

分类号 _____
UDC _____

密级 _____
编号 10741



硕士学位论文

论文题目 沪深 300ETF 期权套利策略的研究
——基于不同波动率模型

研究生姓名: 谢彬

指导教师姓名、职称: 陈芳平 教授

学科、专业名称: 应用经济学 金融工程

研究方向: 金融投资

提交日期: 2021 年 5 月 25 日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的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。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： 谢彬 签字日期： 2021.5.25

导师签名： 陈芳年 签字日期： 2021.5.25

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意” / “不同意”）以下事项：

1. 学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；
2. 学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 谢彬 签字日期： 2021.5.25

导师签名： 陈芳年 签字日期： 2021.5.25

A study on option arbitrage strategies of CSI 300ETF based on three kinds of volatility models

Candidate : Xie bin

Supervisor: Chen fangping

摘 要

2019 年 11 月 8 日,中国证监会正式宣布启动有关沪深 300ETF 期权的试点工作,这是我国为做好股指期货开拓工作的一项重大举措。2019 年 12 月 23 日,沪深 300ETF 期权在上交所上市。沪深 300ETF 期权的问世加速了我国股指期货方面的发展,更意味着我国的交易所在体系完备,功能健全,运行稳定,效率提高等方面有了明显的提高。截至 2020 年末,上交所 ETF 期权合约累计成交 9.82 亿张,上证 50ETF 期权合约全年累计成交 5.19 亿张,沪深 300ETF 期权合约累计成交 4.64 亿张,尽管沪深 300ETF 上市时间只有一年,但其交易活跃程度与 2015 年上市的上证 50ETF 期权一样活跃,这说明沪深 300ETF 在推出后深受广大投资者的青睐。尤其在疫情期间,由于中国防疫措施到位,中国金融市场能平稳运行,规模稳步增长,全球投资者都将沪深 300ETF 期权作为良好的风险管理工具。在期权市场上,我国期权交易中个人投资者和机构投资者大致各占一半。在期权市场上,个人投资者利用期权主要是用来实现增强收益和进行方向性交易,很少去做套利交易。如何给个人投资者提供简单,方便,有效的套利策略是一个值得探究的问题。

本文通过从 Wind 金融数据库中获取的沪深 300ETF 数据和沪深 300ETF 2020 年认购期权的历史数据,建立 GARCH 族类模型,进行预测到期日之前五个交易日的波动率,并将其年化得到年化后的波动率。隐含波动率模型通过建模日相关数据计算认购期权的隐含波动率。高频数据模型先参数检验确定已实现波动率的模型在通过模型计算年化波动率。三种预测波动率与标的 5 日历史波动率差异并不大,说明策略持有期内波动率变化很小,故选择正向买进蝶式套利策略进行套利。通过蒙特卡洛模拟进行对沪深 300ETF 资产价格进行预测,得到一个在行权当日置信区间为 90%的沪深 300ETF 的价格预测区间,根据正向买进蝶式套利策略的损益表计算年收益和年收益率,并在不同的开平仓时机都进行回测。

通过对比不同开平仓时机的回测结果发现,策略构建的时机不同会导致蝶式套利策略的效果。就总的收益率来说,运用隐含波动率模型进行套利的效果最好,收益率达到 12.04%。并提出在期权产品创新,期权交易知识的普及和完善风险预警体系的建议。

关键词: GARCH 族模型 高频数据模型 隐含波动率 期权策略

Abstract

On November 8, 2019, China Securities Regulatory Commission (CSRC) officially announced the launch of the CSI 300ETF options pilot work, which is a major measure for China to do a good job in the development of stock index options. On December 23, 2019, CSI 300ETF options were listed on the Shanghai Stock Exchange. The emergence of CSI 300ETF options has accelerated the development of stock index options in China, which means that China's exchanges have significantly improved in terms of complete system, sound function, stable operation and efficiency improvement. By the end of 2020, a total of 982 million ETF option contracts on the Shanghai Stock Exchange, 519 million ETF option contracts on the Shanghai Stock Exchange and 464 million ETF option contracts on the Shanghai Stock Exchange and Shenzhen Stock Exchange have been traded for the whole year. But it has been as active as the options on the Shanghai 50ETF, which debuted in 2015, suggesting that the CSI 300ETF has been popular with investors since its launch. Especially during the epidemic period, due to China's anti-epidemic measures in place, China's financial market can operate smoothly and its scale grows steadily. Global investors all take the CSI 300ETF options as a good risk management tool. In the option market, the individual investors and the institutional investors account for about half of the options transactions in China. In the option market, individual investors use options mainly to enhance returns and make directional trades, and seldom do arbitrage trades. How to provide simple, convenient and effective arbitrage strategies for individual investors is a problem worth exploring.

Based on the data of CSI 300ETF and the historical data of CSI 300ETF 2020 call options obtained from the Wind financial database, this paper establishes the GARCH family model to predict the volatility of the five trading days before the expiration date, and then annualizes it to get the annualized volatility. Implied volatility model calculates the implied volatility of call options by modeling the relevant data of the day. The high-frequency data model is firstly tested to determine the realized volatility of the model before calculating the annualized volatility of the model. There is no significant difference between the three predicted volatility and the underlying 5-day historical volatility, indicating that volatility changes are very small during the holding period of the strategy, so the positive buying butterfly arbitrage strategy is chosen for arbitrage. Through Monte Carlo simulation, the asset price of Shanghai and Shenzhen 300ETF is predicted, and a price prediction range of Shanghai and Shenzhen 300ETF with confidence interval of 90% on the exercise right day is obtained. According to the income statement of the forward buying butterfly arbitrage strategy, the annual return and annual return rate are calculated, and the backtest is conducted at different opening and closing opportunities.

