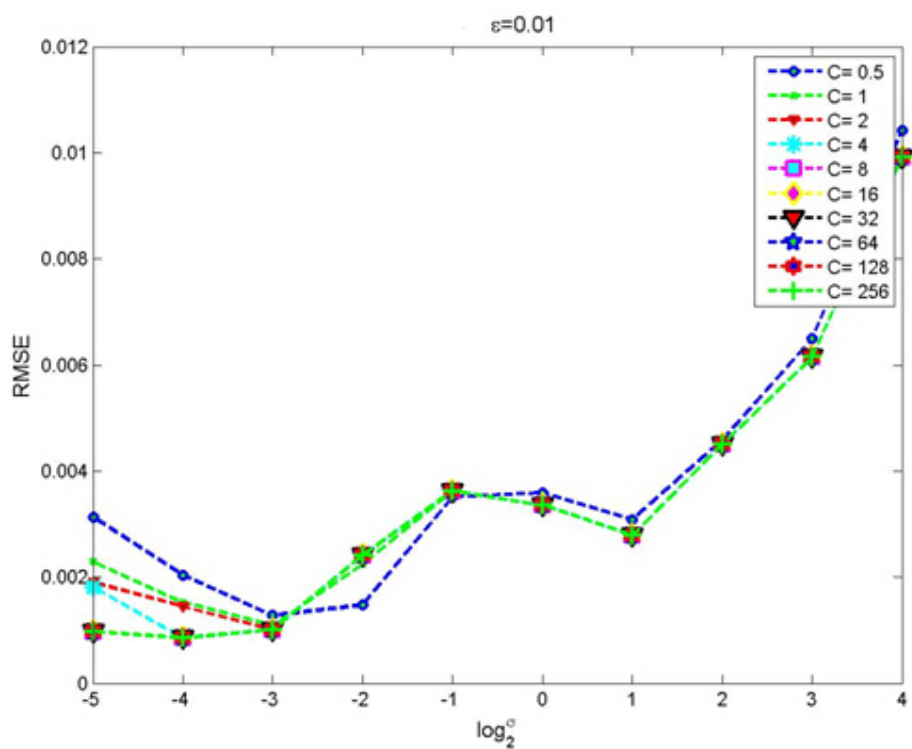


图 5.8 参数 σ 对 SVR 集成模型预测性能的影响($\varepsilon = 0.01$)

