

中图分类号：

密级：公 开

学科分类号：

论文编号：37\_025200\_31045632016300159\_LW

山東財經大學

硕 士 学 位 论 文

（专业学位）

黄金价格动态预测和影响因素研究

作者姓名：刘莉

学科专业：应用统计

指导教师：赵霞（教授）

合作导师：

培养学院：统计学院

二〇一六年五月

# **Dynamic Prediction and Influence Factors Analysis on the Price of Gold**

A Dissertation Submitted for the Degree of Master

**Candidate: Li Liu**

**Supervisor: Prof. Xia Zhao**

School of Statistics

Shandong University of Finance and Economics

中图分类号：

密级：公 开

学科分类号：

论文编号：37\_025200\_31045632016300159\_LW

## 硕 士 学 位 论 文

# 黄金价格动态预测和影响因素研究

作 者 姓 名： 刘莉

申请学位级别：经济学硕士

指导教师姓名： 赵霞

职 称：教授

学 科 专 业： 应用统计

研 究 方 向：风险管理与精算

学 习 时 间： 自 2014 年 9 月 1 日 起 至 2016 年 6 月 30 日 止

学位授予单位： 山东财经大学

学位授予日期： 2016 年 6 月

## 山东财经大学学位论文独创性声明

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

学位论文作者签名：

日期： 年 月 日

## 山东财经大学学位论文使用授权声明

本人完全同意山东财经大学有权使用本学位论文（包括但不限于其印刷版和电子版），使用方式包括但不限于：保留学位论文，按规定向国家有关部门（机构）送交学位论文，以学术交流为目的赠送和交换学位论文，允许学位论文被查阅、借阅和复印，将学位论文的全部或部分内容编入有关数据库进行检索，采用影印、缩印或其他复制手段保存学位论文。

保密学位论文在解密后的使用授权同上。

学位论文作者签名：

日期： 年 月 日

指导教师签名：

日期： 年 月 日

## 摘 要

黄金是一种具有金融属性的产品，是商品市场上较为重要的现货、期货交易产品之一。在 2008 年的世界经济危机中，各种金融产品都表现出疲软的现状，而只有黄金的市场具有突出的表现，近期在中国的股市大幅度波动中，黄金价格变化也不是很大。基于大数据分析方法，探究黄金价格变化规律，无论对投资者、生产者还是政府管理者，都有积极的意义。

本篇文章结构分为五个部分，分别从三个方面对黄金价格进行了研究。在对行文背景、研究思路、内容以及方法分析的基础上，首先利用协整突跃点检测模型、箱线图等方法对黄金价格收益率数据进行了检测，查出了所选数据区间内较异常点所对应的日期；其次分别利用去除异常点和不去异常点的数据进行 ARIMA-GARCH 模型的建立，研究表明去掉异常点之后所建模型的显著性有所提高，接着利用该模型对去异常点之后的黄金价格进行了波动性分析与预测，又利用数据挖掘技术中的支持向量机模型对相同的黄金价格数据进行预测，并将预测结果进行对比，发现支持向量机的预测效果比时间序列模型的预测效果更精确；再次，在对黄金价格进行波动性分析的基础上，研究了黄金价格的一些影响因素，并且引入了两个新的变量，在后续的分析中，我们利用不同的解释变量建立了不同的 VAR 模型。研究发现，就模型的整体效果来说，引入新变量后效果变好了。另外，我们还分别利用月度数据和日数据建立了 VAR 模型并提出了对比结果。最后根据一系列的分析给出了本文的总体结论与建议。

经过理论与实证的分析，结合各种类型的对比分析，对黄金价格的各方面进行了较全面的分析，通过研究文献发现对黄金价格的研究已经从多方面进行了开展，比如波动性分析、影响因素分析等等，但是本文是对中国的黄金价格进行的分析，所以由于中国国情的特殊性，本论文的分析与结论都还需要进一步完善。

**关键词：**黄金价格；波动率；异常点检测；支持向量机；VAR 模型

## Abstract

Gold is one kind of product with financial attributes, and it is one of the important spot and stock contract products in the commodity market. In the world economic crisis in 2008, a variety of financial products have showed weak in the present situation. But only the gold market has outstanding performance. Recently the stock market of Chinese has a substantial fluctuations, but the price of gold has small fluctuation. Based on the data analysis method to explore the rule of gold price's change has positive significance, whatever for Investors、producers or government regulators.

The structure of this article is divided into five parts and studied from different aspects. On the basis of the research for writing background、research ideas、content and method. First, for the yield on the gold price, the article has carried on the testing using the model of cointegration outlier detection、boxplot and so on. And found out the date corresponding to the abnormal points. Secondly, Build the model of ARIMA-GARCH, respectively, using removed and unremoved abnormal point data. The results showed that the model with the removed data has a higher significance. Then we have used the model to analyze the volatility of gold price and carried on a forecast. In addition, the article using support vector machine (SVM) predicted the price of gold in the technology of data mining. Finally we compared the two different results and found that the model of svm has a better prediction effect. Thirdly, on the basis of the analysis for gold price's volatility, we found it is meaningful that to study the factors which influencing the price of gold. We studied the previous literature and summarized some influencing factors. Then we recommend two new factors with users and media index. In follow-up studies, we has build different models of VAR with the different interpretation variables. The study found that the overall effect of model became more better after we recommend new factors. In addition, we also analyzed the different model respectively with monthly and daily data, and then put forward the results of comparison. The last, the conclusions and recommendations of this article was presented.

Through the oretical and empirical analysis and all sorts of comparative analysis, we

have carried on the comprehensive analysis from aspects of the gold price. By studying the literature found that the study of the price of gold has been carried out from many aspects. Such as volatility, influence factor analysis and so on. But this article is the analysis of China's gold price. So because of the particularity of China's national conditions, the analysis and conclusion of this paper are still need further perfect.

**Key words:** The Price of Gold; Volatility; Anomaly Detection; Support Vector Machines; Vector Auto-regressive Model

## 目录

第 1 章	绪论	1
1.1	研究背景和意义	1
1.2	文献综述	3
1.3	研究内容	8
1.4	研究思路和研究方法	9
1.5	创新点和不足	10
第 2 章	模型介绍	12
2.1	GARCH 族模型的介绍	12
2.2	K-means 聚类算法的介绍	14
2.3	支持向量机模型的介绍	14
2.4	ARIMA 模型介绍	15
2.5	VAR 回归模型介绍	16
2.6	本章小结	16
第 3 章	黄金价格动态预测分析	18
3.1	黄金价格的异常点检测	18
3.2	基于 ARIMA-GARCH 族模型的黄金价格预测分析	25
3.3	基于支持向量机的黄金价格预测分析	33
3.4	结果对比分析	34
3.5	本章小结	38
第 4 章	黄金价格影响因素分析	40
4.1	变量选择	40
4.2	数据选取与描述分析	42
4.3	基于 VAR 模型的黄金价格影响因素分析	44
4.3.1	基于月度数据的分析	44
4.3.2	基于日数据的分析	47
4.3.3	对比分析	49
4.4	本章小结	51



第 5 章 结论和建议 .....	51
5.1 结论 .....	52
5.2 建议 .....	53
参考文献 .....	55
致谢 .....	57

## 第 1 章 绪论

这一章的主要内容就是对论文的行文结构进行一个整体的把握，从社会的大环境以及对研究黄金价格的一些文献进行分析的基础上，提出了论文的研究内容与研究思路，并且对论文的一些创新点进行了简单的表述。在分析黄金价格的经济大形势时发现，正因为黄金价格在实际中波动比较大，才显得本文的分析格外的具有理论与实践意义。通过对黄金价格突变点检测以及其影响变量的研究，进而利用差异化的方法对黄金价格进行预测，不但能够规范一下中国的黄金市场，促进其健康发展，而且进一步说还可以完善我们市场的一些制度，使得制度建设的更加利民。另一方面说，不论对我们的国家来说还是对我们个人来说都会有一些利用价值的。根据以上所提意义，这一章对论文的研究内容进行了总结概括，提出了论文创新的地方和论文在行文过程中存在的一些难点与不足。

### 1.1 研究背景和意义

#### 1.1.1 选题背景

黄金是商品市场上较为重要的现货、期货交易产品之一，在美元与黄金脱钩之后，黄金价格开始步入市场化的轨道，黄金价格开始出现越来越激烈的大幅波动。在过去的几年里，黄金价格整体上呈现不断上涨的趋向。2008 年发生的经济危机之中，许多金融资产都表现出异常，只有黄金交易市场不然，并且近期在中国的股市大幅度波动中，其黄金价格并没有表现出特别明显的波动。所以，对于黄金价格影响因素的研究还是很有价值的。

通过研读文章中黄金价格这许多年来的波动趋向，这些趋向表明黄金价格的波动幅度是我们难以想象到的。众所周知黄金具有货币的职能，这一职能让其价格一直比较稳定，然而，随着社会的发展，人类文明的进步，生活水平的提高，黄金的价格也开始出现了大幅的波动。具体数值如表 1-1 所示：

表 1-1 世界黄金价格变化

	20 世纪 70 年代初	80 年代初	90 年代初	本世纪初
黄金价格（多美元/每盎司）	30	约 700	370	270

其中，历年来比较的话，黄金价格最高出现在 2008 年的 2 月份，价格达到了约每盎司 1005 美元，接着又呈现大幅度下降，只有月余的时间价格就下降到约每盎司

845 美元，下降幅度约 15.9%。具体看细节来说国家黄金价格时而上升时而下降，不断交换，然而整体来看，黄金价格还是一路上升的，其中上升幅度最大的当属 2011 年，竟然长到了每盎司 1911 美元。上升速度如此之快，让人吃惊的同时，也引起了大家的研究热情，来纷纷分析出现这种现象是出于何因。

另外，由于黄金不仅具有货币职能，它还是一种金融产品，黄金的投资者以及黄金的生产者对于黄金的价值取向可以直接由它的价格变化所决定，所以利用它具体的运转程式中所生成的价格值中，掌握整体的经济运转规律或分析验证已存在的成熟的理论知识、说明大家普遍默许的经济现象，有助于投资黄金和生产黄金的人更好的了解黄金市场，本文是想要通过一些模型预测黄金市场的行情，并为投资者和生产者提供决策方面的帮助。

### 1.1.2 研究意义

就现在的一些研究而言大部分研究都是针对国际黄金的，影响因素研究以及动态预测研究都有。国际黄金的市场发展的比较早、比较完善是其中很重要的一个原因，重点来说，对黄金价格影响较大的一般是那些西方的比较的发达国家。又由于黄金的比较特殊的定价机制导致了大多数的黄金价格波动分析的文献都是针对国际黄金的，多数文献分析的黄金价格都是基于伦敦黄金市场协会中黄金的下午定盘价数据进行的，而伦敦金银市场的定价机制分为上午定盘价和下午定盘价，此基准价每天更新两次，最终的解释权主要归其定价的银行。但我们行文所分析数据一般采用下午的定盘价。然而，从历年黄金价格一路飙升的特点出发，其背后所反映的问题确是经济的传统与新兴体制之间有关货币定价权的一种博弈。就黄金的产量来说，07 年以来我国已连续四年全球第一，而且不仅是最大的生产国也是最大的消费国。

黄金是一种具有金融属性的产品，是商品市场上较为重要的现货、期货交易产品之一，其价格变化与国内外金融形势、投资者、生产者的价值行为都有密切联系。在 2008 年发生的经济危机之中，许多金融资产都表现出异常，只有黄金交易市场不然，并且近期在中国的股市大幅度波动中，其黄金价格并没有表现出特别明显的波动。基于大数据分析，探究黄金价格变化规律，无论对投资者、生产者还是政府管理者，都有积极的意义。

## 1.2 文献综述

本节通过对国内和国际文献的比较分析发现,目前关于黄金价格的研究,主要从两个角度展开,一是黄金价格的趋势动态预测,二是影响黄金价格的因素分析。本章先对中国黄金价格的趋势动态预测进行了介绍。首先分析了国内以及国际上一些学者对黄金价格的研究思路与方法,让作者对黄金目前的分析方向和分析进程大致上有了了解和认识。其次,对国际上以及国内黄金价格动态预测的分析结果进行研究和汇总,显示出对黄金价格动态预测研究时所用的各类方法,又分析和汇总了所观察学者的文献。接着,下一小节分析研究了黄金价格的有关影响因素的文章,并对国际上以及国内的两方面的相关研究文献进行了总结和研究。最终提出了四个目前研究者们为涉及的研究方向以及存在的一些可以丰富的地方。

### 1.2.1 黄金价格动态预测研究

目前,对黄金价格的定量化预测模型的研究中,国内外都积累了大量的文献,同时也扩展出了多样的预测模型,其中论文中常见的有网络神经元模型分析法、协整模型分析法、资产组合分析方法、时间序列分析法、传统回归分析法、灰色模型(Grey Model)分析法、GARCH 族模型分析方法等。

国内关于黄金价格的预测有许多研究:潘贵豪等(2010)用 1971 年 1 月份到 2008 年 12 月份期间的伦敦黄金交易市场收盘价格为基础,分析了时间序列的相关理论的基础上建立了黄金价格的 ARIMA-GARCH 模型,形象的刻画了黄金价格数据的生成过程,发现平均误差很小,模型能够很好的拟合数据并且预测黄金市场的行情<sup>[1]</sup>。

之后曾黎、李春(2013)引入数据挖掘技术,根据 ARIMA 模型和数据挖掘技术中的 NN 模型,预测黄金价格,发现预测的精确度高于之前的单一模型。但是此文章没有考虑对黄金价格时间序列数据的分解问题<sup>[2]</sup>。赵庆、王志强(2014)发现了预测的一种新思路:先是将时间序列数据用 HP 滤波分解两部分,分别是带有趋势的部分和带有周期性的部分;接着根据两序列不同的性质,在趋势部分建立 AR 模型进行预测分析,在具有周期性的部分建立 ARMA-GARCH 族动态进行预测分析;最后利用两模型预测的综合结果与开始数据进行对比;发现预测精度明显有所提高<sup>[3]</sup>。

魏宇等(2010)对 Au99.95 的日数据进行了分析,分别建立了不同的模型并且在不一样的背景分布下进行了 VaR 预测,这种预测方法是属于外动态的。紧接着运用更

加严格和稳健性比较高的 LR 检验和动态分位数回归检验法, 对上面各模型得到的 VaR 预测精度进行了对比和后验分析。分析结果表明, 就我国而言黄金的直接交易中, 价格波动所具有的杠杆效应并不明显, 但数据本身却表现除了显著的有偏性和尖峰厚尾性。且看出就 VaR 预测精度而言, 较繁杂的 GARCH 族模型没有像想象中比 GARCH 模型更精确<sup>[4]</sup>。曾濂等(2010)利用 BP 神经网络的改善模型, 对黄金价格进行了预测<sup>[5]</sup>。

孙晓莹等(2012)发现在分析股价变化的预测问题时, 因为很多因素都可以影响股票的价格, 使得股票价格具有随机、突变以及非线性的特性, 传统的一个预测模型很难描述股票整体的变化, 也会使得模型的预测精度很低。为了提高股票价格的预测精度, 文中利用了一种组合模型, 即在原始模型中引入数据挖掘技术。根据股票价格自身的特点, 文中首先对股价的线性部分进行了预测建模, 其次利用数据挖掘技术对数据的非线性部分进行了预测建模, 最后把两种不同的预测模型结果结合起来分析, 预测出最终的股票价格。实证分析发现, 与只包含一种模型的股价预测方法相比, 利用组合预测方法进行预测不仅仅能够提高模型的预测精度, 还可以使得模型最终的预测结果误差变小, 所建立的模型也能更好的反映建模对象的特征及其内在的变化规律, 这种股价预测方法无疑是值得推广的<sup>[6]</sup>。

蒋翠清等(2015)在分析社交媒体信息对股票行为的一些时不仅考虑了传统的整体的一些层面, 还将参与社交媒体的人员多样性以及所讨论话题的多样性考虑其中, 她利用社交媒体信息的一些独有的特征比如内容、写作风格等, 根据 EM 聚类技术、主成分分析法、文本特征提取技术等研究社会媒体的参与人以及这些参与人所关心的重点。接着, 该文分别分析这些干系人以及话题, 从不同的方向组建四个有关社会媒体的指标来建立回归模型, 并进行预测。最终, 此文章进行了实证分析, 数据取自雅虎金融论坛的美国银行模块, 对前面所列方法进行了验证<sup>[7]</sup>。

徐海鹏(2010)从数据挖掘中的关联规则算法开始, 基于主、客观的兴趣度约束进行研究, 力求挖掘有关股票收益率的一些消息, 从而达到预测股价的目的。另外将结果与实际的序列进行对比, 发觉关联算法预测的精确度<sup>[8]</sup>。张延利(2012)在文章中将传统的 ARMA 预测模型与马尔科夫模型进行了结合, 使得数据的统计特征与灰色理论进行了很好的结合。其中 ARMA 预测模型部分来模拟序列的线性变化规律, 而马尔科夫状态转移概率的知识用来观察状态转移的动向。研究分析发现, 此组合方程的预测准确度比其中两者中的任意单一模型的预测精度都好<sup>[9]</sup>。

兰秋军等(2005)提出了 TSEOPM, 这是数据挖掘算法中的新思路, 力求找到序列

中拥有预测能力的那部分,并利用我国沪、深 A 股市场近年来收益数据进行了验证,探求了其短期可预测性,研究发现对于征兆模式而言要想获得收益局部模式要比全模式好,并且经过对之前价格的研究在短期内能够获得一些超额的收益,因此可以表明中国的股票市场还不是弱有效的<sup>[10]</sup>。张其联等(2011)通过实证分析发现黄金价格有上升和下降两个状态,二者在转换时是遵循不同的概率的。观察模型结果可以发现,用马尔科夫链知识描述国际现货黄金价格涨跌的特征比经常看到的经典发线性自回归模型要好<sup>[11]</sup>。

放眼国际而言,研究黄金价格预测方法的文章也很多。Shahriar 等(2010)将 ARIMA 模型以及随机游走过程进行了组合,一起分析一些对黄金价格具有影响的因素(比如国际通货膨胀和原油价格等),来预测黄金价格的月收盘价<sup>[12]</sup>。Ismail Z(2009)基于 8 种因素,即 Commodity Research Bureau(CRB)index、M1、NYSE 指数、标准普尔 500 指数、USDIX、国债收益率等因素建立计量经济回归模型来对黄金价格进行预测<sup>[13]</sup>。G Peter Zhang 指出,一个时间序列过程可以看成是由一个线性结构和一个非线性结构两部分组成<sup>[14]</sup>。

### 1.2.2 黄金价格的影响因素研究

对于黄金价格影响因素的分析,国际上以及国内的去多学者也做了一定的分析。根据 Capie 等(2004)的研究结果,黄金价格与主要货币汇率二者中间具有负相关关系<sup>[15]</sup>。Tully and Lucey(2007)分析发现在整体的影响黄金价格的因素中,美元的影响程度是最大的,同时发现黄金能够抵抗其价格的下降<sup>[16]</sup>。Smith(2001)的分析表明了黄金价格的变化走向和一些工业过度股票价格指数具有相反的变动关系<sup>[17]</sup>。Levin 等(2006)通过简单的供求关系理论发现黄金价格与美国物价指数具有长期稳定的正相关性<sup>[18]</sup>。