By comparing the backtest results of different opening and closing opportunities, it is found that different timing of strategy construction will lead to the effect of butterfly arbitrage strategy. In terms of the total return rate, the best effect of arbitrage is to use the implied volatility model, and the return rate reaches 12.04%. And put forward in the option product innovation, the option trading

knowledge popularization and consummation risk early warning system suggestion.

Keywords: GARCH family model; High-frequency data model; Implied volatility; Option strategy

目 录

| | |
|-------------------------------------|----|
| 1 引 言 | 1 |
| 1.1 选题的背景及意义 | 1 |
| 1.1.1 选题背景..... | 1 |
| 1.1.2 研究意义..... | 2 |
| 1.2 国内外文献综述 | 4 |
| 1.2.1 关于 GARCH 族类模型的文献综述..... | 4 |
| 1.2.2 关于隐含波动率模型的文献综述..... | 6 |
| 1.2.3 关于高频数据波动率模型的文献综述..... | 8 |
| 1.3 研究路径与方法 | 10 |
| 1.3.1 研究路径..... | 10 |
| 1.3.2 研究方法..... | 11 |
| 1.4 论文的创新与不足 | 12 |
| 1.4.1 论文的创新..... | 12 |
| 1.4.2 论文的不足..... | 12 |
| 2. 波动率模型以及期权套利策略理论 | 14 |
| 2.1 波动率及其模型相关理论 | 14 |
| 2.2.1 波动率相关理论..... | 14 |
| 2.2.2 GRACH 族类模型 | 16 |
| 2.2.3 隐含波动率模型..... | 18 |
| 2.2.4 高频数据模型..... | 18 |
| 2.2 期权交易策略相关理论 | 19 |
| 2.2.1 牛市价差组合..... | 20 |
| 2.2.2 熊市价差组合..... | 20 |
| 2.2.3 跨式组合策略..... | 21 |
| 2.2.4 蝶式价差组合策略..... | 22 |
| 2.3 蒙特卡洛模拟理论 | 24 |
| 3. 沪深 300ETF 期权套利策略的构建 | 25 |

| | |
|-----------------------------------|-----------|
| 3.1 沪深 300ETF 期权现状 | 25 |
| 3.2 套利策略的思路与方案设计 | 26 |
| 3.2.1 套利策略的思路..... | 26 |
| 3.2.2 套利方案设计..... | 26 |
| 3.3 数据的收集和预处理 | 28 |
| 3.3.1 正态性检验..... | 28 |
| 3.3.2 平稳性检验..... | 29 |
| 3.3.3 自相关性检验..... | 30 |
| 3.3.4 ARCH 效应检验 | 31 |
| 3.3.5 构建 GARCH 族模型计算波动率..... | 31 |
| 3.3.6 基于 B-S-M 模型计算隐含波动率..... | 33 |
| 3.3.7 基于高频数据波动率模型计算波动率..... | 34 |
| 3.3.8 预测持有期标的资产价格..... | 36 |
| 3.4 期权策略选择 | 36 |
| 4. 期权蝶式套利的实施方案 | 38 |
| 4.1 基于不同波动率模型的期权蝶式组合回测收益 | 38 |
| 4.1.1 GARCH 族模型的期权套利策略 | 38 |
| 4.1.2 B-S-M 模型下隐含波动率的期权套利策略 | 39 |
| 4.1.3 高频数据模型下的期权套利策略..... | 41 |
| 4.2 基于不同波动率模型的回测结果对比分析 | 42 |
| 5. 结论与建议 | 43 |
| 5.1 结论 | 43 |
| 5.2 建议 | 44 |
| 5.2.1 应加强期权量化投资产品的创新..... | 44 |
| 5.2.2 量化期权交易策略的市场渗透率有望提高..... | 45 |
| 5.2.3 完善风险预警体系, 保障市场平稳运行..... | 45 |
| 参考文献..... | 62 |
| 后 记..... | 66 |

1 引言

1.1 选题的背景及意义

1.1.1 选题背景

期权的出现初衷是满足投资者风险管理的需求。但由于当时监管不到位，市场不成熟，交易方只会对自己明显有利的时候才去选择进行签署期权合约，这使得当时期权流动性很低，期权当时受到了社会的质疑。1973 年，芝加哥证券交易所的出现，将场外期权进行了标准化并在交易所内交易。ETF 期权是基于交易所交易基金（ETF）而专门设立的标准化期权合约。在鼓励金融创新的新环境下，伴随着计算机飞速的发展和政策的宽松，美国各交易所将发展的重心都放在了期权产品种类的创新上，在不断地创新之下，美国期权市场已经是一个非常成熟的期权市场。

期权在欧美市场已是较成熟的产品，但是在中国期权的发展正刚刚起步，我国的期权市场尚在萌芽阶段。上证 50ETF 期权是我国首支场内期权，它标志着我国进入了期权时代。2019 年 12 月 23 日，上交所和深交所上市我国第二支 ETF 期权——沪深 300ETF 期权合约。沪深 300 指数相比较于上证 50 指数更能全面地反映我国两个市场的表现。样本选自两个证券市场，涵盖大多数的市值，成份股是市场代表性好、流动性好、交易活跃的主流资金投入的股票。该指数可以反映市场的主力资金的收益能力。

虽然在我国期权交易才出现不久，但是期权交易的活跃程度却逐年增加，截至 2020 年末，上交所 ETF 期权合约累计成交 9.82 亿张，上证 50ETF 期权合约全年累计成交 5.19 亿张，沪深 300ETF 期权合约累计成交 4.64 亿张，虽然沪深 300ETF 上市时间只有一年，但是其交易量和上证 50ETF 期权一样活跃。尤其在疫情期间，中国防疫措施到位，中国金融市场平稳运行，规模稳步增长，沪深 300ETF 更是受到了全球投资者的青睐，将沪深 300ETF 期权作为风险管理的良好工具。投资者选择期权当做投资工具，因为期权具有以下几个优点。第一，期权的本质是一种对未来的选择权，期权费的本质是买方为了选择权而付出的成本，