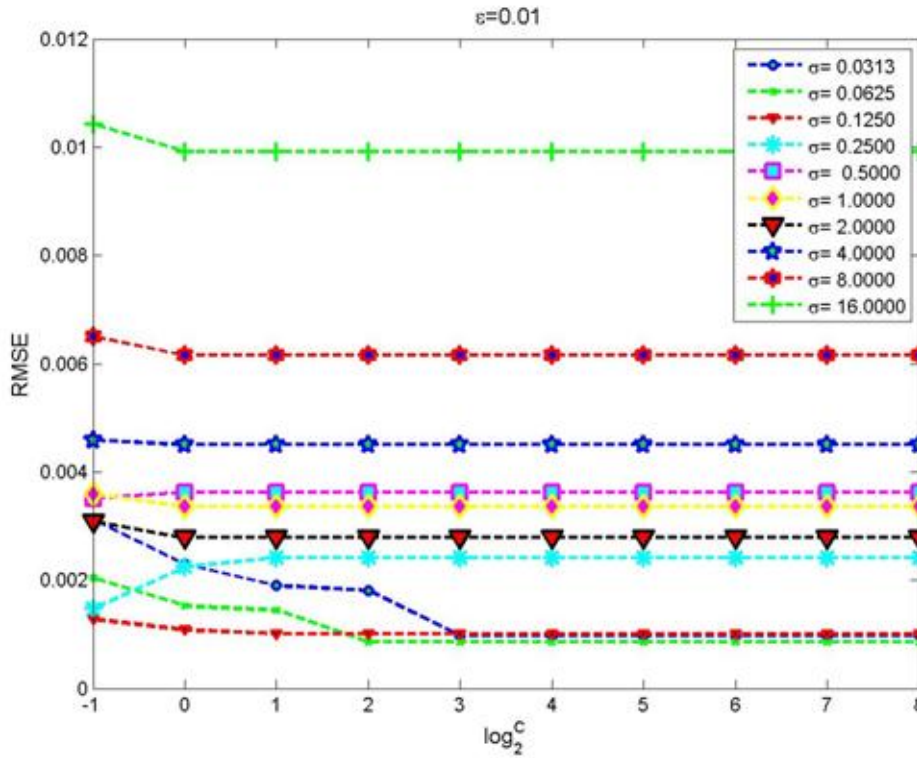


图 5.9 参数 C 对 SVR 集成模型预测性能的影响 ($\varepsilon = 0.01$)

从图 5.7 看出，独立成分的重构模型的预测性能对 ε 不敏感，故选择默认值 $\varepsilon = 0.01$ 。对图 5.8 分析可以看出，在参数 C 的值一定的情形下， $\sigma = 0.0625$ 时预测性能最优；从图 5.9 可以看出，在核参数 σ 不变的情况下，伴随着不断增大的 C ，模型预测效果呈现出逐渐收敛的趋势，确定最优参数 $C = 4$ 。最终确定 SVR 重构模型的最优参数组为 $C = 4, \sigma = 0.0625, \varepsilon = 0.01$ 。

5.4.2 SVR 集成模型的预测结果及鲁棒性分析

以最优参数组 $C = 4, \sigma = 0.0625, \varepsilon = 0.01$ 确定 SVR 集成模型，以三项独立成分的预测值作为输入，得到 WTI 原油现货价格的最终组合预测结果，如图 5.10 所示。基于独立成分分析所建立的国际油价预测模型很好的拟合了国际油价的走势，对于 2013 年 1 月 4 日至 2014 年 6 月 27 日期间的油价走势的预测达到比较高的精度，测试样本的均方根误差为 1.6251。

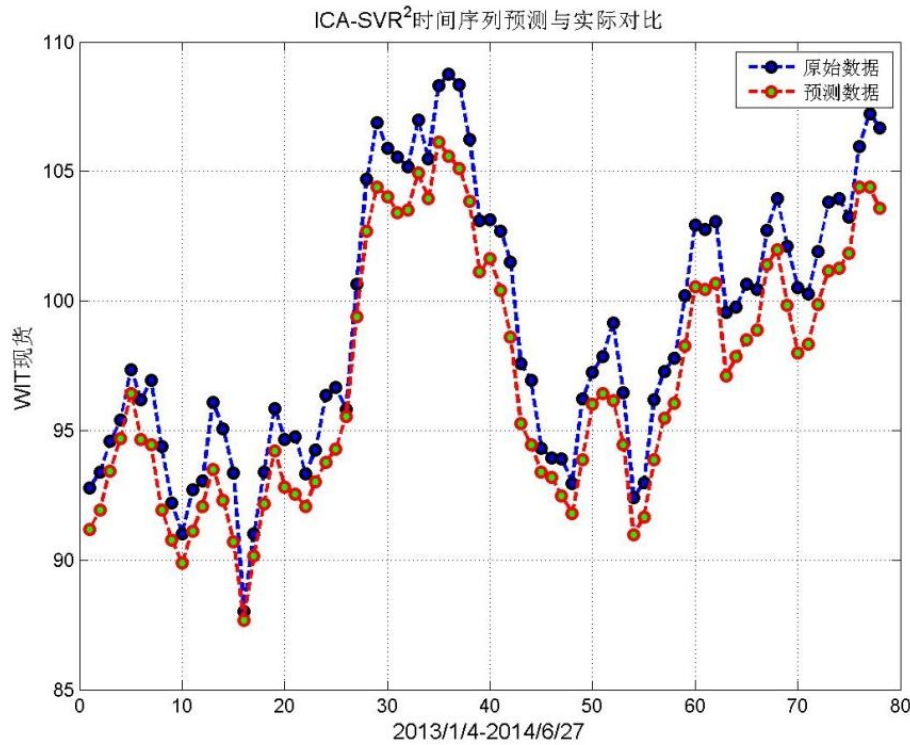


图 5.10 ICA-SVR²对 WTI 原油价格的预测结果(单位：美元/桶)