Colin Lawrence(2003)分析可知,对于投资黄金来说,其投入与产出的比率大小是不同于我们通常所说的经济的循环的,重要的指标(如 GDP 增长率等)和我们所分析的黄金的投入产出比中间在统计意义上没有显著的关系,但黄金回报率与一些金融资产的指数如股票指数、另外的商品如铝、石油、锌等价格之间拥有的相关关系确是显著的<sup>[19]</sup>。Ranson 等(2005)表明黄金价格可以作为判定通货膨胀的变量,也可以作为债券市场的晴雨表,作为长短期名义利率、通货膨胀的强大预报器是绰绰有余的<sup>[20]</sup>。Harmston(1998)也利用 CPI、批发价格指数两个方面对美国、日本、德国和法国的黄

金价格指数之间进行了研究并进行了对比,分析表明,长远来说,是因为黄金继续了其对于消费品以及中间产品的实际购买力才使得其保持了自己的价值。同时还研究了战争时期与非战争时期的区别,发现战争虽然是特殊时期,在此期间物价飞速上涨,但黄金的一些特性比如方便携带性、易流动性等往往比它的可交换性更重要<sup>[21]</sup>。

傅瑜(2004)利用一些简单的统计学描述性分析方法对黄金价格的波动进行了实证研究,他发现黄金价格与许多因素之间都具有负相关的关系,其中对黄金价格影响程度最大的是美元的汇率;另外,黄金价格与通胀率、国际形势等具有正向相关的关系<sup>[22]</sup>。杨柳勇等(2004)对前辈的研究进行了分析,并提出了一种新的计量经济模型,在实证分析时,他采取1973年以来的序列分析黄金价格的长期决定因素,来保证对黄金价格长期性的解释进行一个很好的补充,分析发现在长期的指标里面有美元名义有效汇率、道琼斯价格指数等因素对黄金价格构成影响<sup>[23]</sup>。

在此基础上,刘曙光等(2008)分别研究了三个时间段内对黄金价格具有长期影响的一些因素,研究表明美国联邦基金利率、道琼斯价格指数、美元名义有效汇率、世界黄金储量等都是黄金价格的长期影响因素。但具体到不一样的时期,这些长期因素却各有不同,也就是说对黄金价格具有长期性影响的指标并不是稳定存在的。就长期因素说,美元和黄金的价格波动大概是一样的<sup>[24]</sup>。

之后谢朝阳(2010)又在分析金价的影响因素时加入了政府政策性因素货币供应量M2和美国CPI两个新的变量,其实证分析发现,对黄金价格具有长期影响的因素是美元的实际汇率,并且也是最主要的指标<sup>[25]</sup>。前面所说的所有文献都是分析的影响黄金价格的一些外界的因素,但是除了前面提到黄金价格自身对自身价格也会有一定的影响。邱济洲等(2014)利用VAR模型研究了黄金价格总水平的影响因素,其中创新性的引入了黄金价格自身的影响,研究表明黄金价格本身的一些特性是引起国际上的黄金价格上涨的很重要的一个原因<sup>[26]</sup>。另外,考虑到黄金收益率序列分布具有尖峰厚尾的特性,郑秀田(2008)建立了GARCH-M模型分析研究了我国黄金市场上的收益和风险之间存在的关系。该研究表明:通过黄金价格日度数据所求得的收益率序列存在很大的聚集波动性,对于黄金价格走势的拟合采用GARCH族模型就很好<sup>[27]</sup>。

在此基础上,范为、房四海(2011)综合考虑了黄金的多个属性比如货币职能、避险特质等,把黄金价值分成:主权国家信用违约的风险溢价、基于汇率的“隐性货币价值”、商品基准价值,另外,在定价研究中,其引进了CRB、USDX和CDS利差等指数做代理指标。基于向量自回归(VAR)模型的研究表明:USDX与黄金价格具

有负向相关关系，CDS、CRB 与黄金价格之间具有正向相关关系；其中 CRB 和 USDX 一阶差分平稳，CDS 利差二阶差分平稳且其对黄金价格的影响程度最大。他们在使用 VAR 模型分析黄金价格影响因素的基础上运用 EGARCH 模型来处理金价的尖峰和厚尾现象<sup>[28]</sup>。闵嘉宁(2011)研究了 2008 年经济崩溃以来，在黄金价格一路飙升的背景下，首先研究了黄金交易市场的价格体系组建，分析了黄金价格变化的内部原因以及对其产生影响的一些外部因素，并探究了黄金价格投资的价值和其波动之间具有的相关性，进而提出相关的策略建议<sup>[29]</sup>。

张继德等(2014)如采取研究了上证 T180 股票，利用用户搜索量作为一般投资者对黄金的关注度变量，揭示一般投资人对股票关注度的流动性对于股票收益的影响。研究结果发现，当其他一些变量控制不变时，一般投资黄金的人的关注度和市场的流动性之间具有高度的相关关系，这会推动投资黄金的人更加注重交易市场<sup>[30]</sup>。杨虎等(2004)主要讨论了利用何种方法可以将市场上的时间序列模型转化为相对应的线性模型，接着用线性模型的相关方法去检测异常值，同时想法确定这个异常点的具体性质与特点。如果属于创新型的，那么针对它的分析不但在理论上具有一定的意义，针对具体经济风险分析的现实意义也非常强。最后基于算法进行了实证研究，结果显示文章提到的思路在具体的分析中具有可行性和有效性<sup>[31]</sup>。陈晓(2015)首先研究了货币政策与股票市场之间的影响机制，在此基础上，又研究了二者之间的关系并进行了实证分析，揭示了二者之间存在的计量协整关系，这一结论对于我国的货币政策的调整是有一定作用的<sup>[32]</sup>。

吕书然(2015)以金融经济学界始终重点关注的货币政策和股市之间的关系机制为题进行了研究，首先分析了国际上以及国内众多学者的文献包括现状、方法与取得成果等进行了基础的研究，然后重点分析了股票市场是否受货币政策改革的影响，以及怎么影响的问题，第三点是我国沪深两市对于货币政策变动的反应以及股票市场波动对货币政策带来的调整四个方面探析了货币政策变动与股票市场变动二者的关系问题，希望能够提高货币政策制定时的水平，以便使得通过货币政策的调整，从而更好的为我国股票市场和实体经济的健康发展提高裨益<sup>[33]</sup>。王相宁等(2015)建立半参数时变 Copula 模型，检验人民币汇率与黄金价格之间的变动机制是怎样的，根据二者之间的  $r$  的变化趋势，研究黄金与汇率二者间的关系以及影响。结果表明：黄金价格与间接标价法下的人民币汇率之间是正相关的关系，并且其具有的尾部相依结构也是对称的；在两者相关性的影响方面，人民币汇率改革的影响也是最为显著的，其次是



次贷、欧债危机以及国际主要货币的汇率政策<sup>[34]</sup>。

周云丽等(2014)在探讨黄金价格的影响因素的时候,引入了美元汇率、道琼斯指数等变量利用分位数回归模型进行了实证分析。研究结果显示采取不一样的分位数研究时,各因素对黄金价格的影响程度是不同的。我们知道分位数回归模型能够从黄金价格的以前数据中发现很多有用的信息,这也更有利于黄金投资者在投资的过程中做出更合理更有利的决策<sup>[35]</sup>。

### 1.2.3 总体评价

综合上述文献来说,学者们从宏观经济变量、金融资产的回报率、金融指数等各种影响因素以及模型的改进等方面对黄金价格做了大量研究,但是仍存在一些值得进一步研究的内容。

(1) 目前关于黄金价格影响因素的分析,基本上都是立足于常见的、易于想到的一些因素,比如汇率、利率等,但是考虑到现在的大数据背景以及国内外的大形势,互联网科技对投资者和媒体的信息接受以及根据所有信息做出的决策产生了颠覆性的影响,所以本文在对黄金价格的影响因素进行分析的时候,考虑引入了基于百度指数的用户指数和媒体指数因素。

(2) 目前关于黄金价格预测方面的研究,所建模型很单一,一般都是运用时间序列中的 ARIMA 模型,即使有提到数据挖掘算法的,方法也仅限于神经网络算法。所以我们基于此考虑使用数据挖掘技术中的 SVM 来预测黄金价格。

(3) 纵观以往文献我们发现,对于黄金价格波动中的异常点的检测问题鲜少有一篇文章提到,只见周舞舞(2013)利用协整突跃点检测模型寻找了价格的异常点<sup>[36]</sup>。所以在异常点检测的方法应用方面还有很多可以丰富的地方。

所以本文将从几个方面对黄金价格原始数据以及其收益率数据进行分析研究。具体内容如下 1.3 节所示。

## 1.3 研究内容

本文在分析的过程中遵循了理论与实证的结合,并在此基础上将全文分成了五章进行。

第一章绪论。首先说明论文方向选择的经济背景和研究该方向的理论与实践意义;其次是文献综述部分,分别介绍了国际上以及国内众多学者对于黄金价格的研究深度,

让我对黄金价格的各个研究方向均有了一个较全面的了解与认识。基于此,本章对黄金价格的趋势预测文章进行了研究,并对国内外的相关研究文献进行了总结和分析。最终提出了四个目前研究者们为涉及的研究方向以及存在的一些可以丰富的地方;接着,对国际上以及国内众多学者分析的黄金价格影响因素的文献进行汇总,研究这些因素对黄金价格的影响。再次提出了本文的研究内容和思路、方法;结尾点明了提出了本文的创新之处以及不足。

第二章模型介绍。本章分别对论文中用到的模型进行了简单的介绍。

第三章基于数据挖掘技术和 ARIMA-GARCH 族模型的黄金价格动态分析。本章首先对黄金价格收益率数据进行了异常点检测,然后分别用支持向量机和 ARIMA-GARCH 族模型对去除异常点的黄金价格进行了预测,最后进行了一系列的对比分析,包括:(1) 除去异常点和不去异常点的 ARIMA-GARCH 模型预测结果对比分析;(2) 除去异常点和不去异常点的支持向量机模型预测结果对比分析;(3) ARIMA-GARCH 模型和支持向量机模型预测结果对比分析。

第四章基于 VAR 回归模型的黄金价格影响因素分析。主要是收集国内所有相关研究文献,在研究已有影响因素(主要是货币供应量、利率、汇率、通货膨胀率等)和金融和社会大环境实际环境基础上,引入基于百度指数的媒体关注度以及投资者关注度数据,立足于用户情绪的角度对黄金价格的影响因素进行分析。其中,媒体关注度和投资者关注度分别用媒体指数和用户搜索量指数来表现,所需要数据和指数可以从上海黄金交易所官网、金裕黄金官网、百度指数官网搜集到。然后,讨论各影响因素与黄金价格之间是否相关及其相关程度;另外建立回归模型判定各因素与黄金价格的关系。

第五章结论与建议。本章结合上文实证研究结果,指出黄金价格的动态预测分析结果,以及影响我国黄金价格的主要因素,给出全文的结论以及建议。

## 1.4 研究思路和研究方法

### 1.4.1 研究思路

首先利用黄金价格收益率数据,通过协整突跃点检测法、残差图检测法、箱线图检测法、基于聚类的方法等进行异常点的检测分析。其次,考虑到黄金价格具有的动态、非线性、时变的特点,利用除去异常点的黄金价格数据,分别建立时间序列模型

和支持向量机(LSSVM)模型对黄金价格进行预测。再次基于已有文献和社会、金融实际,考虑引入基于百度指数的用户搜索量指数和媒体指数来分析影响黄金价格的主观因素。其次,利用回归模型分析影响黄金价格的因素及其影响程度,模型中引入常用的货币供应量、利率、汇率、媒体指数和用户搜索指数因素。最后,得出结论,并提出政策建议。

#### 1.4.2 研究方法

##### (1) 理论分析与实证分析

文章首先将的黄金价格基本信息进行了研究,在分析的过程中注意了理论与实证的结合,并且进行了必要的研究和检验。其中,对黄金价格进行突变点检测的时候利用了非参数方法中的箱线图法,协整突跃点检测模型法,聚类等方法;对黄金价格进行预测分析的时候采用了数据挖掘中的支持向量机算法以及 ARIMA 模型;对黄金价格的影响因素进行分析的时候采用了 VAR 模型分析,实证分析分别利用最新数据通过 SPSS、Eviews、R 软件等实现。在研究中将采用定性政策研究与定量模型研究相结合,数学理论推导与数值模拟、实证研究相结合的方法。

##### (2) 对比分析

文章建立在作者熟练学习了计量经济学的一些必要的理论知识的前提下,选取适合的一些典型的变量,研究变量与黄金价格之间的变动机制。在这些工作均属于为定性分析做铺垫,定性分析之后便利用搜集的序列进行定量研究。通过模型的检验,来了解各解释变量对黄金价格是否具有影响以及其影响程度大小。最后将此模型结果与未引入新变量时的模型结果进行对比,查看变量影响程度的变化。

另外,在对黄金价格的预测分析章节中,在对黄金价格异常点检测以后,分别进行了去异常点和不去异常点的建模分析,了解其结果变化。并且分别利用支持向量机和 ARIMA-GARCH 族模型进行预测,对比其预测结果。

### 1.5 创新点和不足

#### 1.5.1 本文的创新点

(1)把基于百度指数的媒体关注度和投资者关注度引入到黄金价格影响因素分析中。

(2)引入了数据挖掘中的支持向量机算法,与常见的时间序列模型结合进行黄金

价格的预测分析。

(3) 采用协整突跃点检测法、残差图检测法、箱线图检测法、基于聚类的方法等多种方法对黄金价格进行异常点检测。

### 1.5.2 本文的不足

(1) 在研究黄金价格影响因素时由于本文涉及的变量比较多,各变量的数据结构不统一。但是鉴于论文中影响因素丰富度的考虑,部分因素不能删减,所以在分析时,我首先利用月平均数据进行了分析,但是最终建模结果却不理想。在改进的时候,若想利用每天的数据进行分析的话,就必须得放弃几个找不到日度数据的变量,这样就使得论文分析的不够全面,所以说在如何调整数据结构这一点上本文还存在需要改进的地方。

(2) 进行异常点检测时,可能用到的数据挖掘算法适用于离散型数据,但是本文所选的各种变量,离散化之后可能解释意义就没有那么明显了,所以最终也没有进行离散化,但是所幸所得结果和其他方法却是一致的。就数据的预处理方面存在的不足还是本文后续需要努力的地方。

## 第 2 章 模型介绍

本章对本文中所用到的模型分别进行了介绍，包括 GARCH 模型、支持向量机模型、VAR 模型、K-means 聚类算法、ARIMA 模型等。由于篇幅的原因，我们并没有进行详细的介绍，只是在能够理解的情况下进行了简单的介绍与说明。

### 2.1 GARCH 族模型的介绍

#### 2.1.1 GARCH 模型概述

GARCH 族模型属于 ARCH 模型的拓展，又被称作广义 ARCH 模型。它是 ARCH 模型的推广。 $GARCH(p, 0)$ 模型，相当于  $ARCH(p)$ 模型。1982 年 Engle 提出利用 ARCH 模型模拟时序列残差存在异方差，1986 年 T. Bollerslev 提出了 GARCH 模型，GARCH 模型是一个回归模型，准们用于分析金融数据，除了和普通回归模型具有很多相似的地方外，GARCH 族模型还对数据尾部进行了建模分析。该模型在数据的波动性分析与预测中特别适用，分析的结果对于投资者和生产者做出决策具有很重要的指导意义，其作用很多的时候超越了对序列本身的研究与预测。很多的分析发现金融产品收益率的分布不但是尖峰胖尾的，其残差对其的影响还不是对称的。每次市场经受负面的压力时，股价就会下跌收益率的条件方差也会扩大，这也致使股票价格的变动以及收益率的变动幅度更明显；相反，当股票价格上涨，各波动就会减小。因此负面压力对序列的条件方差的此影响又叫做杠杆效应。另外，GARCH 族模型不能模拟条件方差波动不对称的情况，是因为 GARCH 族模型默认特性为正负冲击对条件方差具有相同的影响。

GARCH 模型的参数估计可以使用极大似然估计和最小二乘估计。在进行极大似然估计时，GARCH 模型要比 ARCH 模型复杂。其复杂性表现在：GARCH 模型的参数估计不仅没有显示的表达式，而且，其似然函数也没有显示的表达式，只有迭代计算公式。这种特性会使得在进行极大似然估计时叙述变得复杂了，但是不会带来实质的困难。在现实应用时，经常采用 0 来对它们近似，所以，很容易得到近似的极大似然估计值。当然，求解过程又常用迭代算法，这里从略。对 GARCH 模型参数使用最小二乘估计方法，也同样遇到像极大似然估计类似的麻烦。尽管如此，当  $q$  值较大时，其算法也不比极大似然估计更方便。所以，最多使用的仍是极大似然估计方法。

GARCH 模型是目前经常会用到的条件异方差模型，我们常常用它来近似拟合观测数据。那么，在建模后还应当对所得到的模型进行检验，以判断其是否可接受。在建模前和后所进行的模型检验，其方法不一定相同。建模后使用的模型检验方法，还可作为确定 GARCH 模型阶数的辅助手段。在建立 GARCH 模型的过程中， $p$ 、 $d$ 、 $q$  的确定也是一个很有技术含量的点。对于 GARCH( $p$ ,  $q$ )，我们可以使用 AIC 准则方法，估计阶数  $p$  和  $q$ 。这里不再详细介绍了。我们要指出的是，在实际应用中一般不希望阶数太高，特别不希望虚阶数太高。为了尽量减少多余的参数，我们可以进行逐步试验，逐步尝试淘汰多余阶数。比如，用 AIC 方法确定 ARCH( $p$ )模型中的  $p=8$  时，而且检验也已通过，我们可以再尝试拟合 ARCH(7)模型，并检验之，如果仍能通过，则可降低模型阶数。类似地，当模型阶数不太高时，我们也可依次拟合  $p=1, 2$  等低阶 ARCH 模型，并进行检验，以通过的最小阶数作为阶数估计，这也是阶数估计的一种方法。对于 GARCH( $p$ ,  $q$ )模型，此方法也适用。

### 2.1.2 某些新进展

由于我们现在用 GARCH 模型比较多，且其本身的一些好的特性也造就了它前景的广阔。因此一经提出该模型就迅速的被人们所接受并认可，并且可预见的它会有一个长足的发展。事实证明也是如此，以上几节只是介绍了此模型进展的梗概，在此最后一节，我们再概述某些重要的新进展。下面介绍一些改进模型。