这意味着期权买方会拥有无限的收益和有限的亏损，从这方面来讲，期权降低了投资买方的风险。第二，在股票市场上，只能在大涨中做多股票或者大跌中做空股票当中获得收益，但是在期权市场上，投资者不仅可以从大涨大跌的单向走势中获利，而且可以在双边大幅，股价震荡波动中也可以获利。第三，期权交易灵活度高，投资者可以选择多种多样的策略，通过组合认购和认沽期权的数量和方向，构造出满足自己需求的获利策略。第四，期权具有高杠杆行，期权的权利金的成本远远低于持有标的股票的成本。第五，期权交易的手续费也低于股票买卖的手续费。

期权市场上，我国个人投资者的和机构投资者参与交易比例大致相同。从交易目的看，期权主要被应用的目的是增强收益。除了期权的主要作用是增强收益外，机构投资者还使用期权工具以达到套利的目的。个人投资者则是为了利用期权工具进行方向性交易。中国的证券市场参与者具有以下特征：散户投资者数量占比巨大，市场盲目性高，市场巨大的波动性。当市场走低时，负面效应巨大，给监管部门带来了巨大的工作压力和工作难度。目前期权市场还在发展前期，如果监管部门不能及时做好监管工作，期权交易不仅不会对市场起到积极向上的影响，更有可能当市场下跌时，期权会使市场雪上加霜。为了防止期权对中国市场产生不良影响，证监会采用了提高开仓数量，提高保证金率等一系列措施。由于衍生品在中国才刚刚开始发展，衍生品市场还不够成熟，市场参与者的经验和专业程度不够高，金融衍生品无法充分对冲股市中的风险，反而被视为股市下跌的罪魁祸首。作为获利，对冲和投机的工具-期权可以为投资者提供在市场中期权策略的新想法。

1.1.2 研究意义

沪深 300 相比于上证 50 能更全面反映我国两个市场的整体情况。因此，市场上跟踪沪深 300 指数的基金远多于上证 50，截至 2021 年初，市场上跟踪上证 50 的基金有 26 支，资金规模达到 938.82 亿元，跟踪沪深 300 的指数型基金有 172 支，资产规模达到 2838.5 亿远，资金的规模几乎是跟踪上证 50 基金的 3 倍，这足以表现沪深 300ETF 期权具有巨大的发展潜力。

上证 50ETF 期权的上市实现了我国期权市场从零到一的首次跨越，那么沪深

300ETF 期权的上市标志着我国期权种类从单一到繁多的进步，这意味着股票期权多样化的发展，预示着未来会出现更多可以被交易的股票期权种类，投资者可以使用更多的期权设计出更符合投资者需求的期权交易策略，从而进一步推动期权产品的发展，为我国期权市场注入新鲜活力。现在有了沪深 300ETF 期权等相关产品后，投资者可以运用沪深 300ETF 期权工具实现标的资产覆盖面扩大以及风险分散化，进行更有效的对冲风险，实现精准的风险控制。首先，构建标的资产为沪深 300ETF 的期权交易策略，策略的可行性不仅具有实在的现实意义，而且对于以后期权产品的设计和期权市场发展的方向提供了新的思路。其次，有效的套利对于金融市场有着非常重要的现实意义。有效的套利会使金融市场自发的达到均衡状态，市场能稳健，安全及有效的运行。对于投资者来说，制定期权套利策略进行套利可以得到超额收益，并与其他金融工具相结合去实现更多满足投资者的需求的策略。对于衍生品市场来说，套利交易的存在可能会促使更多的市场参与者进入市场寻找利润机会，这提供了更高的流动性，促进了衍生品的合理定价和市场持续良好运行。个人投资者很少利用期权工具进行套利交易。如何给个人投资者提供简单，方便，有效的套利策略是一个值得研究的问题。

本文通过从 Wind 金融数据库中的沪深 300ETF 数据和沪深 300ETF2020 年认沽和认购的历史数据。在 GARCH 族类模型分析中，首先建立分别在正态分布和 t-分布进行下 GARCH 模型，EGARCH 模型和 T-GARCH 模型，通过 AIC 准则选择出每个月的最优模型进行预测，预测日选则在到期日之前五个工作日，预测出这五日的波动率，并将其年化得到年化波动率，下个月再从选择最优模型开始进行操作，即动态 GARCH 族类预测模型。隐含波动率模型则是通过从 Wind 数据库中获取的建模日当日认购期权的数据，通过 B-S-M 公式得到了隐含波动率（已是年化后的）。高频数据模型首先建立 HAR-RV 和 HAR-lnRV 模型进行对比分析，通过对比可决系数选择最终预测模型，从而计算波动率。在得到波动率数据后，将三个预测波动率与标的 5 日历史波动率相比较发现，很适合使用蝶式策略进行套利。进行 1000 次蒙特卡洛模拟得到一个在行权当日置信区间为 90%的沪深 300ETF 的价格预测区间，根据使得 $\max(S_T, X) - \min(S_T, X) + 2c_2 - c_1 - c_3$ ，其中 S 为标的资产预测价格，X 为执行价格，c 为看涨期权的价格，从而进行正向买进蝶式套利策略。在进行正向买进蝶式套利策略实施时，建仓点可以选择在开盘时或者近