为了进一步评估 ICA-SVR² 集成预测模型的有效性，度量该集成预测模型的鲁棒性，本文尝试改变训练样本在整个样本区间(2004/7/2~2014/6/27)的比率，对比不同比率下该模型的预测结果。不同比率下，ICA-SVR² 对 WTI 现货价格的预测结果如表所示：

表 5.5 不同比率下 ICA-SVR² 对 WTI 原油价格测试集的预测结果

训练样本比率	模型	RMSE	MAPE	DS(100%)
60%	ICA-SVR ²	3.3425	0.0351	81.21%
70%	ICA-SVR ²	2.1999	0.0289	83.33%
80%	ICA-SVR ²	1.6251	0.0167	88.31%
90%	ICA-SVR ²	1.5271	0.0150	92.31%

实证结果表明，在不同的训练样本与测试样本比例的情况下，ICA-SVR² 始终保持较高预测精度，在训练样本仅占整个样本区间的 60% 的情况下，该模型对 WTI 现货价格的方向预测准确性仍然能保持在 80% 以上。

5.5 预测结果对比分析

为了检验本文方法的预测精度，基于同样的数据集，本文运用 ICA-ARIMA-SVR 模型、单

一的 SVR 模型建立国际油价预测模型，并与 ICA-SVR² 模型的预测结果相对比。ICA-ARIMA-SVR 利用 SPSS 和 Matlab 工具箱进行建模，将历史油价序列进行独立成分分析，将得到的独立成分分别利用滑动平均进行预测，最后通过 SVR 重构得到最终油价预测值。单一的 SVR 模型在 Matlab7.1.0 环境中进行建模，直接对国际历史油价序列进行预测。不同方法的预测结果对比见图 5.11:

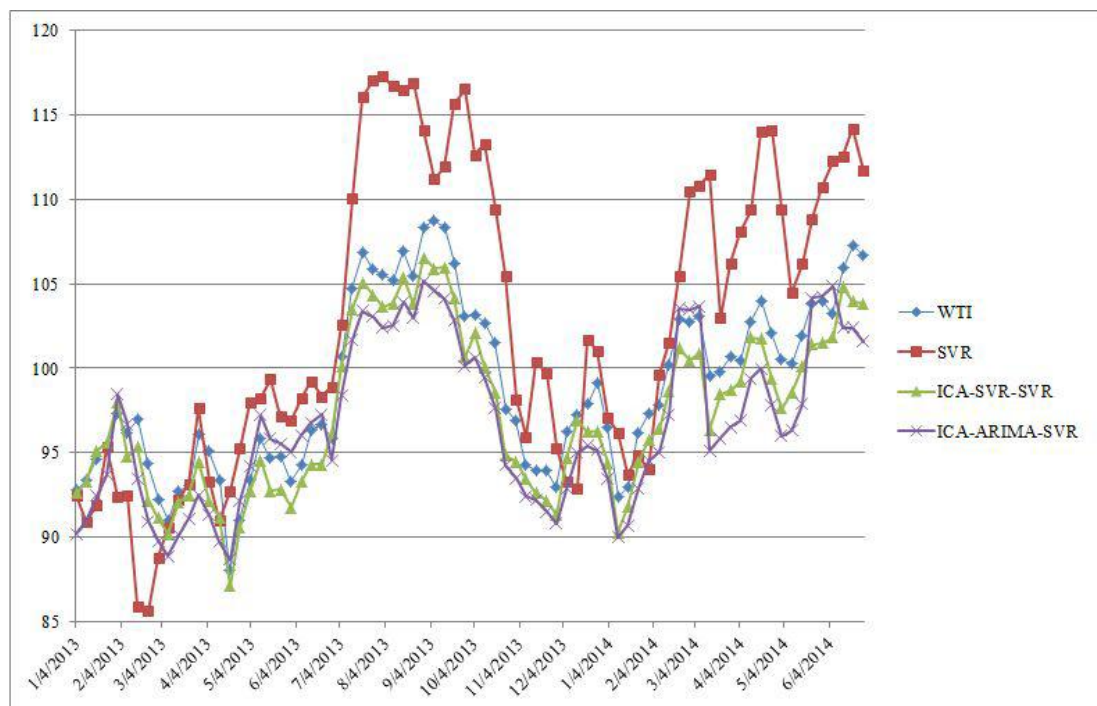


图 5.11 三种方法的原油价格预测结果对比(单位: 美元/桶)