#### (1) $\beta$ -GARCH 模型

从前面叙述 GARCH 模型的平稳性条件来看，所有人都会感到太苛刻了，特别当  $p$  或  $q$  较大时，每个  $\alpha_i$  和  $\beta_j$  都小得很，实用受限制。于是有如下模型出现， $h_t = \alpha_0 + \alpha_1 |y_t - 1|2\beta + \alpha_2 |y_t - 2|2\beta + \dots + \alpha_p |y_t - i|2\beta$ ,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ , (2.1)  
 $i = 1, 2, \dots, p$ 。其中  $0 < \beta \leq 1$ ，当  $\beta=1$  时就是 ARCH 模型，当  $\beta < 1$  时，为  $\beta$ -ARCH 模型，同理可推广出  $\beta$ -GARCH 模型。使用此模型时， $\beta$  值也需要估计，此估计带来较多的麻烦。

#### (2) 指数 GARCH 模型

以上  $\beta$ -ARCH 模型克服了条件限制太强的缺点。于是，有人提出对建立模型。比如：

$$\log h_t = \alpha_0 + \alpha_1 \log h_{t-1} + \alpha_2 \log h_{t-2} + \dots + \alpha_p \log h_{t-p} + \varepsilon_t \quad (2.2)$$

这里 $\{\varepsilon_t\}$ 被假定为 i.i.d. 的随机序列, 且  $E\varepsilon_t=0$ 。此模型有平稳解的条件与熟知的 AR(p) 模型有平稳解的条件相同。记上式的解为  $u_t$ , 则  $h_t=\exp\{u_t\}$ 。

## 2.2 K-means 聚类算法的介绍

数据挖掘技术中的聚类算法一般针对离散数据进行聚类分析, 对于聚类结果而言, 我们可以从多个不同的角度研究, 每一个角度都会得到很多对我们有利的信息, 其中一个角度就是一个很好的查找异常点的方法。

聚类是一个数据集分成几组, 或类的过程中, 在同一组数据对象有很高的相似性, 以及数据对象在不同的组中是不一样的。相似或不相似的描述是基于数据描述属性值。聚类算法是目前非常多, 但具体问题算法的选择主要取决于类型的数据和集群和应用程序的目的。此算法中具体来说又可以有好多不同的原理组成, 基于这些原理又可以把算法进行分解, 分别可以包括: 网格法、网格法、分层方法, 网格法、模型法、密度法。层次聚类方法对给定的数据集分解, 直到某种条件满足, 桦木等典型算法。基于密度的聚类思想是只要一个点的密度大于某个阈值, 就会将它添加到集群, 典型的如 DBSCAN 算法。基于网格的聚类方法是量化空间有限数量的单位, 然后在定量空间聚类操作等典型算法。基于模型的聚类方法的一种尝试的一个给定的数据和数学模型来实现最好的配件, 这些方法通常是基于假设的数据是根据概率分布生成的潜力。可以用建立模型来分析的方法包括统计学以及神经网络方法。以下采用 K-means 聚类的算法来查找异常数据。根据前面小节分析发现在我们选取的数据时间段内, 2011 年和 2013 年黄金价格波动较大一些, 本小节只是想说明 K-means 聚类法在异常点查找中的应用, 所以以下分析只选取 2011 年和 2013 年的数据进行分析。

## 2.3 支持向量机模型的介绍

SVM (Support Vector Machines, 支持向量机) 线性可分情况下的最优分类面。意味着有需求分类最优分类线不仅可以单独的两类错误, 和分类两个类之间的间隔是最大的, 前面的是确保经验风险值达到最小(0), 并通过讨论发现最大的分类间隔实际上是由分类中心决定的。而对于最优分类线而言, 将其扩展到高维空间时就变成了最优分类面。支持向量机的训练过程是根据分类间隔的原则进行的, 它的训练结果取决于我们对数据所做的预处理, 也就是说, 在高维的空间中所表达的原始模式。

SVM 的基本思想可以概括为：首先通过一个非线性的变换将数据所输入的空间映射到一个更高维的空间，接着在此高维空间中来求最优分类线所对应的最优分类面，而所做的非线性变换之所以能够实现是因为定义了合适的内积函数。通过 SVM 求得的分类最优分类面形式上和数据挖掘技术中的神经网络很相似，它输出的是几个线性的组合，这个组合表现的是中间层的节点，然而每一个输出的节点都与一个输入样本和所谓的支持向量机的内积相对应，所以也被叫做支持向量网络。由于训练出的几个向量的内积函数和求和函数仅包含最后的判别，从而判别分类的计算的复杂程度就决定于所训练支持向量的个数。

不难发现，SVM 与经典的统计理论知识以及传统的分析方法相比来说，它是一种完全不一样的思维方式，而不像常见的一些统计方法一样，它第一步就尝试原来所输入的空间的降维方法，而是试着输入空间 ID，在高维度的空间范围内线性可分问题或者是接近线性可分的问题。因为 ID 知识改变了内积操作，不会增加算法的复杂性增加的维度，在高维空间和泛化能力不受影响的维度。

## 2.4 ARIMA 模型介绍

ARIMA (Autoregressive Integrated Moving Average Model, 自回归移动平均模型) 是 70 年代初由博克思 (Box) 和詹金斯 (Jenkins) 提出的一种著名的预测方法。当  $d$  不等于零时，该模型叫做差分自回归移动平均模型，其中 AR 是自回归， $p$  是自回归的项数；MA 表示移动平均， $q$  是移动平均的项数， $d$  表示数据平稳化所需差分的次数。

ARIMA 模型的基本思想是：模型对于预测对象所形成的序列数据随着时间的随机的序列，与一定的近似数学模拟模型来描述序列。一旦确定的模型可以从过去的数值来预测未来。

ARIMA 模型进行预测的基本步骤是：

a. 根据散点图、时间序列的 acf、pacf 图了解其数值从而观察其趋势性以及季节性、用 ADF 单位根检验其是否存在单位根从而判断其平稳性。一般来说，我们通常分析的一些经济金融方面的数据大都是非平稳序列。

b. 光滑的不平稳序列的处理。如果数据序列是平稳，有一定的增长或下降，你需要不同的数据，如果数据是异方差性，需要技术处理数据，直到经过处理数据的自相关函数和偏相关函数值无显著不同于零。



c. 按照时间序列模型固有的定阶原则，构建合适的模型。如果一个光滑的偏相关函数序列截断，自相关函数后，可以得出结论，AR 模型的序列；如果平稳序列的偏相关函数后，自相关函数的截断，可以得出结论，MA 模型的序列；如果一个平稳序列的偏相关函数和相关函数后，序列 ARMA 模型。

d. 进行参数估计，以检测参数是否拥有统计上的意义。

e. 进行假设性检验，判断所得残差序列是否为白噪声序列。

f. 用已经检验过的模型进行预测研究。

## 2.5 VAR 回归模型介绍

VAR 模型全称为向量自回归模型，在计量经济学中用的较为普遍，而且易于掌握。VAR 模型是使用所有的利润中所有变量的滞后变量的回归模型。VAR 模型是用来评估联合内生变量之间的动态关系，和之前没有任何约束。泛化的 AR 模型，这个模型目前已广泛应用。

向量自回归(VAR)是基于数据模型的统计特性，VAR 模型，系统的每个内生变量的滞后值系统所有内生变量的函数来构造模型。VAR 模型用于处理几个具有相关性的经济指标，其分析和预测方面来说是很简单的一个模型，另外，在合适的条件下，多元的 MA 和 ARMA 模型也可转化成 VAR 模型，因此近年来 VAR 模型越来越多的被用到一些经济工作中。

VAR 模型说的是在相同的数据区间内的  $n$  个指标能够作为其历史数据的线性函数，因变量的差分项同样成为解释变量放置在所列方程的右端。这样的既包含因变量本身又包含因变量差分项的方程叫做自回归模型。

在这种模型中，解释变量  $X$  与它之后项之间通常具有高度相关关系，因此会带来严重的多重共线性问题。因此，经常在模型中添加增加一些先验约束条件以便降低需要估计的独立参数的数量，进而避免多重共线性问题，或者至少减少其影响。

## 2.6 本章小结

本章中我们对以下实证分析所用到的模型分别进行了介绍与说明，其中 ARIMA 模型和 GARCH 族模型用于分析黄金价格的收益率波动性分析以及黄金价格的短期预测分析；支持向量机是一种基于数据挖掘技术的预测模型，可以对黄金价格的非线性部分进行预测，并且可以进行长期的预测；向量自回归模型是用来分析黄金价格的

影响因素，在引入新的变量后进行回归分析，分析出各影响因素对黄金价格是否具有影响以及其影响程度大小；**K-Means** 聚类算法主要用于对黄金价格数据的异常点进行检测。本文中还用到了其他的一些方法和算法，基于篇幅的原因我们只简单是介绍这几个模型，其他一些简单的方法就不再一一进行介绍了。

## 第3章 黄金价格动态预测分析

前面章节已经分析了目前对于黄金价格研究的各个领域以及进展程度，并且对我们所用模型进行了简单的介绍，所以本章的重点就是针对前面提到的一部分问题进行实证分析。本章的实证分析数据来源于 AU9999 2011 年 1 月 4 号到 2015 年 11 月 02 号的日收盘价，共收集了 1171 个数据样本，并基于此求得黄金价格的日收益率数据进行后续的分析。下面，首先对黄金价格收益率数据进行了异常点检测，然后分别用支持向量机和 ARIMA-GARCH 族模型对去除异常点的黄金价格进行了预测，最后进行了一系列的对比分析，包括：（1）除去异常点和不去异常点的 ARIMA-GARCH 模型预测结果对比分析；（2）除去异常点和不去异常点的支持向量机模型预测结果对比分析；（3）ARIMA-GARCH 模型和支持向量机模型预测结果对比分析。

### 3.1 黄金价格的异常点检测

在序列分析时，原始数据通常涵盖某些这样的序列对象，他们的行为和模型的一般行为并不一致，我们通常把这样的数据对象称作异常点。异常的定义是指在数据集中存在不同的数据点，这些点脱离了大部分数据，但看规律会使人怀疑这些数据不是随机出现的，这就是异常点。大部分的数据挖掘方法找异常点都是处理噪音或直接丢弃。然而在某些具体事件中，比起正常时间的检测，对于罕见的事件（异常点）的检测通常是更有意义的。

#### 3.1.1 协整突跃点检测模型

为方便分析黄金价格的变动规律，在此我们把模拟模型后得到的残差序列的条件方差和原数据图对应排放。在图 3-1 中，能够明确的看到两个“异常点”，一个是在 2011 年 10 月份，另一个是在 2013 年 4 月份。在 2011 年 10 月份前有一个小的波动，到了 10 月份黄金价格经历了一次直线下跌，然后又有反弹的迹象，整体来看，黄金价格的波动一直很大，在 2013 年 4 月份的时候又出现一次直线下跌，下跌之后虽有所反弹，但是整体来说却没有再回到原来水平了，一直处于较低迷的状态。

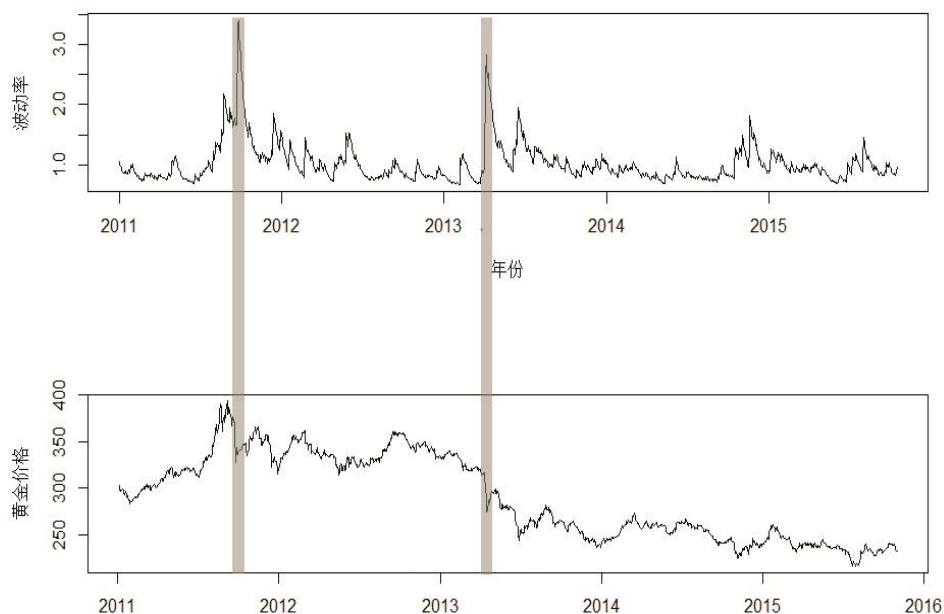


图 3-1 黄金价格的波动率以及金价趋势图

为了寻找突变所在的具体日期,可以利用汇率与黄金价格的单阶协整关系进行查询该突变点所在。首先建立汇率和黄金价格的变结构 Johansen 协整模型,然后利用 R 软件中的 `model.break@bp` 语句提取变结构 Johansen 协整模型的突变点位置,得到第 547 位。

结果表明,协整的突变点位于第 547 个样本数据,即 2013 年 4 月 16 号。下面画出 2013 年各月的箱线图分析具体异常点所在:

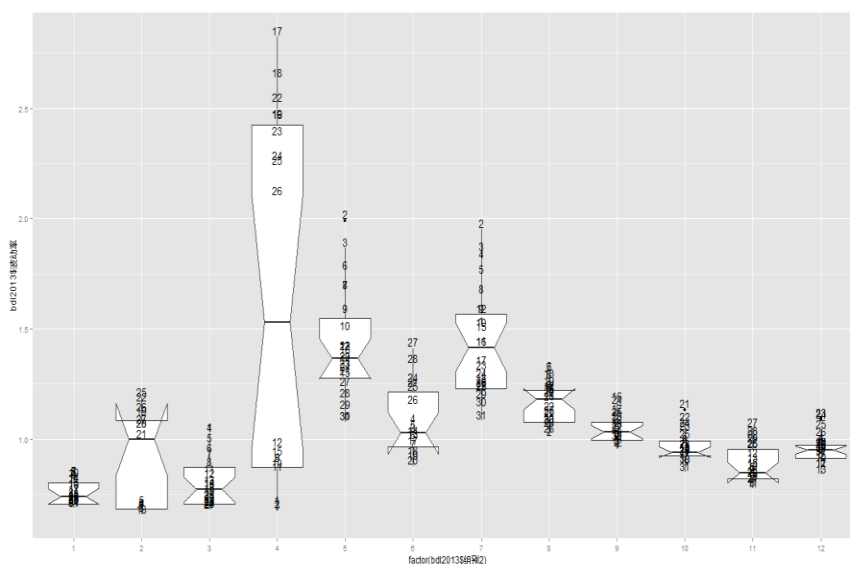


图 3-2 2013 年各月波动率的箱线图

由图 3-2 可以看出,4 月份的整体波动率是非常大的,结合 4 月份的黄金价格数据我们发现,4 月份黄金价格出现了暴跌。究其原因,四月基本面数据繁多,但主要

诱使黄金期货价格下跌的数据就是塞浦路斯抛售黄金储备还债以及美联储考虑提前结束 QE 政策初始黄金价格空头压力加大。黄金不仅可以抵抗通胀风险，在经济市场动荡或者经济危机发生时也非常受风险偏好者的喜爱。所以，不管是美国次贷危机发生时期还是欧债危机发生时期，黄金价格都保持了上涨的态势。然而，本轮塞浦路斯危机爆发以后，以德国为首的政府保持着非常艰苦和恶劣的条件提出了援助，这将导致政府不得不宣布，通过出售 10 吨黄金储备增加纾困资金的一部分。塞浦路斯政府激励市场的想象力，因为当前的南欧国家的债务危机高举黄金储备，特别是意大利政府的黄金储备超过 2000 吨。另外在 5、10、12 月份出现了一个波动率的离群点，即只有某一天的波动出现了异常。由该图还可以看出，2013 年各月的异常日期为 5 月 2 号，10 月 21 号，12 月 22 号，存在三个异常点。并且可以看到 4 月 17 号是波动率最大的一天。

### 3.1.2 对收益率数据取其标准化残差之后分析异常点

由图 3-3 可以看出，所建模型标准残差的时间序列图在 2011 年 10 月以及 2013 年 2 月时存在着小的可疑的标准残差，即存在异常点。下面我们可以利用箱线图具体确定该异常点的日期。

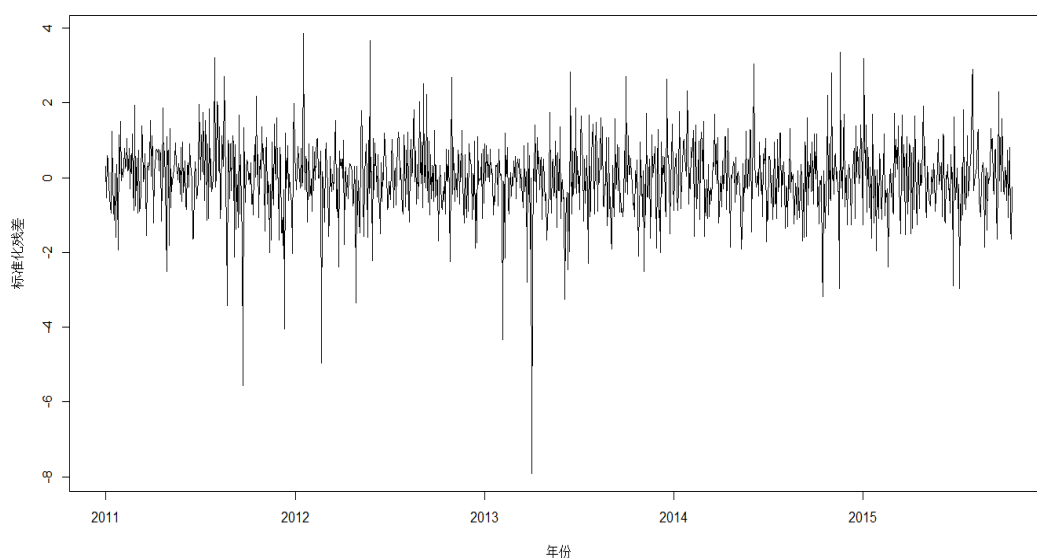


图 3-3 标准化残差趋势图

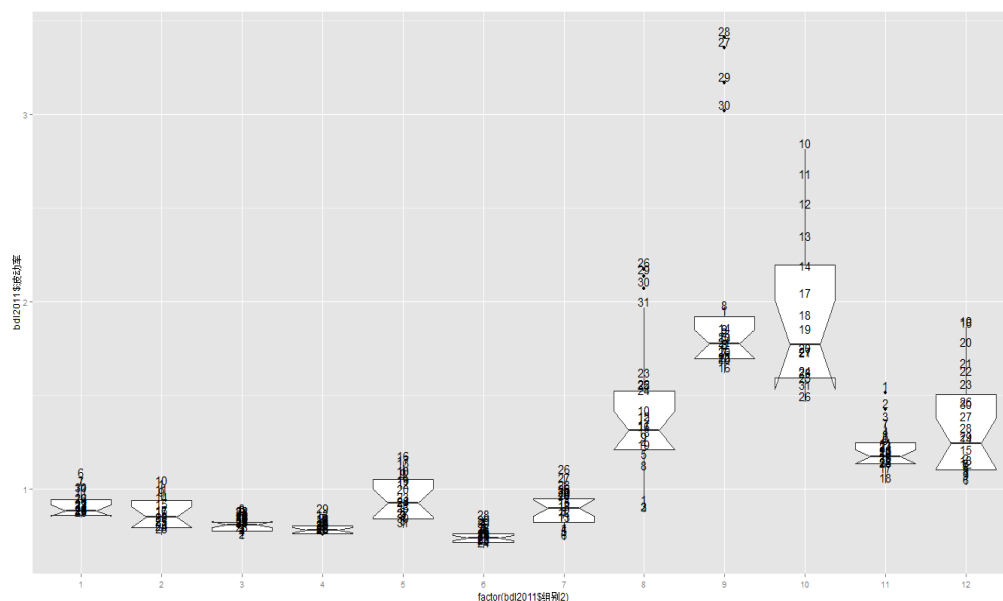


图 3-4 2011 年各月的波动率的箱线图

由图 3-4 可以看出，2011 年整体来看上半年黄金价格的波动率是比较小的，但是从 8 月份开始黄金价格的波动率开始出现大的波动，其中 10 月份的整体波动率是非常大的，具体来说 9 月份的 27、28、29、30 以及 10 月的 10、11、12、13、14 黄金价格波动率是最大的，即国庆节前后的一段时间，黄金价格达到顶峰，因为廉价劳动力供应全球拐点，人口的东欧，俄罗斯，资源是有限的，中国的人口红利，往往是，沿海地区出现短缺，替代劳动资本开始削弱，部分过渡到一个中立的流动性，即使美联储和其他央行实施量化宽松政策，也很难扭转这一趋势，温和的性能也可以从年的全球通胀。和美联储即将退出量化宽松政策的大背景下，全球流动性显然将进一步宽松，缺乏内在价值，仅仅依靠资金流入黄金被打破。流动性宽松，甚至很快就开始向紧的方向进化，金价将继续压力，这将限制金价未来的潜力。而黄金的下一轮牛市，仍然需要大量资金的支持，可能提供充足的廉价劳动力的国家，如印度、印尼加入国际分工，生产力量，黄金将打开一个新的牛市。