收盘时，到期日平仓点也可以选择开盘点和近收盘点两个时间点，故本文会在这四种不同的时点进行回测，从而选出最优的开平仓时机。通过比较在不同模型下，不同开平仓时机的年收益率和年收益，得到最优的投资策略，从而得出最优的预测模型和最优的投资时机。最后总结策略的不足和未来期权市场的展望。

1.2 国内外文献综述

1.2.1 关于 GARCH 族类模型的文献综述

Engle(1982)首次提出了能解决条件异方差的 ARCH 模型，Bollerslev(1986)在前人的基础上提出了 GARCH 模型。GARCH 模型相较于 ARCH 更适合于金融领域的研究，因为其是专门为研究金融数据而设计的计量模型，更适合于进行分析与预测。Kroner (1992) 在其研究中表明当进行样本内测试时，ARCH 族类模型拟合效果都非常显著，但是却没有足够的证据来表明其预测未来波动率的效果与样本内测试效果一样显著。Akgiray(1989)、Franses 和 Van Dijk (1995) 使用样本外的数据对 ARCH 模型的预测能力进行检验，实证结果表明 ARCH 模型的预测能力很差。Nelson D. B. (1991) 提出了符合金融数据尖峰肥尾等特点的 EGARCH 模型。Day 和 Lewis (1992) 通过使用 GARCH 与 EGARCH 模型对标普 500 的研究发现，在用样本内的数据进行测试时，测试的波动率与隐含波动率可以拟合出效果更好的已实现波动率，在使用样本外数据进行回测未来波动率时，发现 GARCH 模型并不显著，但是优于 ARCH 模型。Day 和 Lewis (1993) 运用 GARCH 族类模型对于期货市场进行研究，发现使用 GARCH 族类模型同样优于 ARCH 模型。J Christensen 和 Prabhala(1998) 运用 GARCH 族类模型对标普 500 进行研究得到，隐含波动率之中已经包含 GARCH 族类模型所得到的全部信息。Andersen 和 Bollerslev(1998) 研究发现数据选取日回报数据会带有很多噪音，会影响最后预测的精度，则他们在后来进行对外汇市场进行研究时，多使用日内高频。Blair 等 (2000) 使用日内高频实数代替前人所使用的日数据代入 GARCH 模型去研究标普 500，发现在股票市场使用高频数据所得到的结果精度更高。Blair 等 (2000) 使用 GARCH 模型进行不同时段的预测模拟，研究发现，GARCH 模型在预测 1-5 日的信息，结果比较准确，但是超过这个范围，预测精度便大大降低。此外还得到了另一个结论，

隐含波动率所包含的信息已经包括了 GARCH 模型所预测出来的信息,也就是说 GARCH 族类模型不会得到超出隐含波动率所包含的信息。Martens M. 和 J. Zein (2002) 在研究外汇市场和期货市场时运用了 GARCH 模型,其发现 GARCH 模型在预测波动率时,短期是有效的,长期来看准确度不够高。他们还发现,隐含波动率之所以比 GARCH 模型预测出的结果更有效,是因为市场上大部分投资者都会用 GARCH 模型去做预测分析,这一结果会反映到隐含波动率上,所以隐含波动率的信息包含程度会高于 GARCH 模型所包含的。Filis 和 Floros (2011) 运用 TGARCH 模型和 GED-GARCH 模型研究希腊市场的股票期权,得到了希腊市场在引入股票期权后,市场交易变得更加活跃有效。Hu J. (2016) 通过观察期权推出十年期间,发现期权的出现大大降低了股票市场的信息不对称的问题。

惠晓峰和柳鸿生等 (2003) 为了研究人民币的汇率的问题,使用了 GARCH 模型,结果令人欣喜,用预测结果与现实数据拟合效果优秀,各项误差也较小,由此可以总结, GARCH 模型在预测短期汇率时有效的。Glosten 等 (1993) 尝试去解决金融数据普遍存在的收益率波动不对称的问题,提出了 GJR-GARCH 模型。康建林和朱开永等 (2005) 在预测海尔集团股价的波动性时使用了 GARCH 模型,结论发现若预测期较短时, GARCH 模型能很好地拟好真实数据,一旦预测期过长, GARCH 模型的拟合效果就大大降低。陈芳平和李松涛 (2006) 使用 GARCH 族类模型得出了日经股指期货对日本股市波动率有削弱的作用。黄海南和钟伟 (2007) 在预测上证时使用了 GARCH 族类模型,所得到的结果与已实现波动率为准线进行拟合对比,发现用 GJR-GARCH-T 模型的效果最好。不同于 Martens M. 和 J. Zein (2002) 所得到的结论郑振龙和黄慧舟 (2010) 通过对不同波动率模型所包含信息的研究来比较 GARCH 模型和隐含波动率所包含的信息。研究发现 GARCH 模型在短期内所得到的信息含量大于隐含波动率所包含的信息量,但在长期来看,恰好相反。这就论证了短期内 GARCH 族类模型的预测效果与隐含波动率孰优孰劣,是由不同的市场决定的,但长期而言, GARCH 模型所含的信息不如隐含波动率模型所含的。赵伟雄等 (2010) 以沪铜期货为研究对象,使用 GARCH 族类模型进行预测,实验结果结论证明 FIGARCH 模型预测效果最好, EGARCH 模型效果不佳,表明该市场具有长记忆性。袁臻 (2013) 以恒生指数为对象,比较多种 GARCH 模型和两种波动率模型 (B-S-M 和无模型) 的预测能力,得到了同样的结论,短期内