根据图 5.11 对各个模型的预测结果的分析对比可知，本文选用的三种方法所得到的国际原油价格都呈现出与实际值基本一致的总体趋势，但单一的 SVR 模型的预测结果波动较大，偏离了 WTI 原油价格的实际值。基于独立成分分析的两种方法实现了 WTI 原油实际价格的有效拟合。为了进一步对比分析三种方法的预测精度，本文选取 RMSE、MAPE 和 DS 三个评价准则进行有效性度量。对比结果见表 5.6:

表 5.6 三种方法测试集的预测误差比较

	RMSE	MAPE	DS(100%)
SVR	3.9394	0.0537	64.94%
ICA-ARIMA-SVR	1.8855	0.0189	87.01%
ICA-SVR²	1.6251	0.0167	88.31%

对比分析三种模型的预测结果，可以看出，从预测精度而言：基于独立成分分析的两类集成模型明显高于单一的 SVR 模型；而较之于与线性时间序列模型相结合，独立成分分析与非线性 SVR 模型的重构提高了 13.8% 的预测精度；从对于原油商品投资具有重要意义的方向准确性来考虑，ICA-SVR² 模型能达到 88% 的优度，较之对比模型也有一定程度的提高。

对比分析本文选用的三种预测模型的最终效果,可以看出,独立成分分析方法首次被用作原油价格数据信号的预处理技术,对提高模型对于油价分析预测的效果起到重要的作用,它可以分解出国际油价隐藏驱动因素,更直观的反映油价的变化规律和层次特性。同时,引入独立成分分析可以提高预测模型的解释能力,对揭示国际油价的波动规律具有明显优势。而非线性支持向量回归模型可以很好的捕捉国际原油市场价格的非线性特点,较好的实现国际油价走势的预测。

5.6 本章小结

本章在独立成分分析的基础上建立了 WTI 原油现货价格支持向量回归模型,分解得到有具体经济含义的独立成分,并预测国际油价未来走势,取得了较好的预测效果。首先运用独立成分分析方法从非平稳的国际油价时间序列中提取出 3 个独立成分,并对这些代表不同经济含义的独立成分序列进行分析,找到隐藏在原始数据中的驱动因素。接着针对不同波动规律的独立成分建立不同的 SVR 油价预测模型,实现代表不同经济含义的时间序列数据特征的准确刻画,以期获得更为合理的解释预测结果。SVR 预测模型选用 RBF 核函数,利用交叉验证和网格搜索法相结合的方法进行参数寻优,通过样本数据的不断学习训练,最终选取测试集拟合精度最大化参数对对应的 SVR 模型进行独立成分的预测,最终对三项独立成分的预测值进行 SVR 重构,预测 WTI 现货价格的未来走势。预测均方根误差仅为 1.62。通过改变训练样本在整个样本区间的比率,验证了本文构建的 ICA-SVR² 集成模型拥有较好的鲁棒性,实证分析结果证明在训练集仅占样本区间 60%的情形下,该模型对 WTI 现货价格的方向预测准确性仍然能保持在 80% 以上。最后对比分析本文构建的集成模型与对比模型的预测结果,对比结果表明,该集成模型的精度较之 ICA-ARIMA-SVR 和单一 SVR 模型分别提高了 13.8% 和 58.75%,证明了 ICA-SVR² 模型在分析并预测国际油价走势方面的有效性。在未来分析预测世界原油价格的工作中,可尝试应用本文所构建的基于独立成分分析的支持向量回归模型准备判断油价趋势,为决策者提供重要的参考信息。