### 3.1.3 箱线图查找波动率的异常点

以上两小节中我们已经利用箱线图的知识查找了一些异常点的具体日期所在，下面我们利用箱线图做一下具体分析。将波动率按照年份分成五组，2011 年为第一组，2012 年为第二组，以此类推，得到如下箱线图 3.5（由箱和须两部分组成，须是由箱子的边缘出发延伸至 1.5 倍四分位距内的最高点。若有超过须的数据点，则被视为异常点）。

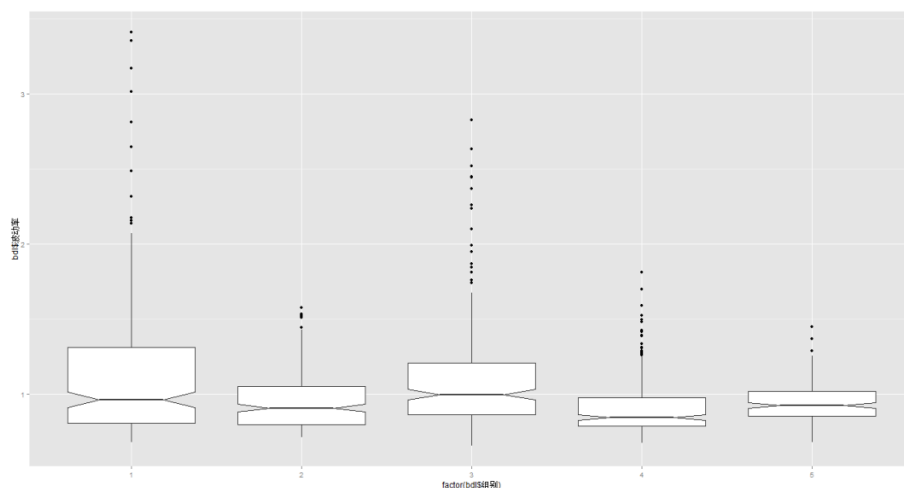


图 3-5 2011~2015 五年的波动率情况

由箱线图的基本知识我们知道图中槽口表示各年波动率的中位数,对于异常点较多的一组数据来说,中位数可以表示该组数据的平均水平。由上图的中位数所处的位置可以看出,综合来说 2013 年的黄金价格波动率平均水平是最高的,其次是 2011 年,由须的最高点上面的点可以看出,2011、2013、2014 年黄金价格的波动异常点较多,说明这三年的黄金价格波动较大,相比较来说,2012 年和 2015 年黄金价格相对较稳定。

### 3.1.4 K-means 聚类法找异常点

首先根据数据选择最好的聚类数,在 R 软件中由 `FitKMeans` 和 `PlotHartigan` 语句直接实现得到图 3-6 所示结果。

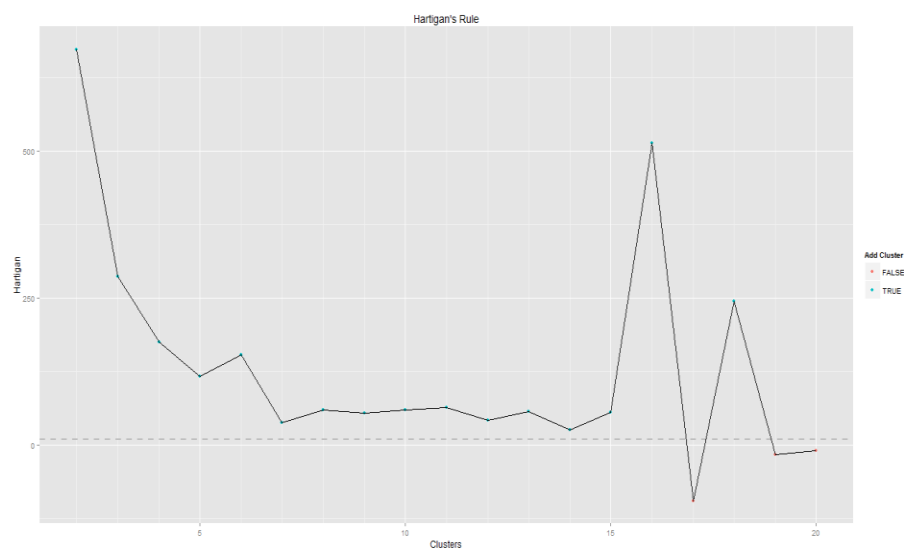


图 3-6 2011 和 2013 年的最佳聚类数结果

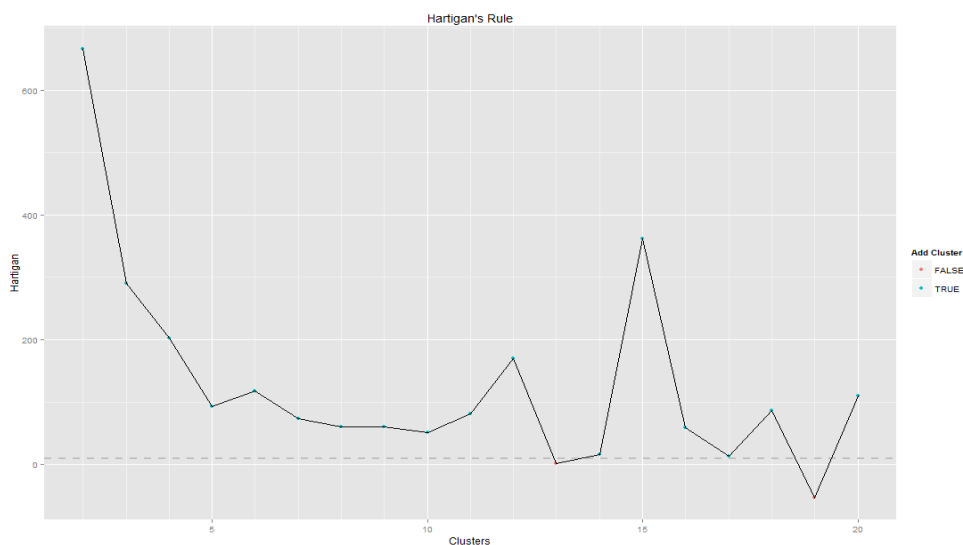


图 3-6 2011 和 2013 年的最佳聚类数结果（续图）

由图 3-6 可以看出 2011 年和 2013 年的最佳聚类数分别为 16 和 12 类。并且可以进一步得到所聚类别的类中心波动率值，如下表 3-1 所示。由表 3-1 可以看出，2011 年中第九类中类中心波动率为 0.7406，是波动率最小的，对应 6 月份，即可知六月份的黄金价格波动率整体来说是最小的。

表 3-1 2011 和 2013 年的类中心波动率值

聚类数	2011 年		2013 年	
	类中心波动率	月份	类中心波动率	月份
1	1.4042	8	1.2292	8
2	0.8843	7	0.8429	2.583
3	1.0507	12	2.4282	4
4	1.3997	12	1.0588	10
5	0.886	1.455	1.2832	6.575
6	1.821	12	0.7115	1
7	1.1237	12	0.9268	10
8	1.5853	12	0.9161	11.5116
9	0.7406	6	0.8059	1
10	1.2572	12	1.2458	4.7097
11	1.7608	9.43333	1.0340	9
12	0.7828	4	1.0660	8
13	0.7987	3		
14	1.2035	11		
15	2.9865	9.42857		
16	0.9437	5		

另外，第十五类的类中心波动率是最大的，达到了 2.9865，对应月份的中心值为 9.42857，也就是说此类波动率出现在九月份和十月份之间的一些天。其次是出现在



12 月份的第六类波动率达到了 1.82115，也是很大的。在 2013 年中波动率最大出现在 4 月份，达到了 2.4282，其次是 6.575 月份，即六月份和七月份之间，达到了 1.2832 的波动率。下面我们利用图形具体定位这些波动较大的日期，如表 3-1 所示。

由图 3-7 可以看出，2011 年中波动较大的日期具体为 9 月 27、28、29、30 以及 10 月 10、11、12 号。而 2013 年波动较大的日期具体为 4 月的 17、18、22、23、24、25、26 以及 5 月的 2、3、6、7、8、9、10，另外还有六月底到七月初的几天。

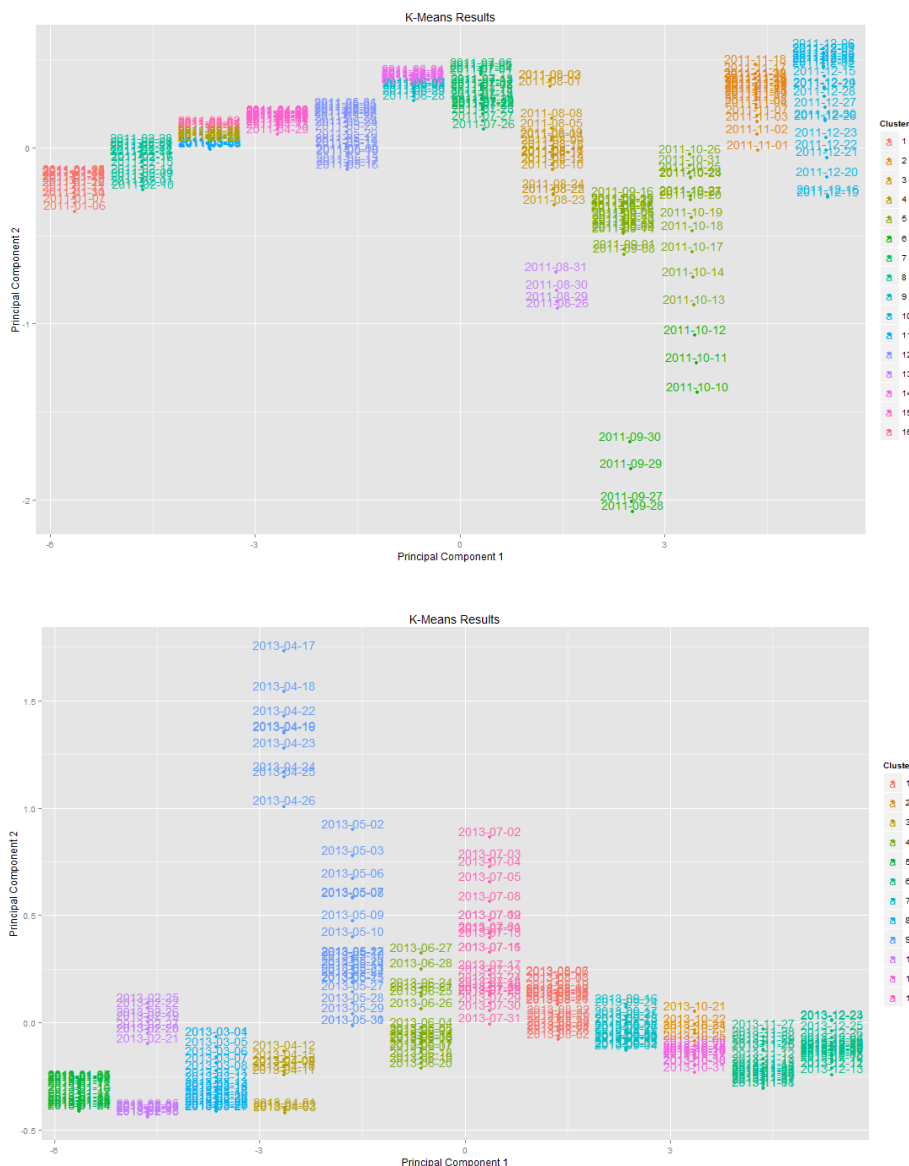


图 3-7 2011 和 2013 年的波动率异常点检测结果图

从检测结果图 3-7 中可以看出，异常的一些数据点大致均在数据集的周边区域，它们离其他数据较远，这写特征和异常点的概念是完全符合的。而对于图中那些正常范围内的数据集中的异常数据点的分布问题而言，这是因为原始数据集中的互相重复

的数据点比较多所造成的，重复数据点在图片之上的分布只是一个点而已，它们基本上大都分布在正常的区域，而异常的一些数据点发生重复的情况却是很少。综合以上分析我们可以知道，用聚类的方法找到的异常点与协整检验、标准残差图、箱线图的方法找到的异常点是一致的。

综合上述分析，黄金价格每年较异常的数据可以归纳如表 3-2 所示：

表 3-2 异常点所在日期

年份	异常点所在日期
2011	9 月 27、28、29、30，10 月 10、11、12 日
2012	5 月 2、3、4、5、6、7、8、9 日，8 月 24 号，11 月 8 日
2013	4 月 17、18、22、23、24、25、26 号，5 月 2、3、6、7、8、9、10 日
2014	6 月 23、24、25、26、27、30 日，9 月 28、29、30 日，10 月 31 号、11 月 18、19 日，12 月 3、4、5、8、11、12 日
2015	3 月 10、11 日，4 月 29 日，8 月 13、14、17、18 日

本小节通过具体的例子分别详细介绍了利用不同的方法进行异常点检测的过程，可以看出，集中的方法在进行异常点的检测问题时是非常有效的。但是，在实际的应用过程中还应该注意一下异常点所拥有的相对性和主观性特点，检测者在对异常点进行挖掘研究时必须依据所选数据以及研究问题的一些特定的挖掘背景，对最终探测到的异常的数据集的确定通常要首先通过一些特定的算法确定范围，然后再结合领域专家的意见方可最终确定。本文在后续的黄金价格动态分析中将分别利用上述所找异常点的结论，使用带有异常点和除去异常点的两种数据进行黄金价格的预测分析，并进行结果对比分析。

## 3.2 基于 ARIMA-GARCH 族模型的黄金价格预测分析

### 3.2.1 黄金价格数据的基本分析

随着经济的迅速发展，股票交易市场已变成居民管理财富的主要途径，但是股票交易由于受多方影响一直被认为是一个高回报和高风险并存的理财方式，想要得到大量的利润，就一定要准确把握股票价格的发展态及其波动规律，所以对于股票价格的预测问题也是金融领域里的不可或缺的研究课题。

中国境内，黄金交易市场产生以后很快引起了许多研究者的青睐，他们纷纷提出了一些预测黄金价格的方法，常用的一些有成本法分析、时间序列分析、线性回归分析、灰色预测算法分析等。其中的时间序列分析中 ARIMA 使用最为广泛，国内外学

者对黄金价格趋势的研究的文献很多,但是鲜少有提到用数据挖掘方法的,并且我们知道常用的时间序列模型是基于线性建模的,而黄金价格是一种非线性、时变的时间序列数据,为了提高预测精度我们提出一种基于数据挖掘的黄金价格预测模型。首先采用线性预测模型 ARIMA 对黄金价格进行预测,然后采用支持向量机(SVM)对黄金价格非线性部分进行预测,最后将两模型的预测值进行对比融合形成黄金价格的预测值。

本文的实证分析数据来源于 AU9999 2011 年 1 月 4 号到 2015 年 11 月 02 号的日收盘价,共收集了 1171 个数据样本,由上一小节我们知道数据每一年都会有一些异常点存在,所以在后续的分析中,我们建模的过程中将每年数据中比较异常的点去掉,进而对黄金价格进行预测分析。所以除去异常点之后,此时数据总数变为 1118,其中前 1000 个数据作为训练集,对黄金收盘价进行建模,后面的 118 个样本数据作为测试集,等预测结果出来以后与其进行对比进而计算预测的误差率,方便量化并与其他模型的精确度进行对比。AU9999 的日收盘价的走势如图 3-8 所示。首先,根据实际数据可以做出黄金价格的走势图 3-8:

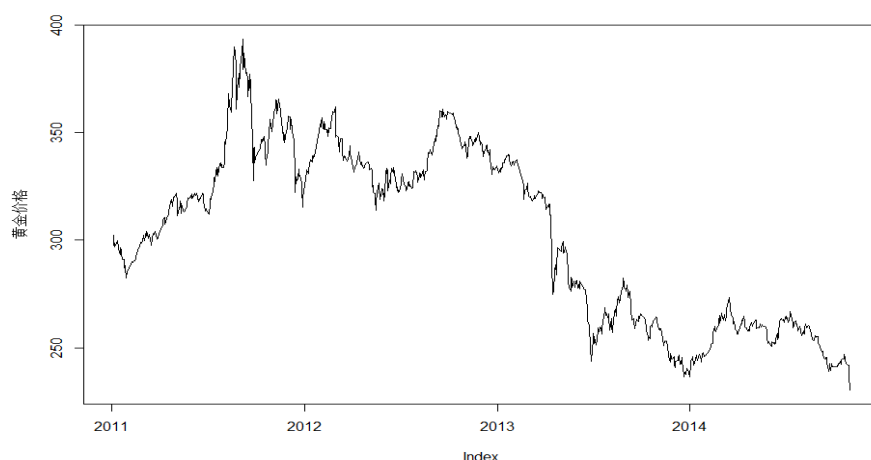


图 3-8 2011 年 1 月 4 号到 2015 年 5 月 4 号中国黄金价格走势

从图 3-8 可以看出,2011 年 1 月 4 号至今,中国黄金价格大体呈现一个先上升后下降的态势。2011 年 5 月底之前黄金价格虽有波动,但是很小。随后黄金价格几乎呈直线上升,直到 8 月底达到一个高峰 389.8 元/克。之后在 2012 年 8 月的一年间黄金价格出现了大幅度的波动,在 9 月初出现了大幅度的下滑,直到 2013 年年中黄金价格逐渐回升的迹象,但始终没有所突破,至今只是小幅的波动而已。下面分别用不同模型对黄金价格的走势进行预测分析。根据数据走势以及数据的特点得到黄金价格