GARCH 模型所预测的结果优于隐含波动率,但在长期来看隐含波动率的信息要优于 GARCH 族类模型,这与之前学者研究的结论相一致,但不同之处在于,她发现不同模型所得出的结果之间是一种交集的关系,并非是之前学者所认为的包含关系。王天一和赵晓军等(2014)对沪深 300 收益率进行预测,结果发现 Realized GARCH 模型能更好的解释沪深 300 收益率尖峰肥尾的分布,其预测效果优于其他传统的 GARCH 模型。洪晶晶和吴鹏跃(2016)以上证指数真实收盘价为样本,使用两种定义下的波动率使用 GARCH 模型进行预测,结果发现 GARCH(1,1)的解释力度最好,对未来波动率的预测更精确。楼迎军(2003)建立 GARCH 模型并使用上证指数数据对我国股市进行了研究,研究结果表明我国股市存在明显的杠杆效应和羊群效应。张瑶(2020)通过 AHBS 模型、GARCH 模型以及 HESTON 模型对沪深 300ETF 的波动率进行估计,结论发现 HESTON 模型更适合于期权定价。尽管如此,HESTON 模型的效果并不是显著,局限在于数据选取的频率过低,选用日内高频数据可能会有更好的效果。

1.2.2 关于隐含波动率模型的文献综述

隐含波动率模型是使用 B-S-M 模型中,6 个变量分别为期权价格,行权日时标的资产价格,行权价,无风险收益率,到期时间以及波动率。其中,除了波动率以外的所有数据均可以在现实市场中获取,从而倒推出波动率,这种算法所得到的波动率叫做隐含波动率。隐含波动率的实质是利用现实可得的数据去对未来不可知的结果进行的一次预测。自从期权定价公式问世以来,很多学者都着重研究隐含波动率,认为隐含波动率在研究期权中占有非常重要的地位。

Black 和 Scholes(1973)开创性的提出了期权定价模型,该模型被认为是期权研究领域基石的存在,但是其假设过于苛刻,导致其在现实运用过程中存在着困难。其中一个假设,B-S-M 公式中波动率是常数,但是在现实中他却是波动的。很多学者对此问题都进行了深入地研究,结论都表明期权的理论价格与实际价格存在着偏差,从 B-S-M 模型可以直到,标的资产价格,到日期,行权价这些都是已知且固定的,利率在期权的存续期内一般很稳定,由此可以推断出,造成理论价格与实际价格之间有误差的只有波动率。这表示隐含波动率与历史波动率之间是有差异的。根据模型分析,波动率应该是个常数,但是现实中不同行权价

的相同到期日的相同标的的波动率是不一致的,而且离行权日越近越大。学者们将期权价格与隐含波动率之间的关系放在一个坐标里,得到了一个类似微笑的形状,这就是有名的微笑曲线。一个期权会形成一个曲线,多个期权则会形成一个曲面。学者们基于此发现了波动率微笑曲线,他们也在不断地放松 B-S-M 的假设,使其更符合现实,使其有更强的现实指导意义。Lamoureux 和 Lap-strapes(1993)在研究期权交易时,因其使用的研究对象交易量较少,价格不能很好地反映现实,所计算出来的隐含波动率带有很多的噪音。这是由于 Lamoureux 和 Lap-strapes (1993)所研究的是个股期权,个股期权交易并不活跃,很难提出噪音的干扰,基于此等原因,后续学者都转向去研究股指期权的隐含波动率。

Canina 和 Figlewski (1993)研究隐含波动率时,得出结论,隐含波动率对于预测波动率是有贡献的,但是隐含波动率是否含有过去的信息却不能有明确的结论。很多学者在这一点上,都有不同的看法,争议很多。Ederington 和 Guan (2005)在研究隐含波动率时,发现有 Implied Volatility 里没有任何的历史信息。但是 Blair 等(2001)、Szakmary 等(2003)则在研究结果中得出结论,隐含波动率里含有历史信息,且包含了全部的历史信息。John(2008)用数据验证了期权价格与隐含波动率有函数关系,期权这个特性可以被利用来使其成为一个良好的控制风险的工具。

Poon 和 Granger (2005)研究表明在进行期权定价时,波动率是一个非常重要的因素,要想先预测标的资产的价格得先预测期波动率。他通过研究衍生品市场的交易特点,发现市场上看涨期权更受青睐,因为大部分投资者会通过购买看涨期权的方式来对自己的资产进行风险管理,所以市场中认沽期权与认购期权的隐含波动率有着差别。所以,如果期权市场交易有成本的话,看涨期权和看跌期权的隐含波动率在相同条件下也是有所差异的。更有结论表明,通过对美国衍生品交易市场的研究发现,认沽期权与认购期权之间的隐含波动率的差异与已实现波动率之间是有关系的,且存在明显的正向关系,这种关系被定义为跳跃。

骆桦和王爽(2016)以香港恒生指数期权为研究对象,使用真实历史数据分析不同期间的 GARCH 族类模型和隐含波动率的预测能力,并以已实现波动率为基准。研究结果证实,在短期内,GARCH 模型的预测能力更好,在长期,大于一周的期限内,隐含波动率模型的预测效果会更好。因为香港市场发展较完善,市场

成熟, 所以其隐含波动率所含的信息量更多。郑振龙和黄慧舟(2010)同样以香港市场为研究对象, 在短期内。GARCH(1,1)的预测能力最强, 超过一周以上, 一个月以内, 隐含波动率预测能力比 GARCH(1,1)效果更强。因此, 在较长的预测期内, 隐含波动率所包含的信息更加全面。在相对成熟的期权市场, 研究对象的期权交易越活跃, 隐含波动率的预测效果更好。杨晓辉和王裕彬(2019)同时使用 GARCH 模型研究和隐含波动率模型上证 50ETF 的隐含波动率。实证结果表明, 上证 50ETF 波动率存在“尖峰肥尾”, 长记忆性等特点。