第六章 结论与展望

6.1 研究结论

本文以国际原油市场为研究对象，实现了独立成分分析和支持向量回归模型的有机结合，较为准确的把握和判断了国际基准油价 WTI 原油现货价格的走势。本文对于国际原油市场价格的研究主要围绕以下 4 个方面详细展开：

1) 总结梳理国际原油价格的基本影响因素。根据原油的商品属性和金融属性，将影响国际油价的众多影响因素归纳为三大类：基本供需关系、经济金融因素以及突发事件。其中供需基本因素作为油价波动的基础条件，反映了原油的商品属性，原油库存是解决供需不平衡的稳定器，而石油的金融属性日趋显著，国际石油期货市场日益显著的地位，致使经济金融因素成为造成世界油价短期频繁波动的主要原因。而包含重大突发事件、地缘政治冲突和突发天气在内的突发事件，对国际原油市场产生中短期的影响。

2) 对比分析各类国际油价预测方法。按照不同的建模依据，国内外油价分析预测方法可划分为国际油价分析预测理论方法、计量经济学预测方法和数据驱动型方法。油价分析预测理论方法侧重于对国际油价走势的理论分析，但无法给出合理的定量分析，缺乏充分的现实意义；计量经济学模型有利于定性研究国际油价与各类影响因素之间的关系以及影响的程度，但是如何实现计量方程中变量的有效选取，是众多学者不断研究的难题；数据驱动型方法强调可以完全依赖过去值推断将来值，包括传统的时间序列方法和人工智能模型在内的数据驱动型方法可以较好的预测油价未来走势，但其缺乏合理的经济解释。

3) 在详细阐述独立成分分析和支持向量回归模型基本理论和适用性的基础上，建立基于独立成分分析和支持向量回归的国际油价预测集成模型 ICA-SVR²。首先根据特征值确定独立成分个数，使用 Fast-ICA 将 4 列国际原油价分解成重大事件项、短期市场波动项、趋势项 3 个独立成分，并针对独立成分分别建立 SVR 预测模型，SVR 模型中选取 RBF 核函数，参数寻优采用交叉验证法和网格搜索法。最后以估计得到的三组预测值为基础，构建 SVR 模型重组并处理后获得油价的最终预测值。

4) 评价构建的 ICA-SVR² 的有效性：通过改变 ICA-SVR² 模型中训练样本占据整个样本空间的比率，从而进一步分析该模型的鲁棒性。实证结果表明，该模型在训练集比例仅为 60% 的情况下，对国际油价的方向正确性的判断仍能达到 80%。将 ICA-SVR² 模型的油价预测结果同 ICA-ARIMA-SVR 和单一的 SVR 模型进行对比，从 RMSE、MAPE 和 DS 三个角度评价三种方法的性能。实证结果显示，基于独立成分分析方法的两种集成模型无论是在预测精度还是在方向准确性判断上均远远高于单一 SVR 模型；而较之与线性 ARIMA 相结合，独立成分分析方法

与支持向量回归方法提高了 13.8% 的预测精度，对油价走势的方向判断上也有略微的提升。ICA-SVR² 有效的提升了预测结果的可解释性，克服了传统数据驱动型方法缺乏经济解释的弱点，实现了国际油价趋势的准确预测，证明了 ICA-SVR² 预测模型具有预测精度高、学习与泛化能力强的特点。

6.2 不足和展望

本文对国际原油价格分析预测进行了一定的讨论，取得了一定的成果，鉴于国际原油市场本身的复杂性，这篇文章仍然存在一些问题需要进一步深化：

1)核函数和参数的选择问题：本文在构建 SVR 模型的过程中，均选用 RBF 函数，没有与其他类型的核函数进行预测结果对比。目前尚未存在统一的确定最优参数的方法，本研究应用经验法，确定足够大的参数搜索范围，带来了明显的计算问题，大量重复运算大大加大了建模时间。如何正确设置推广误差的界以及如何提高参数空间搜索的效率等工作还需要进一步展开。