数据的分解图如下：

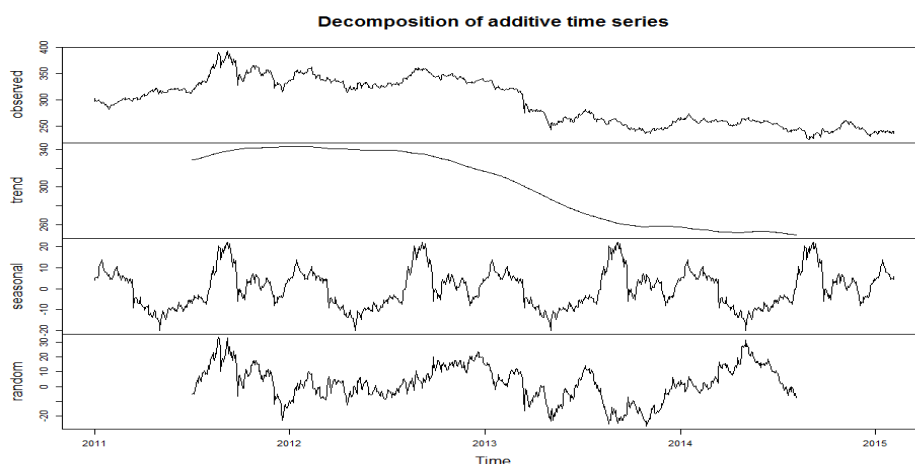


图 3-9 黄金价格数据分解图

由图 3-9 可以看出，模型长期趋势为递减的，并且有明显的季节趋势，随机性波动幅度也很大。周期为 244 天。所以在后续的分析中，要根据数据周期 244 天的特点展开。

### 3.2.2 ARIMA 模型的拟合

时间序列方法是指经过对历史数据的分析来揭示序列随着时间变化的内在规律性，并通过一定的模型规则将这种规律具体呈现出来并延伸到未来，即预测。在众多的预测方法中 ARIMA 方法是最常用也是很有效的一种方法，这种方法的精髓之处在于用自身和随机误差来解释自身。若时间序列数据不是平稳的，则需要对序列进行  $d$  阶差分，以使序列变成平稳的序列，然后对平稳化后的序列用 ARIMA 模型进行建模。

在具体的分析过程中，首先本文对原序列做时间序列图，观察序列的规律，我们发现黄金价格数据  $h$  具有异方差性，非平稳性。在对黄金价格平稳化的过程中，本文对黄金价格数据做了以下处理：（1）利用了 R 软件中 forecast 包中的最优差分次数函数，确定了最优差分次数为 1。（2）对一阶差分后的数据  $dh$  进行 ADF 检验，另外画出差分后数据的时间序列图，发现该数据为平稳序列。（3）利用 forecast 包中自动确定模型结构的 auto.arima 函数，该函数是根据 AICc 准则(对模型复杂性的惩罚函数做了修正的 AIC 准则)得到模型 ARIMA(1, 1, 1)。（4）对该模型拟合后的残差进行白噪声检验，发现其残差符合白噪声的特征，证实了我们的模型选择是合适的。从而确定最终模型为 ARIMA(1, 1, 1)。与上面分析的结果也是相符的，所以最终确定模型的阶数为 ARIMA(1, 1, 1)。（5）利用语句 coef 确定 ARIMA 模型的系数，结果如表

3-3。

表 3-3 模型系数结果

	mu	ar1	ma1
系数	302.32	-0.3003	0.2148
P 值	0.0000	0.0000	0.0745

由上表可以得到模型的最终表达式为：

$$dH=302.32-0.3003*dH_{t-1}+0.2148*\varepsilon_{t-1} \quad (3.1)$$

此时所建立的模型并没有考虑黄金价格数据的尖峰厚尾的特点，所以为了使预测结果更准确，这里我们首先进行 ARCH 效应的检验，因为若数据存在 ARCH 效应，那么我们在建立 ARIMA 模型时要加入 GARCH 族模型。所以在此利用残差或者数据平方构造的 B-L 统计量来判断其 ARCH 效应，结果如图 3-10 所示：

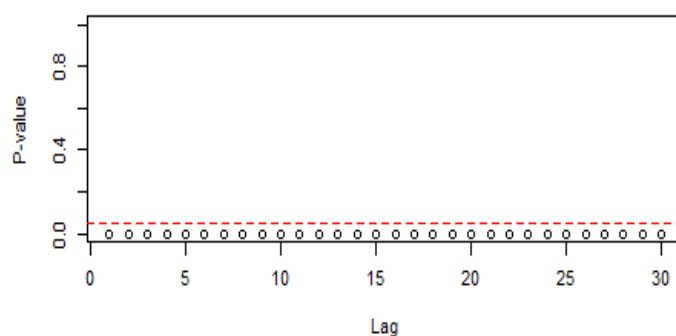


图 3-10 ARCH 效应判别图

由图 3-10 可以看出，数据的 P 值均在 0.05 以下，拒绝原假设，即认为黄金价格数据存在显著的 ARCH 效应，即存在明显的异方差性，可以建立 GARCH 模型。所以在后续建立 ARIMA 模型时我们加入 GARCH 族模型的考虑。

### 3.2.3 基于黄金价格收益率的 GARCH 族模型的建立

#### (1) GARCH 族模型的建立

本文对黄金价格进行动态分析的收益率采用日对数收益率。在价格序列变化很小的情况下，对数收益率与收益率近似相等，并且因为对数收益率具有可加性，因此可以在后续处理时更好地拟合数据。所求对数收益率时序图如图 3-11 所示。

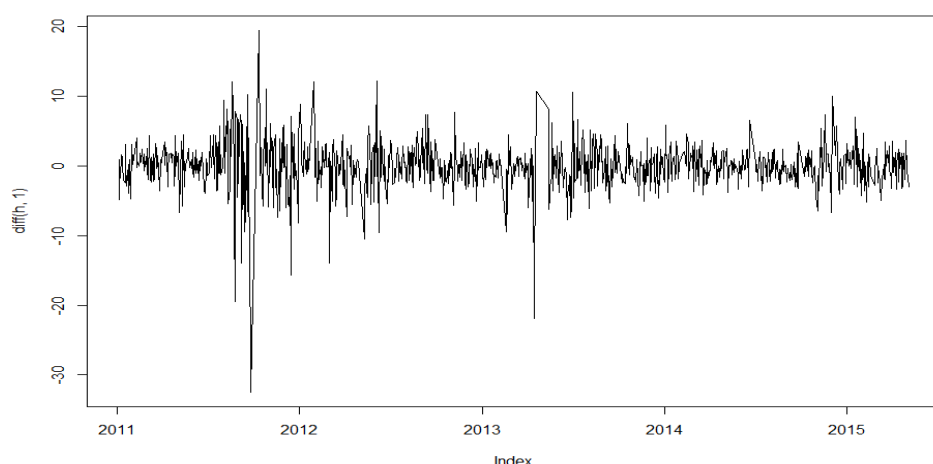


图 3-11 黄金价格对数收益率的时序图

从以上对数收益率序列的时序图中可以观察到对数收益率波动的集群现象：收益率在零周围频繁上下波动，波动在一些时间段内较小，在有的时间段内非常大，并且在大的波动后往往伴随着另一次大的波动。

收益率的直方图和一些基本的统计量如图 3-12 所示。收益率序列均值为-0.00019，标准差为 0.011175，偏度为-1.2354，小于 0，说明该序列的分布特征有比较明显的长的左拖尾，说明投资亏损的概率要比盈利的概率大。我们知道正态分布的峰度值为 3，而此分布的峰度为 13.142，比正态分布的高，这也说明黄金价格收益率数据具有高峰厚尾的特点。Jarque-Bera 统计量为 3682.135，对应的 P 值为 0.000000，因此认为该收益率序列不服从正态分布的假设。

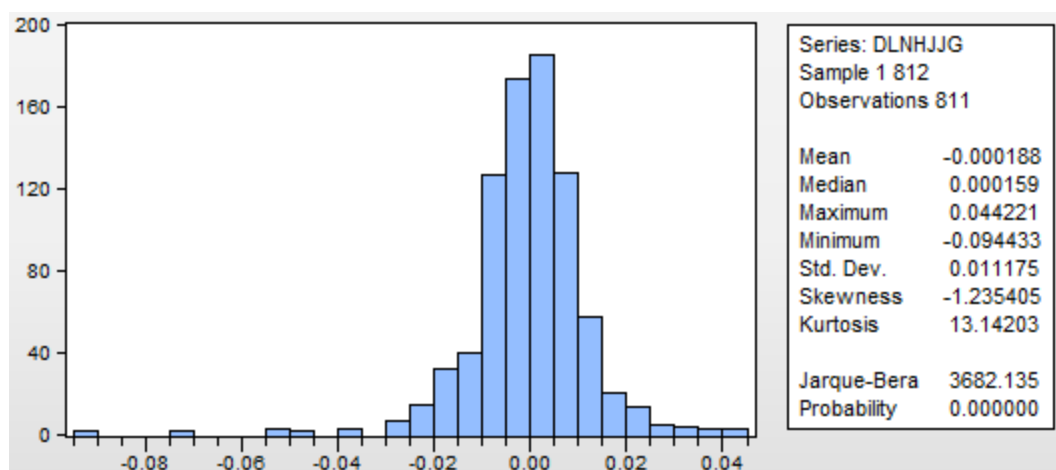


图 3-12 收益率的直方图

由上面的分析我们知道黄金价格收益率数据是可以建立 GARCH 模型的，那么接下来我们就利用收益率数据的 acf, pacf 检验来判断所建立的 GARCH 模型的阶数。

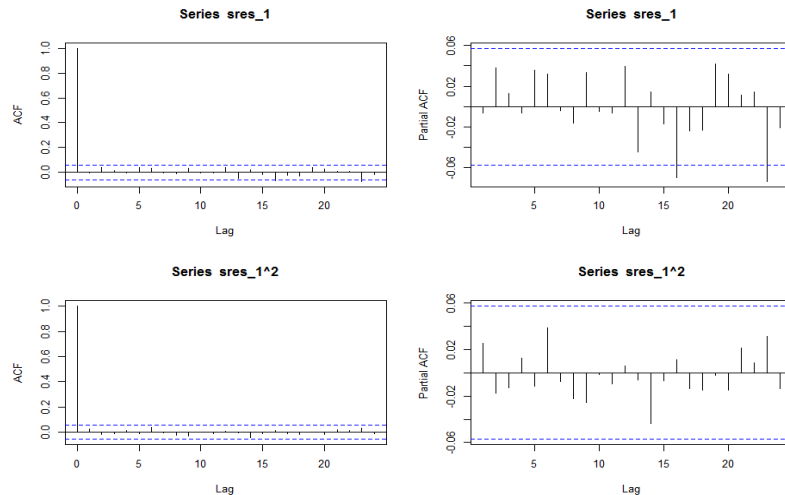


图 3-13 收益率数据的 acf, pacf 检验结果

由图3-13可以看出,我们可以分别利用GARCH(1,1)、GARCH(2,1)和GARCH(1,2)建模。模型的结果如表3-4所示。

表 3-4 GARCH 模型的建模结果

		系数	P 值
GARCH(1, 1)	mu	322.2914	<2e-16
	omega	5.2394	1.53e-07
	alpha1	0.9008	<2e-16
	beta1	0.11	0.0383
GARCH(1, 2)	mu	322.2914	<2e-16
	omega	5.251	5.55e-06
	alpha1	0.9006	<2e-16
	beta1	0.1094	0.518
	beta2	1.00e-08	1.0000
GARCH(2, 1)	mu	322.2914	<2e-16
	omega	5.967	0.0128
	alpha1	0.8914	<2e-16
	alpha2	0.1195	0.8145
	beta1	1.00e-08	1.0000

从表3-4中可以看到模型GARCH(1,1)中3个系数对应的P值都小于0.05,也就是说通过了t检验,而GARCH(2,1)和GARCH(1,2)模型中都有系数无法通过t检验,因此可以初步选定拟合的模型为GARCH(1,1)。

## (2) GARCH 族模型的修正与检验

在确定了GARCH模型的阶数后,可以继续对残差进行TGARCH、APARCH和GARCHM建模,能够比较好的模拟收益率数据的波动性以及尖峰和厚尾的现象。建模具体结果见表3-5。

表 3-5 TGARCH、APARCH 和 GARCHM 模型结果

	系数	P 值
APARCH(1, 1)	mu	322.2936
	omega	5.2380
	alpha1	0.9022
	gamma1	-0.0021
	beta1	0.1097
	delta	2.0000
TARCH(1, 1)	mu	322.71
	omega	1.2064
	alpha1	0.7691
	beta1	0.1948
	beta2	0.0146
GARCHM(1, 1)	mu	335.772
	archm	-0.5141
	omega	6.9649
	alpha1	0.9046
	beta1	0.0944

由表 3-5 可以看到 TGARCH 和 APARCH 模型的系数以及其 P 值, 此处系数的 P 值均有大于 0.05 的, 说明模型系数无法通过显著性检验, 而 GARCHM 模型的每一个系数所对应的 P 值都小于 0.05, 能够通过显著性检验, 所以可以建立 GARCHM 模型作为原模型的修正。接着对建立的 GARCHM(1, 1)模型进行残差 ARCH 效应检验, 观察拟合该模型后的序列是否还存在 ARCH 效应。滞后阶数分别选取 1, 4, 8, 12 以滞后 4 阶为例, 输出结果见表 3-6。

表 3-6 残差 ARCH 效应检验表

Heteroskedasticity Test: ARCH			
F-statistic	1.594424	Prob.F(4,583)	0.1742
Obs*R-squared	6.362787	Prob.Chi-Square(4)	0.1736

滞后 4 阶时, F 值为 0.1742; 在各阶下, F 值均大于 0.05, 不显著, 说明模型已经不存在 ARCH 效应了。

### (3) 建模结果分析

对所选取的时间段内的上海黄金交易所日收盘价的数据建模分析表明:

a. 对数收益率时间序列的均值方程是一个白噪声过程, 而其残差能用 GARCHM(1, 1)模型进行较好的拟合。

b. 收益率序列呈现明显的高峰厚尾和集群现象, 存在异方差性, 说明风险对收益率的影响不显著。



c.模型中非对称系数为-1.2354，是很大的，并且各系数均通过了显著性的检验，这就说明数据波动性是有长期的记忆性的，也就是说历史价格的波动和它长期的价格波动的大小均是有一定关系的。同时这也表明了黄金市场的灵敏性是比较高的，意思是各投资者对市场中的各类信息的反应是很剧烈的，在实际操作中，他们买入和卖出的速度很快，这同时也非常符合黄金延期交收业务的期货特点。

d.模型中的参数 $-1.2354 < 0$ ，说明信息的作用不是对称的，有一定的杠杆效应，也就是说市场受负冲击要比正冲击引起更大的波动，即“涨得慢，跌得猛”。

### 3.2.4 ARIMA-GARCH 族预测模型的建立

由以上分析我们得出，黄金价格数据的均值部分模拟 ARIMA(1, 1, 1)是比较好的，然而进过一系列的筛选其尾部特征显示 GARCHM(1, 1)模型对数据模拟较贴切。所以我们最终建立模型结果为：

$$r_t = 302.32 - 0.3003 * dH_{t-1} + 0.2148 * \varepsilon_{t-1} + \varepsilon_t \quad (3.2)$$

$$\varepsilon_t = \sqrt{h_t} e_t \quad (3.3)$$

$$h_t = 0.33 + 0.089\varepsilon_{t-1}^2 + 0.879\delta_{t-1}^2 \quad (3.4)$$

### 3.2.5 预测结果分析

ARIMA 模型的预测就是根据黄金价格收盘价时间序列的过去数据，运用 ARIMA 预测模型对之后一段时间的黄金价格进行推断，现在 ARIMA 建模方法在很多领域都得到了应用，之所以会这样很大程度上是由于它在预测方面的有效性，尤其是在短期预测的方面。但随着预测时间序列的延长，所预测的数值的方差会变得越来越来大。ARIMA 模型的预测结果与实际值的偏差也会越来越大，所以 ARIMA 模型更适用与序列的短期预测。在这里，我们用第二小节估计出的模型做短期预测，结果如下图 3-14 所示。其中，实线为点预测，阴影部分为预测的置信区间。由前面的分析我们知道黄金价格数据的周期为 244，所以我们将预测未来 244 期的黄金价格数据，并将其与之前建模数据进行结合，得到预测分析图如 3-14：

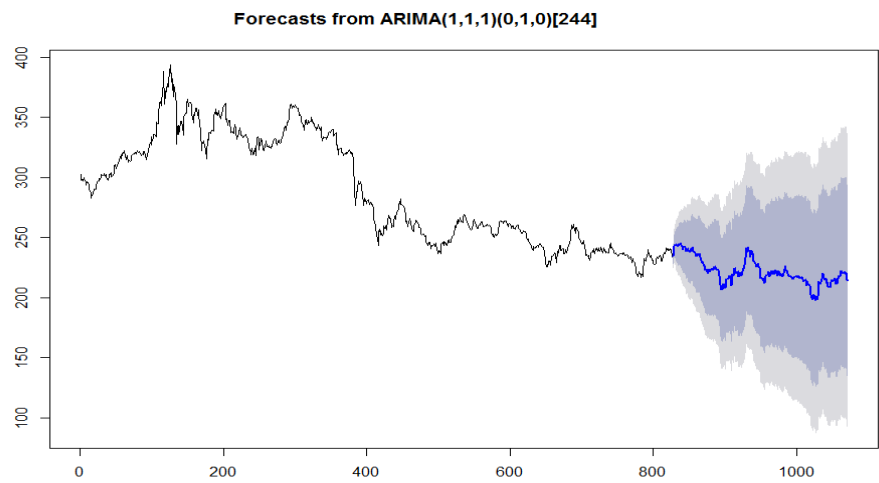


图 3-14 ARIMA 模型预测结果图

由图 3-14 可以看出，预测数据和建模数据的走向大致是相同的，并且整体来看预测误差的置信区间也是可以接受的，所以预测结果还是很好的。

3.3 基于支持向量机的黄金价格预测分析

本文中主要利用支持向量机进行黄金价格时间序列的预测检验。在用 R 语言进行数据分析时，分别给各变量一个代码以便分析时变量在软件中的显示。在具体的数据分析过程中，支持向量机的建立过程为了最终预测结果的精确性，首先要进行模型参数的选择：不同的参数对应会产生不同的训练误差，经过一系列的试验，得到如表 3-7 的结果。