1.2.3 关于高频数据波动率模型的文献综述

Anderson 和 Bollerslev(2000)首次提出 Realized Volatility 的概念, Realized Volatility 具有非对称性和长记忆性的特点。Bollerslev 和 Diebold(2001)在 Anderson, Bollerslev(2000)的基础上, 将已实现波动率(RV)进行更加详细地分解, 分为连续和跳跃两种类型。他们使用 HAR-RV-CJ 模型预测金融市场标的资产的波动性, 从而极大地提高的预测的精度。Patton 和 Sheppard(2015)在 HAR-RV 模型的基础上, 按将已实现波动率分解成 upside realized semi-variance 和 downside realized semi-variance, 得到 signed jump variation, 并以此设立了 HAR 类模型, 实证结果表明这三个模型都有较好的预测能力。Luo 和 Chen(2018)使用 HAR 族模型预测波动率, 实证表明 HAR 族波动率预测模型对未来波动率具有较好的预测能力, 且在 HAR-RV 的衍生模型的拟合效果都好于基本模型。

文凤华等(2012), 依托于微观市场理论, 并结合市场的杠杆波动性特点, 设立了 LHAR-RV-V。并使用 LHAR-RV-V 模型对沪深 300 指数未来一段时间期进行实证检验并预测。实验结果表明, 该模型能很好地反映我国股市的长记忆性和杠杆性, 模型效果要强于传统模型。陈浪南和杨科(2013)为了预测中国市场的波动率对传统 HAR 模型进行了改进, 设立了 HAR-CJ-DFIGARCH 模预, 并以上证指数高频数据为测试样本。实验结果表明, 中国金融市场具有长记忆性和结构突变的特点, 以及用样本外数据进行检测后发现改模型的预测精度是最高的。刘广应等(2013)在文章中阐述了五种高频数据模型的理论, 并使用上证指数 1 分钟的高频数据作为样本数据进行实证分析, 表明 LHAR-RV-EVT 预测能力最强。闫会强等

(2017) 使用了多种 HAR 模型对沪深 300 指数五分钟数据和日已实现波动率, 周已实现波动率和月已实现波动率进行了预测, 我国股市表现出一定的长记忆性, 对数形式的高频数据模型拟合效果最好, 预测能力越精确。朱万锐和唐大为(2017) 在使用高频波动率模型对中国市场进行实证分析时, 以已实现波动率为指标, 最后得出高频波动率模型在短期预测能力很强, 但是在长期来看并不适合中国市场的结论。李航和何枫(2017) 投资者想做好投资组合, 风险控制和衍生品定价这一系列都离不开对金融资产波动率的准确预测。作者建立 HAR 模型对其进行实证研究。得出结论, 基于现货市场数据所设立的模型的预测能力要强于基于期货市场数据所设立的模型。刘晓倩等(2017) 将现有的 HAR 模型进行了梳理和整理, 在此基础上, 将市场波动率指数引入其中, 该模型既包含了股票市场的高频信息, 又包含了期权市场的模拟信心, 最大程度的吸取了两个市场的全部信息, 因此, 该模型的预测能力很好, 拟合效果很好。他们在研究沪深 300 指数时, 使用了 GARCH 模型, 隐含波动率模型和 HAR 模型对沪深 300 指数进行了回测, 结果表明基于高频数据下的 HAR 模型的效果最好。张波和蒋远营(2017), 研究了 15 年的高频交易数据研究, 发现在开盘期间回报率显著为负, 收盘前回报率显著为正。由于对随机波动时的高频数据进行了建模对计算机的要求较高, 计算机要处理的数据量异常庞大, 因此对估计方法的进行改进了, 最终得到了合理性更好的模型, 并且该模型能够预测该状态, 从而证明 SVJt-s 模型的效果比其他模型的预测能力更优秀。龚旭和文凤华(2017) 基于 HAR-RV 模型将模型中实现的波动率分为三种不同频率的已实现波动率, 并引入跳跃波动率。然后, 以沪深 300 指数和沪深 300 股指期货为样本, 并以其五分钟交易频率数据为样本数据, 对常见的 HAR 模型进行样本内和样本外的分析, 研究表明, 在 HAR-RV-EMD-J 模型中, 不同频率的波动率所容纳的信息量要大于跳跃波动率和趋势已实现波动率所包含的信息。龚旭和曹杰(2020) 基于高频交易数据的 HAR 模型已经在预测金融市场不同类型波动的研究中显示出良好的预测。本文基于四个经典的 HAR 模型, 这些模型解释了杠杆和结构突变对波动性的预期影响, 创建了四个具有杠杆和结构突变的 HAR 模型。然后, 作为检验高频交易数据的模型, 该模型基于对上述模型进行分析后五分钟内获取的数据, 以检验样本内和样本外, 以检验各个模型的预测市场波动的效果。实证结果表明, 在更有效地预测股市波动方面的模型是已实现波动