2)国际原油价格多步预测问题：本研究中仅仅给出了原油价格的单步预测，如何对原油价格进行有效的多步预测问题，仍然需要进一步的研究。

参考文献

- [1] 韩亮亮, 周德群, 李宏伟. 石油价格波动与中国宏观经济运行关系的协整分析[J]. 价格理论与实践, 2006(6):57~58.
- [2] 郁琴. 世界石油价格波动及其形成机制的研究[D], [硕士学位论文]. 北京: 对外经济贸易大学, 2004.
- [3] 程伟力. 影响国际石油价格因素的定量分析[J]. 国际石油经济, 2005(8): 40~43.
- [4] 周海宝. 国际原油价格的影响因素分析及其预测研究[D], [硕士学位论文]. 南京: 南京理工大学, 2012.
- [5] 温渤, 李大伟, 汪寿阳. 国际石油期货价格预测及风险度量研究[M]. 北京: 科学出版社, 2009: 22~25.
- [6] 梅孝峰. 国际市场油价波动分析[N]. 北京: 北京大学中国经济研究中心研究报告, 2001: 24~26.
- [7] 杜光年. 石油价格的波动分析与预测方法研究[D], [硕士学位论文]. 成都: 西南石油大学, 2006.
- [8] 张珣, 汪寿阳. DAC 方法论及其在国际原油价格波动分析与预测中的应用[M]. 北京: 科学出版社, 2010:9~11.
- [9] 张殉, 余乐安, 黎建, 汪寿阳. 重大突发事件对原油价格的影响[J]. 系统工程理论与实践, 2009, 29(3):3~6.
- [10] Hotelling, H. The economics of exhaustible resources[J]. The Journal of Political Economy, 1931:137~175.
- [11] Deshmukh, S. D, Pliska, S. R. Optimal consumption and exploration of nonrenewable resources under uncertainty[J]. Econometrica: Journal of the Econometric Society, 1980:177~200.
- [12] Arrow K J, Chang S. Optimal pricing, use, and exploration of uncertain natural resource stocks[J]. Journal of Environmental Economics and Management, 1982, 9(1): 1~10.
- [13] Salant S W. Exhaustible resources and industrial structure: A Nash~Cournot approach to the world oil market[J]. The Journal of Political Economy, 1976: 1079~1093.
- [14] Cr ́ner J, Salehi~Isfahani D. The rise and fall of oil prices: a competitive view[J]. Annales d'Economie et de Statistique, 1989: 427~454.
- [15] Cremer J, Weitzman M L. OPEC and the monopoly price of world oil[J]. European Economic Review, 1976, 8(2): 155~164.

- [16] F G, Marquez J. Petroleum price elasticity, income effects, and OPEC's pricing policy[J]. The Energy Journal, 1984: 115~128.
- [17] Gately D. OPEC: Retrospective and prospects 1973–1990[J]. European Economic Review, 1983, 21(3): 313~331.
- [18] Solow R M, Wan F Y. Extraction costs in the theory of exhaustible resources[J]. The Bell Journal of Economics, 1976: 359~370.
- [19] Zamani M. An econometrics forecasting model of short term oil spot price[C].6th IAEE European Conference. 2004.
- [20] Dees S, Karadeloglou P, Kaufmann R K, et al. Modelling the world oil market: Assessment of a quarterly econometric model[J]. Energy Policy, 2007, 35(1): 178~191.
- [21] Ye M, Zyren J, Shore J. Forecasting crude oil spot price using OECD petroleum inventory levels[J]. International Advances in Economic Research, 2002, 8(4): 324~333.
- [22] Ye M, Zyren J, Shore J. Forecasting short-run crude oil price using high-and low-inventory variables[J]. Energy Policy, 2006, 34(17): 2736~2743.
- [23] Zhang X, Lai K K, Wang S Y. A new approach for crude oil price analysis based on empirical mode decomposition[J]. Energy Economics, 2008, 30(3): 905~918.
- [24] Zhang X, Yu L, Wang S, et al. Estimating the impact of extreme events on crude oil price: An EMD-based event analysis method[J]. Energy Economics, 2009, 31(5): 768~778.
- [25] Cifarelli G, Paladino G. Oil price dynamics and speculation: a multivariate financial approach[J]. Energy Economics, 2010, 32(2): 363~372.
- [26] Sanders D R, Boris K, Manfredo M. Hedgers, funds, and small speculators in the energy futures markets: an analysis of the CFTC's Commitments of Traders reports[J]. Energy Economics, 2004, 26(3): 425~445.
- [27] Zhang Y J. Speculative trading and WTI crude oil futures price movement: An empirical analysis[J]. Applied Energy, 2013, 107: 394~402.
- [28] Chaudhuri K, Daniel B C. Long-run equilibrium real exchange rates and oil prices[J]. Economics Letters, 1998, 58(2): 231~238.
- [29] Zhang Y J, Fan Y, Tsai H T, et al. Spillover effect of US dollar exchange rate on oil prices[J]. Journal of Policy Modeling, 2008, 30(6): 973~991.
- [30] Quan J. Two - step testing procedure for price discovery role of futures prices[J]. Journal of Futures Markets, 1992, 12(2): 139~149.
- [31] Gülen S G. Efficiency in the crude oil futures market[J]. Journal of Energy Finance &