表 3-7 支持向量机模型参数的选择

参数选择	prob=c(0.7, 0.3)	c(0.8, 0.2)	c(0.6, 0.4)	c(0.9, 0.1)
支持向量个数	39	44	42	46
E	0.1	0.1	0.1	0.1
训练误差	0.002648	0.002933	0.003311	0.002818

由表 3-7 可以看出，选择参数为 prob=c(0.7, 0.3)时训练误差是最小的，此时训练误差为 0.002648。所以在后续建立支持向量机模型的过程中选择参数为(0.7, 0.3)。即本部分在数据选择时，选取样本数据的 70%为训练样本，30%为测试样本。

在选定最优参数以后就可以来估计支持向量机模型的参数最优解得到结果 3-8：

表 3-8 模型估计值

	Gamma	Cost	模型误差	离差	$\epsilon$	支持向量
参数最优解	1e-05	1e+05	1.8899	0.4823	0.1	8

根据表3-8利用训练数据建立的模型的估计结果,可以得到测试数据的预测数值,此预测数值在图中表现为图 3-15:

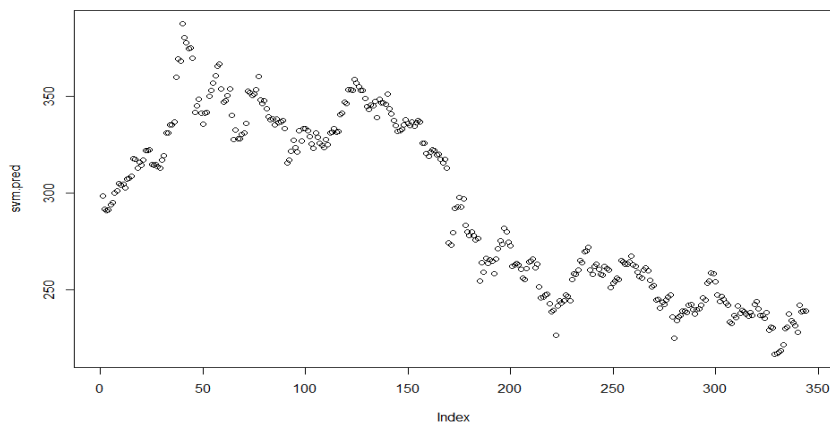


图 3-15 测试数据的预测值

只根据图形 3-15 来判断预测效果是不明显的,所以下面我们将预测数据与实际的测试数据进行一个连接,以便在同一幅图中出现,如图 3-16 所示:

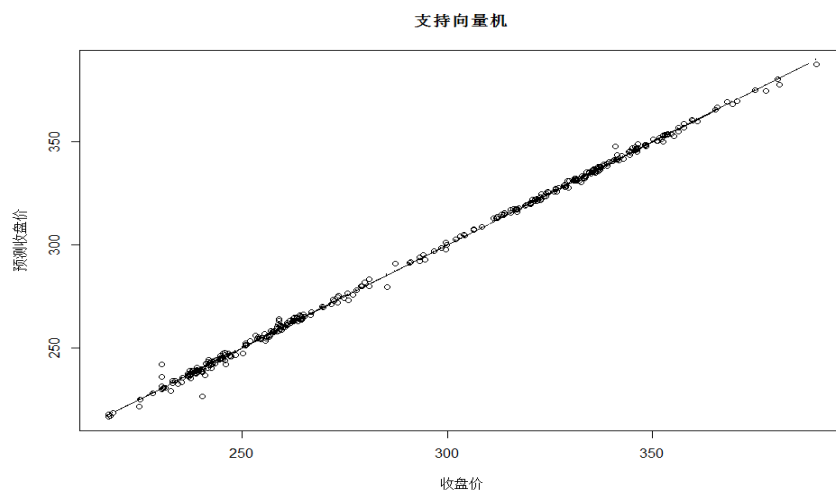


图 3-16 预测值与实际值的散点图

由图 3-16 可以看出,预测值与实际值基本上在一条直线上,除了部分异常点以外,所以整体来说该预测还是比较精确的,预测效果是很好的。

### 3.4 结果对比分析

#### 3.4.1 去异常点前后的 ARIMA-GARCH 模型预测结果对比分析

由上文分析我们知道,去除异常点之后所建 ARIMA-GARCH 的结果我们已经知道。在此我们利用未去除异常点的数据做进一步的分析,所得结果与之前的结果进行

对比如表 3-9:

表 3-9 不同均值方程的 GARCH(1, 1) 模型拟合结果

	AIC	BIC
ARMA(1, 1) (去除异常点)	4.8704	4.9048
ARMA(1, 1) (未去除异常点)	4.8715	4.9059

由表 3-9 可以看出, 去除异常点之后所建 ARIMA-GARCH 模型的 AIC 和 BIC 准则降低了一点, 所以说去除异常点之后模型模拟的更好了点。模型的估计结果对比如表 3-10。

表 3-10 模型系数结果

	mu	ar1	ma1	omega	alpha1	beta1
系数 (未去异常点)	302.32	-0.3003	0.2148	0.33	0.089	0.879
P 值	0.0000	0.0000	0.0745	0.0449	0.0029	0.0000
系数 (去异常点)	302.30	1.00	-0.067	0.25	0.076	0.901
P 值	0.0000	0.0000	0.0254	0.0584	0.005	0.0000

由表 3-10 对比结果发现, 除去异常点之后数据的系数虽然变化不大, 但是其 P 值均显著减小了, 说明模型的显著性提高了。

### 3.4.2 去异常点前后的支持向量机模型预测结果对比分析

本小节中分别利用去除异常点和不除异常点的数据来建立支持向量机模型, 得到的模型估计值分别为:

表 3-11 模型估计值

	Gamma	Cost	模型误差	离差	$\varepsilon$	支持向量
参数最优解 (去除异常点)	1e-05	1e+05	1.8899	0.4823	0.1	8
参数最优解 (未去除异常点)	1e-03	1e+02	3.9511	1.9072	0.1	17

由表 3-11 可以看出, 去除异常点以后模型的估计误差变小了, 离差值也变小了, 所以说异常点对数据还是有较大影响的, 下面将两种数据条件下得到的模型进行黄金价格的预测, 得到预测值后, 将两组值进行一个链接, 绘制散点图如 3-17。

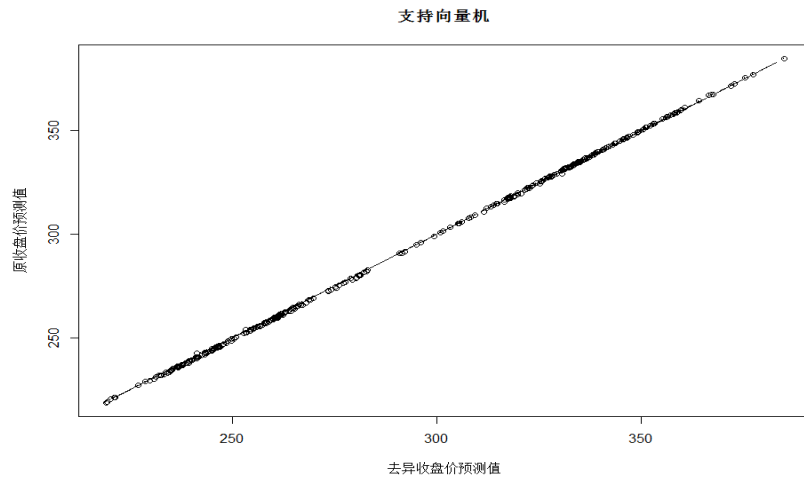


图 3-17 去异收盘价预测值与原收盘价预测值的散点图

由图 3-17 可以看出，去异收盘价预测值与原收盘价预测值大体在一条直线上，虽然模型误差值变小了，但是预测值却大致相同。整体来说，预测效果都是很好的，比较精确的。

### 3.4.3 ARIMA-GARCH 模型和支持向量机模型预测结果对比分析

由上一节的内容可以知道支持向量机模型下的黄金价格预测值，我们将预测值与真实值做一个链接，得到数据的对照表如表 3-12。

表 3-12 支持向量机模型的预测值与真实值对照表

序号	预测值(p)	实际值(r)	序号	预测值	实际值	序号	预测值	实际值
5	298.69	298.53	124	319.53	319.2	218	347.71	348.47
14	291.58	291.01	131	330.98	331.55	220	350.57	350.99
16	290.88	287.3	135	331.26	329.7	227	353.93	354.6
26	291.40	290.95	137	335.32	334.5	231	340.22	339.31
28	293.94	293.4	140	335.16	334	234	327.69	329.42
29	295.11	294.1	142	336.67	336.4	237	332.63	333.2
36	300.23	299.79	147	359.78	360.87	240	328.07	328.78
39	301.25	299.72	149	369.25	368.2	245	328.24	328.72
40	305.04	304.05	154	368.35	369.37	247	330.45	332.58
50	304.42	303.15	156	387.46	389.8	249	331.08	332.05
53	304.72	304.3	158	380.53	380.45	250	336.13	335.5
58	302.57	302.16	169	377.58	380.9	261	352.60	355.1
60	307.25	306.47	171	374.65	377.5	267	351.94	351.74
61	307.61	306.54	173	375.08	374.98	270	350.30	351.2
66	308.85	308.54	176	369.80	370.63	272	351.01	352.16
72	317.81	317.48	185	341.60	342.91	274	353.29	353.65
74	317.39	316.07	187	345.15	346.22	277	360.16	359.79

表 3-12 支持向量机模型的预测值与真实值对照表（续表）

序号	预测值(p)	实际值(r)	序号	预测值	实际值	序号	预测值	实际值
81	313.04	311.39	190	348.38	348.29	283	348.33	348.09
86	316.07	317	192	341.43	340.8	284	346.12	346.01
90	314.54	313.7	194	335.80	336.74	285	347.87	340.99
92	317.12	315.5	195	341.22	341.46	290	343.73	341.35
100	321.95	321.38	196	341.78	341.5	293	339.31	339
111	322.00	320.67	197	350.12	352.5	296	337.92	336.77
113	322.48	321.78	198	353.07	352.97	302	338.50	339.04
116	314.89	313.95	199	356.85	356.34	305	335.40	335.4
117	314.36	313.62	204	360.73	359.49	308	338.33	337.97
120	314.70	313.33	206	365.55	365.2	310	336.60	335.43
121	313.70	312.27	210	366.79	365.7	311	336.67	335.55
122	313.01	312.13	214	353.83	353.7	320	337.58	336.46
123	317.20	316.51	216	346.93	346.37	322	333.49	332.84

在具体的建模过程中，我们随机抽取了 70% 的黄金价格数据作为训练集，训练处支持向量机模型，利用剩余的 30% 数据（即 344 个样本数据）作为测试集数据，并将预测值和实际值进行了对比。由于篇幅发原因，这里我们只是展示部分预测值与实际值的对照图，从表 3-12 得到，支持向量机预测的最大误差率为： $\max(\frac{|r_i|-|p_i|}{|r_i|})=4.7765\%$ 。

其均方误差  $RMSE=\sqrt{\frac{(X_1-\hat{X}_1)^2+(X_2-\hat{X}_2)^2+...+(X_n-\hat{X}_n)^2}{n}}=\sqrt{\frac{166.8279}{90}}=1.3615$ 。为了做预测模型效果的比较，下面我们将 ARIMA-GARCH 模型得到的预测值和实际值做一个展示，如表 3-13。

表 3-13 ARIMA-GARCH 模型的预测值与真实值对照表

天数	实际值 r	预测值 p	天数	实际值 r	预测值 p	天数	实际值 r	预测值 p
828	230.93	235.07	857	242.9	228.67	887	258.39	231.13
829	229.5	237.06	858	243.1	227.26	888	259.16	229.72
830	225.9	234.88	859	239.09	227.12	889	257.5	226.47
831	225.38	233.95	860	239.3	227.93	890	253.19	233.68
832	225.3	233.86	861	240.35	228.51	891	257.94	232.73
833	230.71	236.14	862	240.93	236.92	892	258.26	233.71
834	227.79	230.61	863	240.25	236.00	893	254.7	233.85
835	230.39	231.64	864	236.95	233.40	894	255.26	233.67
836	228.89	233.40	865	236.57	231.87	895	254.93	229.92
837	227.2	235.38	866	236.69	231.99	896	249.74	230.03
838	234.57	238.30	867	238.2	233.87	897	250.28	232.21
839	234.4	240.63	868	239.96	233.51	898	248.88	232.09
840	235.2	240.50	869	237.42	234.23	899	245.9	229.97
841	233.8	243.85	870	240.59	235.49	900	247.6	230.36

表 3-13 ARIMA-GARCH 模型的预测值与真实值对照表（续表）

天数	实际值 r	预测值 p	天数	实际值 r	预测值 p	天数	实际值 r	预测值 p
842	235.41	243.48	871	240.6	235.72	901	248.6	231.16
843	237.15	244.16	872	242.5	233.05	902	247.23	231.87
844	237	239.97	873	243.14	230.95	903	244.52	230.56
845	237.39	239.07	874	242.75	230.17	904	244.59	230.82
846	236.42	236.32	875	242.92	230.20	905	244.83	232.79
847	235.5	236.13	876	245.8	229.02	906	247.39	231.58
848	228.84	234.28	877	247.38	231.63	907	245	230.67
849	238.86	234.28	878	247.7	228.53	908	244.18	230.72
850	237.65	234.76	879	245.01	229.42	909	243	230.97
851	238	232.07	880	252	228.75	910	242.31	230.88
852	239	232.87	881	255.8	228.57	911	237.4	229.95
853	236.99	230.38	882	255.99	230.63	912	235.45	231.50
854	240.38	229.37	883	260.29	232.13	913	235.3	230.43
855	246.15	230.60	884	257.28	232.18	914	235.79	229.16
856	243.17	228.84	885	259.8	232.76	915	234.64	225.77
			886	260.8	230.96	916	234.85	222.94

在具体的建模过程中，我们利用前 927 个数据进行了建模，并且往后预测了 244 个黄金价格数据，与实际值进行了对比。同样由于篇幅的原因，这里我们只是展示部分预测值与实际值的对照图，从表 3-13 得到，ARIMA-GARCH 模型预测值的最大误差率为： $\max(\frac{|r_i|-|p_i|}{|r_i|})=13.7102\%$ 。其均方误差  $RMSE=\sqrt{\frac{(X_1-\hat{X}_1)^2+(X_2-\hat{X}_2)^2+\dots+(X_n-\hat{X}_n)^2}{n}}=\sqrt{\frac{18831.554}{89}}=14.5462$ 。综合各计算结果对比如表 3-14 所示。

表 3-14 模型预测结果对比

	支持向量机	ARIMA-GARCH
最大误差率	4.7765%	13.7102%
均方误差（RMSE）	1.3615	14.5462

综合对比来说，支持向量机的预测效果比 ARIMA-GARCH 模型的预测更准确一点。

### 3.5 本章小结

本章通过研究发现：（1）在进行异常点的检测时可以通过集中不同的方法进行，并且所得结论大致相同。但是，在实际的应用过程中还应该注意一下异常点所拥有的相对性和主观性特点，检测者在对异常点进行挖掘研究时必须依据所选数据以及研究问题的一些特定的挖掘背景，对最终探测到的异常的数据集的确定通常要首先通过一

些特定的算法确定范围，然后再结合领域专家的意见方可最终确定。(2) 对数收益率时间序列的均值方程是一个白噪声过程，而其残差能用  $APARCH(1, 1)$ 模型进行较好的拟合；收益率序列呈现明显的尖峰厚尾和集群现象，存在异方差性，说明风险对收益率的影响不显著，波动性是有长期的记忆性的。在对原数据以及除去异常点的数据分别进行分析的结果看，异常点的存在对于黄金价格预测模型的精度有一定的影响。

(3) 在分别对相同的一组黄金价格数据进行支持向量机和  $ARIMA-GARCH$  模型建模的研究中发现支持向量机的预测效果是比  $ARIMA-GARCH$  模型的预测效果要好的。



## 第4章 黄金价格影响因素分析

通过以上对黄金价格波动率的分析已经大概了解了黄金价格的一些特性,所以接下来对于影响黄金价格的一些因素的研究显得尤为重要。本章在收集了国内所有相关研究文献,在研究已有影响因素(主要是货币供应量、利率、汇率、通货膨胀率等)和金融和社会大环境实际环境基础上,引入基于百度指数的媒体关注度以及投资者关注度数据,立足于用户情绪的角度对黄金价格的影响因素进行分析。其中,媒体关注度和投资者关注度分别用媒体指数和用户搜索量指数来表现,所需要数据和指数可以从上海黄金交易所官网、金裕黄金官网、百度指数官网搜集到。然后,讨论各影响因素与黄金价格之间是否相关及其相关程度;另外建立回归模型判定各因素与黄金价格的关系。

### 4.1 变量选择

本节重点从理论与实践两个方面讲述了影响黄金价格的影响因素,并进行了分析研究。经过研究黄金价格的影响因素,本文发现影响黄金价格的主要因素是黄金供给和需求,还可以是汇率、利率、国际政治形势等因素。对于中国的黄金价格波动的影响因素,除了注意以上几个因素还必须结合国内的实际情况分析,黄金在中国也受到自己的供给和需求的影响,国内货币政策,中央银行的黄金储备,通货膨胀,市场投机的动力学因素是国内黄金价格。而且本文认为用户情绪也是影响黄金价格的重要因素,所以引入基于百度指数的媒体关注度以及投资者关注度数据,立足于用户情绪的角度对黄金价格的影响因素进行分析。

#### (1) 黄金供求

黄金虽然具有特殊的属性,但既然称之为商品,它就会具备一般商品的那些特征,因此,供给和需求是决定黄金价格变动的最基本的决定因素。黄金交易市场上体现的供应主要来源于:一是经常性供给;二是诱发性供给;三是调节性供给。而,黄金市场中的需求则来自于:一是黄金的实物需求;二是外汇储备中的黄金储备;三是黄金的投资需求;四是黄金的投机需求。

#### (2) 用户指数和媒体指数

百度指数是立足于百度中的大数据,在百度指数官网上不仅可以对关键词的搜索热度进行研究,还能够对于市场需求、舆情信息等多方面的数据特征进行深层次的挖

掘。百度指数官网的数据每天不定期更新，还可以找到自 2006 年 6 月至今任意时间段的 PC 端搜索数据以及 2011 年 1 月至今的移动端无线搜索数据。百度指数可以反映一种需求，一种主动获取的需求，一切能够对居民的搜索行动产生影响的活动均会映射到百度指数的具体数据中。

而黄金价格的用户指数表示网民在百度上的关键词搜索量为数据基础，本文中以“黄金价格”这一关键词为统计对象，科学的分析并计算出了在百度网页搜索中“黄金价格”搜索频次的加权。根据搜索来源的不同，搜索指数分为移动搜索指数和 PC 搜索指数。另外，媒体指数表示各大互联网媒体报道的新闻中，与“黄金价格”相关的一些搜索，并被百度新闻频道收录的数量，采用新闻标题涵盖关键词这一标准来统计，数据的具体来源、计算的方法和搜索指数最终的数值没有直接的联系。分析二者与黄金价格变动之间的关系还是很有意义的。