性，连续波动性，带杠杆和结构性突变的模型，而跳跃波动率和符号跳跃方差在预测市场波动性方面效果不佳。

1.3 研究路径与方法

1.3.1 研究路径

本文研究路径如下：第一章为引言，其中，第一部分阐述了沪深 300ETF 期权的背景及本文的研究意义，第二部分收集了有关国内外对 GARCH 模型，高频数据模型以及隐含波动率模型学者的研究成果，将其整理为文献综述。第三部分讲述了研究的思路 and 路径，绘制技术路线图并且讲述了研究的方法。第四部分讲述了本文的创新与不足之处。第二章着重介绍理论基础。第一部分相关理论基础，GARCH 模型的理论基础，隐含波动率的理论基础和高频数据模型的理论基础。第二部分介绍了期权交易策略的种类以及其理论基础。第三部分介绍了蒙特卡洛模拟的理论基础。第三章着重介绍套利策略的构建。第一部分介绍了沪深 300ETF 期权合约的基本情况，第二部分套利策略的思路与方案，大致地叙述套利策略的方案设计。第三部分首先进行数据的处理，三种不同的波动率模型对历史数据进行建模。使用构建好的波动率模型去预测波动率，比较三种预测出来的波动率与建模日当日的历史波动率相比较发现，未来五日内波动率的变化不大，遂决定使用正向买进蝶式套利策略进行套利，通过蒙特卡洛模拟方法进行标的资产价格预测，得到一个置信度 90%的预测的价格区间。第四章期权蝶式套利方案的实施。第一部分统计不同投资时间（开仓平仓的时间）下的投资回报结果，从而统计出 2020 年每个月的收益和 2020 年度收益率。第五章为结论。通过进行不同时机的回测结果与不同模型下的回测结果进行对比发现，不同的开仓平仓时机会对策略的收益产生影响。在模型的选择上，隐含波动率模型的收益率最高，达到 12.04%。在未来波动率变化不大的情形下，正向买进蝶式套利策略是有可行性，且有一定的收益。

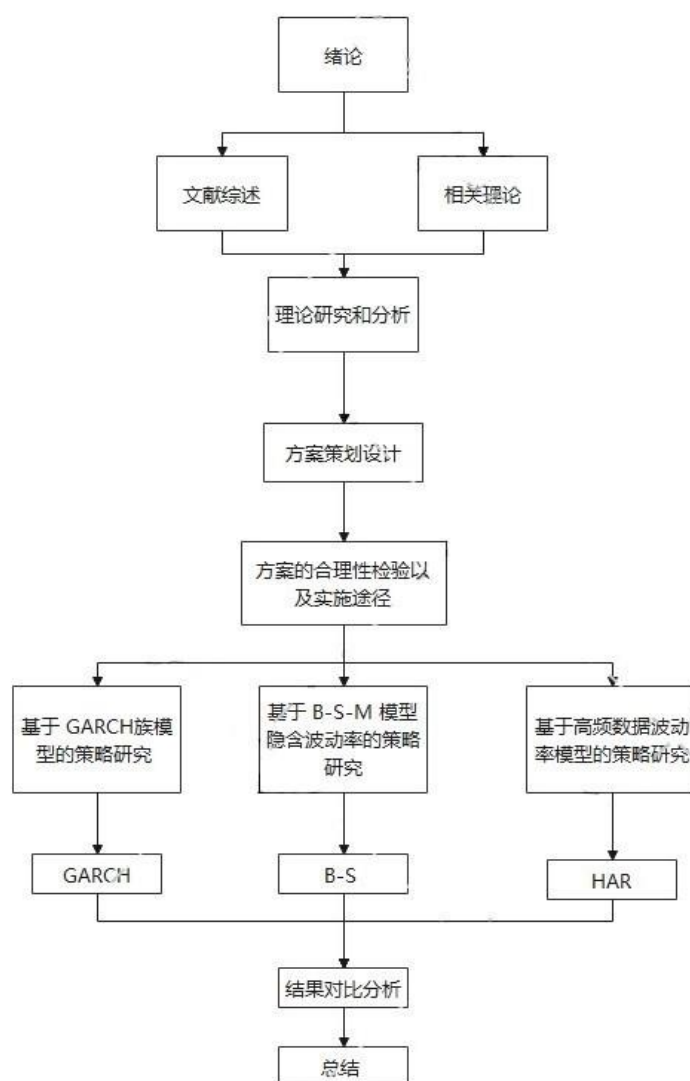


图 1.1 论文技术路线

1.3.2 研究方法

本文主要用到的方法有：

(1) 文献研究法

通过大量阅读国内外学者的研究成果，总结目前学术界主流波动率模型的发展，阐述波动率在金融市场进行投资组合，期权定价以及风险控制的必要性和重要性。初步分析适合我国股市的波动率预测模型和期权套利策略。

(2) 实证研究法

本文根据我国股市的走势选择了合适的套利策略，并通过综合使用 Matlab

2018a 语言、Wind 和 Excel 等软件，对沪深 300ETF 的每日收盘数据分别使用 GARCH 模型，隐含波动率模型和高频数据波动率模型。并将得到的波动率进行年化后，使用蒙特卡洛模拟进行对沪深 300ETF 资产价格进行预测，得到一个在行权当日置信区间为 90%的沪深 300ETF 的价格预测区间以及均值，并进行期权套利策略的实施，比较了不同时机下不同模型下的收益情况。

(3) 比较分析法

本文在对波动率进行预测时，使用了三种不同的波动率模型，GARCH 模型，IV 模型和高频数据波动率模型，将不同模型下所模拟到的价格区间进行回测，比较不同模型下策略的效果以及相同模型下不同开仓平仓的策略效果。

1.4 论文的创新与不足

1.4.1 论文的创新

中国期权市场的投资者偏好主观经验判断和基本面分析，加上期权交易制度的限制导致投资者十分缺乏基于模型预测并有稳定收益的交易策略。本文设计的套利方案可以为市场投资者提供新的策略思路，为中国商品期权市场的金融创新做出一份贡献。

在众多现存研究中，以宽跨式套利为主，本文通过上述的论述，首先进行构建波动率模型计算出预测波动率，分析预测到的三种波动率与五日的历史波动率相比，走势贴合与历史波动率变化不大，依次为依据选择了合适的期权套利策略——正向蝶式套利策略。

在沪深 300ETF 期权推出的一年后，结合期权最新的交易数据和沪深 300ETF 的相关数据，经过模型回测给予了投资者一个方便快捷，操作难度低且可行性高，获利能力尚可的期权套利策略，具有很强的时效性。