- Development, 1998, 3(1): 13~21.
- [32] Abosedra S, Baghestani H. On the predictive accuracy of crude oil futures prices[J]. Energy policy, 2004, 32(12): 1389~1393.
- [33] David Cabedo J, Moya I. Estimating oil price 'Value at Risk' using the historical simulation approach[J]. Energy Economics, 2003, 25(3): 239~253.
- [34] Morana C. A semiparametric approach to short-term oil price forecasting[J]. Energy Economics, 2001, 23(3): 325~338.
- [35] 陈益芬, 徐诺. 国际原油价格市场波动的模型分析[J]. 甘肃联合大学学报: 自然科学版, 2010, 24(2): 36~39.
- [36] Abramson B, Finizza A. Using belief networks to forecast oil prices[J]. International Journal of Forecasting, 1991, 7(3): 299~315.
- [37] Kaboudan M A. Compumetric forecasting of crude oil prices[C]. Evolutionary Computation, 2001. Proceedings of the 2001 Congress on. IEEE, 2001, 1: 283~287.
- [38] Mirmirani S, Li H C. A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil[J]. Advances in Econometrics, 2005, 19: 203~223.
- [39] Radchenko S. The long-run forecasting of energy prices using the model of shifting trend[J]. University of North Carolina at Charlotte, Working Paper, 2005.。
- [40] Xie W, Yu L, Xu S, et al. A new method for crude oil price forecasting based on support vector machines[M]. Computational Science-ICCS 2006. Springer Berlin Heidelberg, 2006: 444~451.
- [41] Fan Y, Liang Q, Wei Y M. A generalized pattern matching approach for multi-step prediction of crude oil price[J]. Energy Economics, 2008, 30(3): 889~904.
- [42] LAI K K. Crude oil price forecasting with TEI@ I methodology[J]. 系统科学与复杂性: 英文版, 2005, 18(2): 145~166.
- [43] 李彬. 国际市场原油价格的非线性组合预测研究[D], [硕士学位论文]. 天津: 天津大学, 2006.
- [44] 吴虹, 尹华. ARIMA 与 SVM 组合模型的石油价格预测[J]. 计算机仿真, 2010 (5): 264~266.
- [45] 梁强, 范英, 魏一鸣. 基于小波分析的石油价格长期趋势预测方法及其实证研究[J]. 中国管理科学, 2005, 13(1): 30~36.
- [46] Yu L, Wang S, Lai K K. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm[J]. Energy Economics, 2008, 30(5): 2623~2635.
- [47] Hyvarinen A, Karhunen J, Oja E. 独立成分分析(周宗潭译)[M]. 北京: 电子科学出版社, 2007.
- [48] Karhunen J, Hyvarinen A, Vig ário R, et al. Applications of neural blind separation to signal and