### （3）银行间拆借利率（SHIBOR）

银行间拆借利率是指利用银行间拆借市场利率。它使用报价系统，根据贷款利率，涉及银行贷款品种提供每一段每一天，公布的加权平均报价处理，每个周期的平均贷款利率。利率作为影响黄金价格的因素引入模型，因为利率可以用作衡量其他资金的价格，利率波动是代表其他资产，利率波动，黄金的价格有一定的关系。模型选择的利率是中国银行间的隔夜贷款利率为。贷款利率是银行间市场的基础上，确定资本供给和需求之间的关系，它可以比较灵活地反映资金供求的变化情况，是最具代表性的短期金融市场利率和其他短期借款利率通常是银行间拆借利率加上一定的保证金。

CHIBOR 是我国货币市场中形成的利率，其能够很好地体现资金供求关系，目前最能代表我国的市场利率。

### （4）上证综合指数（SH）

股票市场和黄金市场是中国金融市场不可分割的一部分，作为两个重要的投资产品，必须有一定的相关性。本文选择重要的国内股票指数：上证综合指数。这是因为这两个股票交易所和深圳有高度的相关性，在规模和交易活动，上海证券交易所更具代表性，加上证综合指数能够描绘了上海证券交易所交易的实际情况，用 Index 来表示。

### （5）货币供应量（M2）

M2 是对中国黄金现货价格具有影响的几个主要的宏观影响因素之一。货币政策以降低存款准备金率、降低再贴现率、在公开市场上购买证券业务的最好方法是增加

货币供应，以货币政策和在其他条件不变的情况下，现货金价格会上涨，货币政策在更高的提高存款准备金率，再贴现率和出售证券的公开市场操作等都是减少货币供应的方式，采取的货币政策和其他条件不变的情况下，现货黄金价格将会下降。由于金融危机的爆发，各国政府采取了应急救援的政策。 $M2$  能够全面概括货币供应的情况，所以本文将采用广义货币供应量来诠释中央银行中货币的供给。

#### (6) CPI 指数

CPI 指数是一个与通货膨胀水平相联系的指标，而通货膨胀对黄金价格又有一定的影响，通货膨胀也是国内宏观经济影响现货黄金价格的因素，一方面，通货膨胀表现为货币购买力下降，从而导致黄金价格上升；另一方面，不适度的通货膨胀会导致居民担心我们的经济前景，同时也会引起居民担心钱的购买力，也增加了具有良好的保护功能的黄金财富的价值，黄金需求将增加，促使黄金价格的上涨。所以我们将此引进影响因素的体系内。许多历史事实证明，黄金是对抗通货膨胀最有效的工具，当有迹象表明，通胀来了，人们将争夺黄金和黄金，为了保值、增值，黄金的价格上涨在这一时期。所以，通货膨胀率的波动与黄金价格的波动之间存在着很大的联系，同时也可以说明 CPI 指数引入的必要性。

## 4.2 数据选取与描述分析

### 4.2.1 数据选取

分析以前学者的研究发现，目前对于黄金价格的影响因素的分析基本上都是基于各变量的月度数据进行的，并且所分析的变量也大都是那些典型的宏观变量如利率、货币供应量、CPI 等，本章在此基础上引入了新的影响黄金价格的因素，即基于百度指数的用户搜索量和媒体搜索量，立足于投资者的情绪因素进行分析。新引入的变量为日数据，但是一些典型的宏观影响因素如利率、货币供应量、CPI 等只能搜索到月度数据，所以本文将所有的指标数据都转换成月平均值进行分析。另外，由于 CPI 指数只能找到 2015 年 4 月份的数据，所以本章数据均采用从 2011 年 1 月至 2015 年 4 月的数据。本文在具体的分析过程中，首先像其他学者一样利用月度数据进行了分析，对于日数据进行了月平均值的处理。又由于本章引入的新变量是日数据，平均化处理以后使其失去了引入的价值，所以本章后续只分析了能够找到日数据的一些影响因素，最后对结果进行了对比分析。

### （1）用户指数和媒体指数指标

本章所引入的这两个新的变量的原始数据来自百度指数官网，由于本章分析的是中国黄金价格的主要影响因素，因此本章选择了百度指数，而不是谷歌趋势。另外，百度指数的搜索关键词是“黄金价格”。在百度指数的官网输入“黄金价格”后会出现一系列基于黄金价格的数据，包含整体趋势、PC 趋势、移动趋势条件下的搜索指数和媒体指数，由于投资者对于黄金市场的关注既有 PC 端也有移动端，所以我们选取整体趋势下的搜索指数和媒体指数引入模型。至于日期的选取，由于百度官网上对于“黄金价格”只能搜索到 2011 年 1 月 1 号至今的数据，所以我们所有的变量在日期选取时均从 2011 年 1 月 1 号开始。

### （2）银行间拆借利率指标

由于本章首先对黄金价格的各影响因素进行了月平均数据的分析，之后为了体现用户指数和媒体指数的作用，本章着重分析了各变量的日数据对黄金价格日收盘价的影响，所以我们引入的利率指标是 shibor 利率，而不是统计年鉴上的人民币存贷款利率。数据也是从 shibor 官网上下载得到。

### （3）其他指标

就中国黄金市场而言，由于上海黄金交易所的 AU99.99 的交易量是最有代表性的，所以本节分析黄金价格影响因素时利用 AU99.99 的日收盘价数据来代表我国的黄金价格，其单位是元/克，2011 年 1 月 4 号到 2015 年 11 月 2 号的数据均来源于锐思数据库的每日收盘价。另外，由于 M2 现在是中国人民银行对于中国经济进行宏观跳空的猪腰参考制表，所以用 M2 来代表货币供应量，数据亦来源于锐思数据库。对于上证综合指数、CPI、汇率等变量的数据也都是下载于锐思数据库。

## 4.2.2 描述分析

本章采用基于百度指数的用户搜索量、媒体搜索量、银行间拆借利率、中国人民银行货币供应量 M2、居民消费价格指数 CPI、汇率、上证综合指数这七个变量作为分析影响黄金价格变化的解释变量。其中 M2 和 CPI 是月度数据，其余为日数据，这些变量在所选时段的走势图如 4-1、4-2 所示：

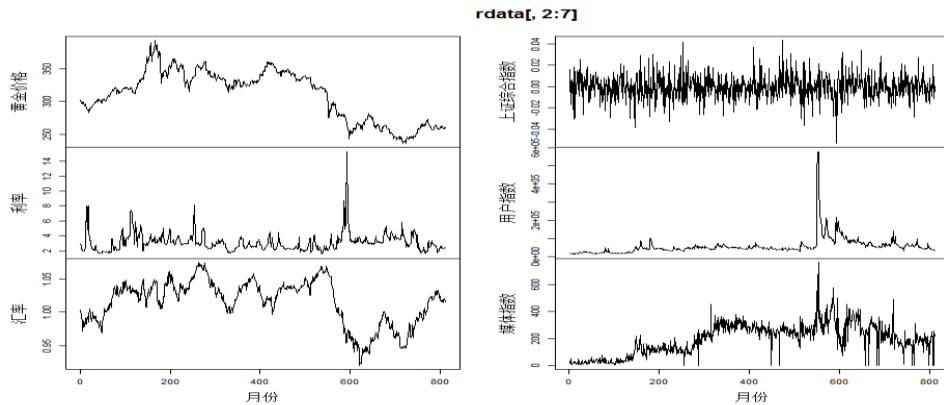


图 4-1 从 2011 年 1 月 04 号到 2014 年 5 月 16 号五个独立变量的走势图

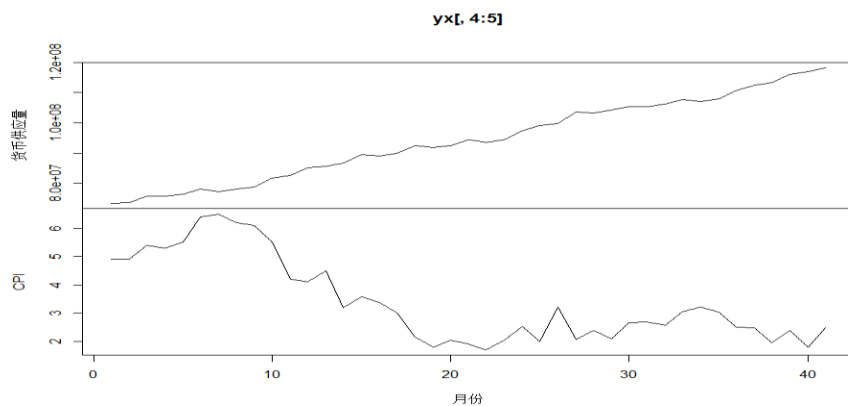


图 4-2 从 2011 年 1 月到 2014 年 5 月 CPI 和 M2 的走势

由图 4-1 可以看出黄金价格的走势一直处于变动幅度较大的状态，其中 2011 年底 2012 年初的时候黄金价格达到最大值。大概从 2013 年初开始，黄金价格呈现明显的下降趋势，之后虽有上升的动势，但是幅度很小。通过黄金价格的分解图可以看出，黄金价格的长期趋势是下降的，并且具有一定的季节性趋势，随机性趋势也很明显。并且由图 4-1 大致可以看出六个指标均不是平稳的时间序列，只有上证综合指数可能是平稳的，我们需要采取一定的措施将数据进行平稳化。另外，如图 4-2 所示，月度变量 CPI 和 M2 的走势变动也很大，所以在后期建模的过程中要进行差分平稳化。

### 4.3 基于 VAR 模型的黄金价格影响因素分析

#### 4.3.1 基于月度数据的分析

##### (1) 相关性分析

在对数据进行建模之前，首先对数据进行基本的分析，了解各变量之间的相关性，使得所建模型能够更准确的反映各解释变量对黄金价格的影响。表 4-1 是各个变

量之间的相关系数，由表中数据可以看出，各变量之间的相关系数中有较大的，所以我们认为各变量之间可能存在多重共线性，为了克服多重共线性的影响，我们采用逐步回归法重新建立回归模型。

表 4-1 从 2011 年 1 月到 2014 年 5 月各变量相关系数

变量	利率	M2	CPI	上证综指	汇率	用户指数	媒体指数
利率	1.00						
货币供应量	-0.06	1.00					
CPI	0.26	-0.78	1.00				
上证综指	-0.32	0.01	-0.04	1.00			
汇率	-0.25	-0.37	0.12	0.01	1.00		
用户指数	0.13	0.49	-0.42	-0.14	-0.19	1.00	
媒体指数	-0.03	0.62	-0.79	-0.10	-0.16	0.66	1.00

将黄金价格对每一个解释变量进行回归，得到 7 个方程的 $R^2$ 以及各变量的系数和 P 值，依据变量的重要性并比较各变量的回归结果中的 $R^2$ 以及 t 值的绝对值，可知首先选择的是黄金价格与用户指数的回归模型作为基础模型。将其余变量逐个引入基础模型，估计结果如表 4-2 所示。

表 4-2 黄金价格数据逐步回归的结果

	用户指数 ( $X_1$ )	媒体指 数( $X_2$ )	CPI ( $X_3$ )	M2 ( $X_4$ )	汇率 ( $X_5$ )	利率 ( $X_6$ )	$R^2$
$f(X_1)$	-9.34E-05 (6.8E-05)						0.93
$f(X_1, X_2)$	-7.72E-05 (7.9E-05)	-0.02 (0.06)					0.93
$f(X_1, X_2, X_3)$	-9.8E-05 (8.0E-05)	0.005 (0.06)	5.69 (3.39)				0.937
$f(X_1, X_2, X_4)$	-9.75E-05 (8.3E-05)	-0.008 (0.06)		-2.43E-06 (1.6E-06)			0.936
$f(X_1, X_2, X_4, X_5)$	-8.66E-05 (8.8E-05)	-0.003 (0.06)		-2.33E-06 (1.7E-06)			0.938
$f(X_1, X_2, X_4, X_6)$	-0.0001 (8.1E-05)	-0.002 (0.06)		-2.36E-06 (1.6E-06)	101 (134)		0.943
$f(X_1, X_2, X_4, X_7)$	-8.87E-05 (8.6E-05)	-0.02 (0.06)		-2.45E-06 (1.7E-06)		0.88 (2.17)	0.938

从表 4-2 可以看出，在基础模型中引入媒体指数和货币供应量之后 $R^2$ 有所提高，变量系数估计值的符号合理且都能通过 t 检验，因此保留媒体指数和 M2。在此基础上引入其他变量所得的模型的估计结果中他们的系数估计值不能通过 t 检验，因此模型中不再引入这些变量。所以最终我们保留解释变量用户指数、媒体指数和 M2。

### (2) 单位根检验

单位根检验结果如表 4-3 所示 (Augmented Dickey- Fuller 检验和 Phillips-Perron 检验)。所有的变量都是  $I(1)$ ，没有长期记忆。进行一阶差分以后，黄金价格、用户指数、媒体指数和 M2 的数据都变成平稳且没有长期记忆。

表 4-3 所有变量的单位根检验

变量(P 值)	Different	ADF 检验	PP Test	Order
货币供应量	Level	0.99	0.064	I(1)
	First	0.02	0.01	
用户指数	Level	0.2846	0.05332	I(1)
	First	0.01	0.01	
媒体指数	Level	0.5384	0.7779	I(1)
	First	0.01	0.01	
黄金价格	Level	0.4837	0.778	I(1)
	First	0.01	0.01	

### (3) 实证结果

#### ①未引入新变量的实证结果

利用黄金价格、M2 的对数差分数据建立 VAR(1)模型，得到结果如表 4-4:

表 4-4 VAR(1)模型显著性检验

R <sup>2</sup>	0.9187
误差项	11.1662
F-统计量	96.0365
AIC	7.7829
SC	7.9961

由表 4-4 可知利用黄金价格和 M2 所建立的向量自回归模型的  $R^2$  为 0.9187，说明模型的拟合优度为 91.87%，AIC 和 SC 分别为 7.7829 和 7.9961 都是很小的，极大似然估计量为 -146.7659，说明我们的模型整体来说是很好的。

#### ②引入新变量的实证结果

在上述建模引入变量的基础上，引入新的变量基于百度指数的用户搜索量和媒体搜索量，建立 VAR(1)模型，得到结果如表 4-5 所示。

表 4-5 VAR(1)模型显著性检验

R <sup>2</sup>	0.9364
误差项	10.5096
F-统计量	55.2531
AIC	7.7416
SC	8.1255

由表 4-5 可知利用黄金价格、媒体指数、用户指数和 M2 所建立的向量自回归模

型的 $R^2$ 为 0.9364, 说明模型的拟合优度为 93.64%, 相比较未引入新变量之前有所提高。AIC 和 SC 分别为 7.7416 和 8.1255, 均有所下降。极大似然估计量为-141.9618, 比未引入前增加了约 4.8, 整体来说引入新的变量之后模型拟合的更好一些, 所以, 下面我们将引入新变量之后的模型结果展示如表 4-6。

表 4-6 黄金价格影响因素的 VAR(1)模型输出结果

变量 (-1)	M2	媒体指数	用户指数	C
系数	-2.43E-06	-0.0075	-9.75E-05	56.9771
P 值	1.6E-06	0.0568	8.3E-05	48.9579
T 值	-1.5257	-0.1327	-1.7185	1.1638

因此, 我们可以根据表 4-6 的结果将建立的黄金价格影响因素的向量自回归模型表示如下:

$$\text{LnH} = 56.9771 + 0.7989\text{LnH}_{t-1} - 2.43\text{E} - 06\text{LnM}_{t-1} - 0.0075\text{LnT}_{t-1} - 9.75\text{E} - 05\text{LnY}_{t-1} \quad (4.1)$$

其中, H 表示黄金价格、M 表示 M2、T 表示媒体指数、Y 表示用户指数。根据方程可以看出各变量相比黄金价格自身对价格的影响是最大的, 其余变量的影响程度依次为媒体指数、用户指数, M2 的影响是最小的。另外, 黄金价格自身对黄金价格的影响是正向的, 而媒体指数、用户指数和 M2 与黄金价格之间具有负向相关关系。

#### 4.3.2 基于日数据的分析

##### (1) 相关性分析

在对数据进行建模之前, 首先对数据进行基本的分析, 了解各变量之间的相关性, 使得所建模型能够更准确的反映各解释变量对黄金价格的影响。表 4-7 是各个变量之间的相关系数, 由表中数据可以看出, 各变量之间的相关系数只有用户指数和媒体指数的大一些。

表 4-7 从 2011 年 1 月 4 号到 2014 年 5 月 16 号各变量相关系数

变量	利率	汇率	上证综指	用户指数	媒体指数
利率	1.00				
汇率	-0.18	1.00			
上证综指	-0.018	0.011	1.00		
用户指数	0.069	-0.15	0.01	1.00	
媒体指数	-0.034	-0.13	-0.01	0.52	1.00

我们可以利用辅助回归法进一步验证多重共线性的存在。将每个解释变量对其他解释变量进行回归, 考察他们的可决系数 $R^2$ 。为节省篇幅这里只给出辅助回归的 $R^2$ 及



由此得到的方差膨胀因子 VIF，如表 4-8 所示。其中汇率的方差膨胀因子大于 10，说明多重共线性将会影响回归的结果。

表 4-8 辅助回归的 $R^2$ 值和 VIF

被解释变量	利率(X1)	汇率(X2)	上证综指(X3)	媒体指数(X4)	用户指数(X5)
$R_j^2$	0.7441	0.9764	0.1127	0.7168	0.8983
$VIF_j(\frac{1}{1-R_j^2})$	3.9078	42.3729	1.127	3.5311	9.8328

则我们将黄金价格对每一个解释变量进行回归，得到 5 个方程的 $R^2$ 以及各变量的系数和 P 值，依据变量的重要性并比较各变量的回归结果中的 $R^2$ 以及 t 值的绝对值，可知首先选择的是黄金价格与媒体指数的回归模型作为基础模型。将其余变量逐个引入基础模型，估计结果如表 4-9 所示。

表 4-9 黄金价格数据逐步回归的结果

	媒体指数 (X <sub>1</sub> )	用户指数 (X <sub>2</sub> )	上证综指 (X <sub>3</sub> )	汇率 (X <sub>4</sub> )	利率 (X <sub>5</sub> )	$R^2$
$f(X_1)$	-0.0018 (0.035)					0.9913
$f(X_1, X_2)$	-0.0022 (0.002)	-3.14E-06 (8.1E-06)				0.9913
$f(X_1, X_2, X_3)$	-0.0023 (0.002)	-1.91E-06 (8.1E-06)	-7.4827 (11.2744)			0.9914
$f(X_1, X_2, X_4)$	-0.0021 (0.002)	-3.54E-06 (8.0E-06)		90.72 (24.25)		0.9915
$f(X_1, X_2, X_5)$	-0.0023 (0.002)	-2.14E-06 (8.1E-06)			-0.2657 (0.2037)	0.9914

从表 4-9 可以看出，在基础模型中引入用户指数之后 $R^2$ 有所提高，变量系数估计值的符号合理且都能通过 t 检验，因此保留用户指数。在此基础上引入其他变量所得的模型的估计结果中他们的系数估计值不能通过 t 检验，因此模型中不再引入这些变量。所以最终我们保留解释变量用户指数、媒体指数。

## (2) 单位根检验

表 4-10 所有变量的单位根检验

变量(P 值)	Different	ADF 检验	PP Test	Order
用户指数	Level	0.2846	0.05332	I(1)
	First	0.01	0.01	
媒体指数	Level	0.5384	0.7779	I(1)
	First	0.01	0.01	
黄金价格	Level	0.4837	0.778	I(1)
	First	0.01	0.01	

单位根检验结果如表 4-10 所示(Augmented Dickey- Fuller 检验和 Phillips- Perron 检验)。所有的变量都是 I(1)，没有长期记忆。进行一阶差分以后，黄金价格、用户指数、媒体指数的数据都变成平稳且没有长期记忆。

### (3) 实证结果

经过上述分析，我们知道最终建立 VAR(1)模型时只保留了用户指数和媒体指数。建模结果如表 4-11 所示。

表 4-11 VAR(1)模型显著性检验

R <sup>2</sup>	0.9913
误差项	3.5589
F-统计量	15325.3
极大似然估计量	-2174.072
AIC	5.3854
SC	5.4260

由表 4-11 可知利用黄金价格用户指数、媒体指数所建立的向量自回归模型的R<sup>2</sup>为 0.9913，说明模型的拟合优度为 99.13%。AIC 和 SC 分别为 5.3854 和 5.4260，说明模型拟合较好，下面我们将模型结果展示如表 4-12。

表 4-12 黄金价格影响因素的 VAR(1)模型输出结果

变量 (-1)	黄金价格	媒体指数	用户指数	C
系数	0.9564	-0.0022	-3.14E-06	0.9508
P 值	0.0367	0.002	8.1E-06	1.1430
T 值	26.0624	-1.0897	-0.3893	0.8319

因此，我们可以根据表 4-12 的结果将建立的黄金价格影响因素的向量自回归模型表示如下：

$$\text{LnHJ} = 0.9508 + 0.9564\text{LnHJ}_{t-1} - 0.0022\text{LnM}_{t-1} - 3.14\text{E} - 06\text{LnY}_{t-1} \quad (4.2)$$

其中，HJ 表示黄金价格、M 表示媒体指数、Y 表示用户指数。根据方程可以看出各变量相比黄金价格本身的变化对其影响是最大的，黄金价格上一时刻每变动 1%，当前黄金价格变动 95.64%。其余变量的影响程度依次为媒体指数和用户指数。另外，黄金价格自身与黄金价格具有正向相关关系，媒体指数和用户指数对黄金价格的影响均是负向的。

### 4.3.3 对比分析

#### (1) 基于月度数据的新变量引入前后对比分析

在利用月度数据进行建模分析时，经过一系列的检验与剔除，最终我们保留了变

量 M2，在此基础上我们又引入了新的变量媒体指数和用户指数，研究发现新变量的引入并不影响原变量的显著性，所以后续的分析我们将三个变量均留在了模型中。所建模型输出结果对比如表 4-13。

表 4-13 基于月度数据的结果对比图

变量 (-1)	黄金价格	M2	媒体指数	用户指数		
引入前系数	0.9602	-1.21E-06				
引入后系数	0.7989	-2.43E-06	-0.0075	-9.75E-05		
	R <sup>2</sup>	误差项	F-统计量	极大似然估计量	AIC	SC
引入前模型检验	0.9187	11.1662	96.0365	-146.7659	7.7829	7.9961
引入后模型检验	0.9364	10.5096	55.2531	-141.9618	7.7416	8.1255

由表 4-13 可以看出，引入新的变量之后黄金价格自身的影响由 0.9602 降为 0.7989，下降了很多，M2 对黄金价格的影响变化不是很大，另外引入新的变量后模型的拟合优度有所提高，并且误差项减少，说明引入新变量后模型整体效果有所改善。

#### (2) 基于月度数据与日数据的模型结果对比

研究以前学者文献发现，大部分学者在研究黄金价格的影响因素时针对各宏观变量以及黄金价格指标，分析时均利用的月平均数据。本章中引入的新变量媒体指数和用户指数立足于用户每天的情绪变化对黄金价格的影响，所以在分析的过程中我们采用了日数据，这种改变给模型造成的影响我们做了一个对比分析如表 4-14：

表 4-14 月度数据与日数据的模型结果对比

变量 (-1)	黄金价格	媒体指数	用户指数	C		
月度数据	0.8942	-0.0181	-7.72E-05	39.5797		
日数据	0.9564	-0.0022	-3.14E-06	0.9508		
	R <sup>2</sup>	误差项	F-统计量	极大似然估计量	AIC	SC
月度数据	0.9303	10.6555	71.1984	-143.7579	7.7312	8.0298
日数据	0.9913	3.5588	15325.3	-2174.07	5.3854	5.4260

由表 4-14 可以看出，利用日数据进行分析时，和月平均数据相比只有常数项的系数变化比较大，其余各变量的系数变化并不大。但是，就模型整体的效果来看，利用日数据所建立的模型的拟合优度较月平均数据来说有所提高，并且其误差项下降很多，AIC、SC 准则也都有明显的下降，所以利用日数据对黄金价格的影响因素进行分析效果更好一些。

## 4.4 本章小结

本章通过建立黄金价格与各指标之间的向量自回归模型表明：（1）黄金价格本身就具有比较大的波动性，这一波动对黄金价格总水平本身就有一定的影响。不管利用月平均数据还是利用日数据进行分析，发现黄金价格自身的影响都是最大的。可能是由于黄金价格的波动属于集聚性的，致使黄金价格波动的强度大小直接影响黄金价格总水平的高低。（2）利用对比的方法对黄金价格的影响因素进行了全面的分析，发现不管在何种条件下我们引进的新的变量基于百度指数的用户搜索量和媒体搜索量对黄金价格均是有影响的，并且影响都是负向的。说明用户搜索量越大，黄金价格越低，媒体关注度越高，黄金价格越低。（3）通过对比分析发现，引入新的变量媒体指数和用户指数之后，模型的整体效果变好。

## 第5章 结论和建议

本章主要是对前面几个章节理论和实证分析所得到的结果进行一个总体的分析与展示,以便能够更好的展现本文的研究所发现的一些问题以及提炼的一些重要信息。并且,根据研究所得的结果提出几个具有针对性意见和建议。

### 5.1 结论

从上文的研究中,我们可以得出,在我们所选时间段内,黄金价格的波动是比较大的,并且其收益率数据存在集群现象,在比较大的波动后往往伴随着另一波大的波动,于是在检验 ARCH 效应后对数据建立了 GARCH 模型,并且根据数据的特点对模型进行了完善。基于原始数据的特点,本文中我们尝试了四种不同的方法来定位异常点对应日期,发现其结果是一致的。接着在对黄金价格预测的模型中发现利用支持向量机和 ARIMA 模型的预测结果大致也是相同的,说明短期预测中二者都是可取的。最后对黄金价格影响因素的分析发现引入的新的变量对黄金价格还是有一定的影响的,但是影响较大的还是我们常见的利率等影响因素。

从以上模型的分析结果能够分析出以下结论:

第一,异常点检测。通过具体的例子分别详细介绍了利用不同的方法进行异常点检测的过程,可以看出,集中的方法在进行异常点的检测问题时是非常有效的。但是,在实际的应用过程中还应该注意一下异常点所拥有的相对性和主观性特点,检测者在对异常点进行挖掘研究时必须依据所选数据以及研究问题的一些特定的挖掘背景,对最终探测到的异常的数据集的确定通常要首先通过一些特定的算法确定范围,然后再结合领域专家的意见方可最终确定。

第二,用来分析黄金价格波动性的序列是一种白噪声过程,价格的风险值可以用 GARCHM(1, 1)模型进行较好的拟合;收益率序列呈现明显的高峰厚尾和集群现象,存在异方差性,说明风险对收益率的影响不显著;模型中非对称系数为-1.2345,是很大的,并且各系数均通过了显著性的检验,这就说明数据波动性是有长期的记忆性的,也就是说历史价格的波动和它长期的价格波动的大小均是有一定关系的。同时这也表明了黄金市场的灵敏性是比较高的,意思是各投资者对市场中的各类信息的反应是很剧烈的,在实际操作中,他们买入和卖出的速度很快,这同时也非常符合黄金延期交收业务的期货特点;模型中的参数 $-1.2345 < 0$ ,说明信息的作用不是对称的,有一定

的杠杆效应，也就是说市场受负冲击要比正冲击引起更大的波动，即“涨得慢，跌得猛”。另外，在对原数据以及除去异常点的数据分别进行分析的结果看，异常点的存在对于黄金价格预测模型的精度有一定的影响。

第三，分别对相同的一组黄金价格数据进行支持向量机和 ARIMA-GARCH 模型建模。支持向量机预测了 344 个数据，所得预测的最大误差率为 4.7765%，均方误差 RMSE 为 1.3615，而利用 ARIMA-GARCH 模型预测了 244 个数据，所得预测值的最大误差率为 13.7102%，均方误差 RMSE 为 14.5462。所以总体来说根据我们所选时间段的数据得出支持向量机的预测效果是较为精确的，这可能是因为在支持向量机模型中我们引入了外生变量，并且就 ARIMA-GARCH 模型本身特点来说，此模型较适合进行短期预测。

第四，就黄金价格影响因素的 VAR 模型结果来看：（1）黄金价格本身就具有比较大的波动性，这一波动对黄金价格总水平本身就有一定的影响。不管利用月平均数据还是利用日数据进行分析，发现黄金价格自身的影响都是最大的。可能是由于黄金价格的波动属于集聚性的，致使黄金价格波动的强度大小直接影响黄金价格总水平的高低。（2）利用对比的方法对黄金价格的影响因素进行了全面的分析，发现不管在何种条件下我们引进的新的变量基于百度指数的用户搜索量和媒体搜索量对黄金价格均是有影响的，并且影响都是负向的。说明用户搜索量越大，黄金价格越低，媒体关注度越高，黄金价格越低。（3）通过对比分析发现，引入新的变量媒体指数和用户指数之后，模型的整体效果变好。

## 5.2 建议

第一，充分利用市场信息，自主做出决策，避免跟风行为。在我们对黄金价格影响因素的分析中发现，引入的基于百度指数的用户搜索量和媒体搜索量两个立足于用户情绪的解釋变量对于黄金价格的影响程度还是挺大的。因此，投资者应避免时尚，理性看待市场氛围，尽可能充分利用市场信息，学习使用情感，而不是被情绪，在复杂的信息，找到最对自己有用的信息，自由作出决定，相信自己做出决定的能力。

第二，充分利用用户指数，加快市场消化信息的速度，让信息转换更加迅速。正确使用用户指数这种工具，使之发挥有效的作用，例如引导市场走向合适的方向、加快市场消化某些特殊信息的速度、让政府政策更加有效的实施等。

第三，在经济市场中有很多种变量都会影响黄金的价格，因此黄金价格具有了一

种复杂的非线性变化的特点，常用的经典预测模式单一导致结果只能映射部分信息，直接影响就是预测精度较低。为更加准确全面的描述黄金价格变化规律，构建了基于数据挖掘的黄金价格预测组合模型。实证结果表明，支持向量机模型能够使预测精度得到一定的提高，因此也说明支持向量机在未来黄金价格的预测中有着非常广泛的应用前景。

第四，据本文以上的分析，对于黄金价格的影响，立足于用户情绪的新闻媒体是一个很重要的因素，媒体报道有助于将一些特质信息融入黄金价格，由此降低了黄金市场金价的同步性。在政策方面，本文的启示是，正处于转型期的我国应大力推进新闻媒体业的市场化改革，使得媒体信息的信息传递和舆论监督的特点能够得到还原，以便更好的发挥其市场资源配置的功能。

## 参考文献

- [1] 潘贵豪, 胡乃联, 刘焕中, 李国清. 基于 ARIMA-GARCH 模型的黄金价格实证分析[J]. 黄金, 2010, 31(1) : 5-8
- [2] 曾黎, 李春. 融合线性、非线性模型的黄金价格预测模型研究[J]. 黄金, 2013, 34(3) : 7-10
- [3] 赵庆, 王志强. 基于 HP 滤波 AR-GARCH 族模型对黄金价格预测研究[J]. 黄金, 2014, 35(3) : 4-8
- [4] 魏宇, 黄登仕, 王建琼, 等. 我国黄金现货市场的动态 VaR 预测模型研究[J]. 管理评论, 2010, 22(8) : 30-38
- [5] 曾濂, 马丹頔, 刘宗鑫. 基于 BP 神经网络改进的黄金价格预测[J]. 计算机仿真, 2010, 27(9) : 200-203
- [6] 孙晓莹, 李晓静. 数据挖掘在股票价格组合预测中的应用[J]. 计算机仿真, 2012, 29(7) : 375-378
- [7] 蒋翠清, 梁坤, 丁勇, 等. 基于社会媒体的股票行为预测[J]. 中国管理科学, 2015, 23(1) : 17-24
- [8] 徐海鹏. 基于关联规则的股票预测方法研究[J]. 计算机与数字工程, 2010, 245(3) : 150-153
- [9] 张延利. 新马尔科夫模型的黄金价格短期预测[J]. 黄金, 2012, 9: 6-8
- [10] 兰秋军, 马超群, 甘国君, 等. 中国股市弱有效吗?----来自数据挖掘的实证研究[J]. 中国管理科学, 2005, 13(4) : 17-22
- [11] 张其联, 胡波, 叶思雄. Markov 机制转换模型在黄金价格波动中的研究[J]. 2011, 33(3) : 66-67
- [12] Shafiee S, Topal E. An overview of global gold market and gold price forecasting[J]. Resources Policy, 2010, 35(2) : 178-189
- [13] Ismail Z, Yahya A, Shabri A. Forecasting Gold Prices Using Multi-ple Linear Regression Method[J]. American Journal of Applied Sciences, 2009, 33(6) : 1509-1514
- [14] Peter Z.G. Time series forecasting using a hybrid ARIMA and neural network model [J]. Neuro computing, 2003, 50: 159-175
- [15] Forrest C, Mills T.C., Wood G. Gold as a Hedge against the US Dollar. Working Paper, World Gold Council, 2004
- [16] Brian M.L., Tully E.S. risk and return in daily COMEX gold and silver data 1982-2002 [J]. Applied Financial Economics, 2006, 16: 4, 319-333
- [17] Graham S. The Price of Gold and Stock Price Indices for the United States. The World Gold Council, 2001
- [18] Levin, Ghosh E.J., MacMillan D.P., Wright R.E. Gold as an Inflation Hedge? Studies in Economics and Finance. 2004, 22: 1-25
- [19] Lawrence C. Why is gold different from other assets, An empirical investigation. The World Gold Council, 2003



- [20] Ranson.D. Is The Superior Predictor of Inflation. Working Paper, World Gold Council, 2005
- [21] Harmston S. Gold a sastore of value. The World Gold COUNCIL, 1998
- [22] 傅瑜. 近期黄金价格波动的实证研究[J]. 产业经济研究, 2004, 01: 30-40
- [23] 杨柳勇, 史震涛. 黄金价格的长期决定因素分析[J]. 统计研究, 2004(6) : 21-24
- [24] 刘曙光, 胡再勇. 黄金价格的长期决定因素稳定性分析[J]. 世界经济研究, 2008, 02: 35-41
- [25] 谢朝阳. 国际黄金价格长期变迁中的美元因素分析[J]. 中国流通经济, 2010, 24(8) : 70-73
- [26] 邱济洲, 额日敦塔娜. 基于 VAR 模型的黄金价格影响因素分析[J]. 财经理论研究, 2014, 06: 10-18
- [27] 郑秀田. 基于 GARCH-M 模型的黄金市场风险与收益关系研究[J]. 黄金, 2008, 29(5) : 4-7
- [28] 范为, 房四海. 金融危机期间黄金价格的影响因素研究[J]. 经济与金融, 2012, 03: 8-16
- [29] 闵嘉宁. 影响黄金价格波动的原因分析与投资策略研究[J]. 生产力研究, 2011, 10: 61-62
- [30] 张继德, 廖微, 张荣武. 普通投资者关注对股市交易的量价影响---基于百度指数的实证研究[J]. 会计研究, 2014, 8: 52-59
- [31] 杨虎, 李强. 金融时序中异常数据挖掘算法设计及实证分析[J]. 中国管理科学, 2003, 12(3) : 7-11
- [32] 陈晓. 我国货币政策对股票市场协整关系实证研究[J]. 时代金融, 2015, 577(1) : 154-155
- [33] 吕书然. 货币政策变动及其对股票市场的影响效应探析[J]. 时代金融, 2015, 589(5) : 116-117
- [34] 王相宁, 张思聪. 基于汇改视角的人民币汇率与黄金价格相关性研究[J]. 财政金融, 2015, 2:1-3
- [35] 周云丽, 汪金菊. 基于分位数回归的国内黄金价格影响因素分析[J]. 大学数学, 2014, 30(4) : 33-40
- [36] 周舞舞. 黄金价格影响因素的实证分析[J]. 生产力研究, 2013, 05: 54-58

## 致谢

首先，衷心的感谢在论文的选题、思路、整改过程中我的导师赵霞教授给予我的帮助，在数据的整理与分析思路以及论文形成以后的完善和最终定稿方面，赵老师都曾耐心以及细心的给我了很多可行性的建议，另外在论文做不下去时赵老师也给我提供了很多新的思路。就是在赵老师悉心的教导下，我才得以完成我的论文。不仅如此，在日常的生活中赵老师也教授了我很多的人生道理，鼓励我勇于追求自己想要的，这对我的现在以及未来的学习、工作方面都有很大的帮助。

其次，我非常感谢曾经教授过我以及给予我很多帮助的代课老师，在石玉峰老师、林春艳老师、金玉国老师、戴洪帅老师、田金方老师、裴海峰老师、李忆唯老师、孔令涛老师、杨冬梅老师、倪洁老师以及其他老师的身上学到了很多知识，论文最终能够顺利完成也是由于在校期间从这些老师的课堂上学到了很多实用的知识和有用的技巧。我从老师们身上学到的东西将终身受益。同时非常感谢张健老师、许薇老师、程燕群老师、赵世民老师以及其他办公室的老师在生活中和学习中给予我的帮助。

再次，对于我的同师门李行同学我也要表示由衷的感谢，研究生两年的时间里，我们朝夕相处，不管是在生活上还是在学习上，她都给予了我很多的帮助，我从她的身上学到了很多令我受益匪浅的知识，也让我明白了友情的可贵之处。

最后，我一定要感谢永远在背后给予我支持的家人，正是有了家人的支持，我才能走到今天，顺利毕业走向人生的下一个阶段。