- image processing[C].Acoustics, Speech, and Signal Processing, 1997. ICASSP~97, 1997 IEEE International Conference on. IEEE, 1997, 1: 131~134.
- [49] Cardoso J F, Laheld B H. Equivariant adaptive source separation[J]. Signal Processing, IEEE Transactions on, 1996, 44(12): 3017~3030.
- [50] Vigário R, Jousmäki V, Haemaelaeninen M, et al. Independent component analysis for identification of artifacts in magnetoencephalographic recordings[J]. Advances in neural information processing systems, 1998: 229~235.
- [51] Torkkola K. Blind separation for audio signals—are we there yet[C]. Proc. Workshop on Independent Component Analysis and Blind Signal Separation. 1999: 11~15.
- [52] 胡月. 基于主成分分析和独立成分分析的人脸识别[D], [硕士学位论文]. 吉林大学, 2010.
- [53] Bell A J, Sejnowski T J. The “independent components” of natural scenes are edge filters[J]. Vision research, 1997, 37(23): 3327~3338.
- [54] Jutten C, Herault J. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture[J]. Signal processing, 1991, 24(1): 1~10.
- [55] Comon P. Independent component analysis, a new concept?[J]. Signal processing, 1994, 36(3): 287~314.
- [56] Kiviluoto K, Oja E. Independent Component Analysis for Parallel Financial Time Series[C]. ICONIP. 1998, 2: 895~898.
- [57] Malaroui S, Kiviluoto K, Oja E. Time series prediction with independent component analysis[C]. Proceedings of international conference on advanced investment technology. 2000: 895~898.
- [58] Lu C J, Lee T S, Chiu C C. Financial time series forecasting using independent component analysis and support vector regression[J]. Decision Support Systems, 2009, 47(2): 115~125.
- [59] Valmicki V. 统计学习理论的本质 (张学工译) [M]. 北京:清华大学出版社, 2000.
- [60] Gunn S R. Support vector machines for classification and regression[J]. ISIS technical report, 1998, 14.
- [61] 孔繁钰, 徐瑞华, 姚胜永. 交通量的支持向量回归预测及参数选择研究[J]. 计算机工程, 2007, 33(05): 20~22.
- [62] Fernandez V. Forecasting commodity prices by classification methods: The cases of crude oil and natural gas spot prices[J].
- [63] 祝金荣. 基于支持向量机的石油期货价格预测[J]. 工业技术经济, 2007, 26(2): 59~61.
- [64] 刘立霞, 马军海. 基于 LS-SVM 的石油期货价格预测研究[J]. Computer Engineering and Applications, 2008, 44(32).

致 谢

首先，对于本人导师周鹏教授以及师母范丽伟教授表达真诚的谢意。为期两年半研究生生涯，针对于学习及生活的各个方面，周老师和范老师均给予本人极为精心的指导。在整个论文的完成过程中，两位老师对于选题、开题以及修改、定稿等各个环节都给予很多指导和建议。针对于生活工作方面，两位老师也给出诸多指引，我将谨记周老师和范老师的教诲，努力做人做事，成为有社会价值的大好青年。

同时，我要对南航能源软科学研究中心的全体成员们表达深深的谢意。每周三的研讨会都是自我学习和提升的良好契机。感谢周德群教授、张力菠副教授、张钦副教授、章玲副教授和查东兰副教授等老师对论文研究工作提出的宝贵意见和建议。感谢张明明师兄和丁浩师弟对每次会议的组织安排，给我们提供了学习交流的平台。

此外，父母、亲人和朋友在研究生期间对我的支持和帮助致以深深的谢意。最后，向各位在百忙之中审阅此论文的专家和教授致以最真诚的谢意！

作者在攻读学位期间发表的学术论文目录

1. Hui-ping Li, P. Zhou, Li-wei Fan. *What Are the Driving Factors behind the Fluctuation of Crude Oil Prices?: An Independent Component Analysis*. **Advanced Materials Research** (已录用, 第一作者)
2. P. Zhou, Hui-ping Li, Li-wei Fan. *Volatility Analysis and Forecasting Models of Crude Oil Prices: A review*. *International Journal of Global Energy Issues* (已接受, 第二作者)
3. 李慧萍, 周鹏, 范丽伟。“原油期货价格影响因素研究”, 第七届中国能源资源开发利用战略学术研讨会暨第四届能源经济与管理学术年会, 2013(分组报告)