1.4.2 论文的不足

沪深 300ETF 期权是我国 2019 年年底新上市的期权品种，上市时间较短，可供研究的数据量不太多，且在期权上市初期，可能会存在市场参与者数量较少、流动性不足等情况，这些因素均有可能影响实证结果。伴随着沪深 300ETF 期权

上市时间越来越久，数据量也越来越大，对其套利的研究应该更加成熟。

在对沪深 300ETF 进行计算隐含波动率时，其理论基础基于经典的 B-S-M 期权定价模型，B-S-M 模型虽然是一个非常经典的模型，在衍生品研究中是举足轻重的地位，但是其对使用条件比较苛刻，因能力有限，本文并没有选择其他替代 B-S-M 模型来进行分析预测。

2. 波动率模型以及期权套利策略理论

2.1 波动率及其模型相关理论

2.2.1 波动率相关理论

波动率是对资产价格不确定性的衡量，是衡量标的资产风险的一项指标。沪深 300ETF 的波动率就是沪深 300ETF 价格变化的程度，价格变化幅度越大，波动率越大，波动率越大，针对波动率进行获利的可能性也越大。但是在现实中，我们并不以价格的变化来作为波动率的定义，而是选用标的资产收益率的变化程度来衡量波动率。这是因为标的资产价格是非负的，且价格是绝对数，并不具备可比性。收益率可正可负，具有可比性，且收益率的分布很好研究。那么沪深 300ETF 的波动率则会被定义为沪深 300ETF 收益率的标准差。计算波动率，那么沪深 300ETF 的波动率就被定义为对数收益率的标准方差。现代金融理论中，波动率是投资者进行投资组合，风险控制和期权策略交易的核心。波动率我们可以从历史数据中计算出来，也可以应用著名的 B-S-M 期权定价公式，将标的资产价格，行权价，期权价格，无风险收益率和到期日等数据从市场中搜集到，反推出波动率。尤其对于衍生品市场而言，投资者和交易员进行期权交易的本质就是对波动率进行研究，很多期权策略都是基于对波动率准确的预测而制定的。市场上所有的期权交易的基石都是研究波动率。波动率对衍生品市场的影响意义非常重大。

因为本文研究对象为沪深 300ETF 期权，对于衍生品市场来说，所有的交易策略都是基于对波动率的准确预测，因此有必要首先介绍波动率的理论基础和分类，为后续的研究奠定扎实的基础。

波动率会具有自相关性，即如果之前是高波动率，那么之后一段时间也会是高波动率；反之，依然如此。波动率具有“惯性”。其次，一般情况下波动率的变化是连续的，但是在特殊的时候比如危机，波动率会跳跃。再者，波动率是均值复归的，即波动率的变化是在一个均值上上下下一个有限的范围内进行运动的。最后，波动率是非对称的，在牛市和熊市之中，标的资产价格的上涨和下跌，波动率对其变化程度的反应是不同的。

通常上，我们将波动率分为四类，每一种都有独自的特点，定义和计算方法。

(1) 历史波动率

历史波动率，是统计市场上一定期间内标的资产价格变化的数据，计算标的资产收益率的变化，从而计算出来的波动率。历史波动率很好的反应了统计期间内，标的资产价格变化的程度。不仅如此，历史波动率也是其他波动率的基础，是预测其他波动率的基石。因此，历史波动率的计算方法十分重要，一般来说，首先，要锁定研究对象即标的资产。其次，人为的规定一个频率或者一个期间，一般以每日，每周或者每月为一个期间，获取期间两端的价格数据，将期末的价格与起初的价格想比，并且取其自然对数，最后计算这些数值的方差，并将其年化。这样就能得到年化后的历史波动率。

(2) 隐含波动率

隐含波动率是基于经典布莱克舒尔斯模型所衍生出来的一个概念。隐含波动率模型是使用 B-S-M 模型中，6 个变量分别为期权价格，行权日时标的资产价格，行权价，无风险收益率，到期时间以及波动率。其中，除了波动率以外的所有数据均可以在现实市场中获取，从而倒推出波动率，这种算法所得到的波动率叫做隐含波动率。隐含波动率的实质是利用现实可得的数据去对未来不可知的结果进行的一次预测。自从期权定价公式问世以来，很多学者都着重研究隐含波动率，认为隐含波动率在研究期权中占有非常重要的地位。如果在一个成熟的市场中，价格能真实地反应市场的供需状态，因此隐含波动率本质是市场中双方博弈的结果。

(3) 未来波动率

未来波动率是金融资产在未来一段时间内的波动率。与历史波动率有所不同，历史波动率是基于金融资产的历史交易数据通过上述的步骤和公式能计算得到，未来波动率则是要靠各种估计方法预测得到。如何运用科学有效的估计方法精准的预测未来波动率是研究波动率过程中至关重要的一步。

(4) 预期波动率

预期波动率是指使用模型对波动率预测的结果。期权理论定价中实际使用的波动率是预期波动率。这意味着在讨论期权时使用的波动率通常意味着预测波动率。历史波动率是预期波动率的基础，预期波动率是人们对实际波动率的理解和认识。

2.2.2 GRACH 族类模型

(1) GRACH 模型

$u_{n-1}^2, u_{n-2}^2, \dots, u_{n-m}^2$ 中的每一项都等权重。研究的意义是估计出当前波动率 σ_n 的水平，因此应该给予预测日较近的数据更高的权重，公式更符合现实意义，一种这样的模型为：

$$\sigma_n^2 = \sum_{i=1}^m \alpha_i u_{n-i}^2$$

变量 α_i 为第 i 天前观察值所对应的权重， α 恒为正值。当选择这些变量时如果对 $i > j$ 选择 $\alpha_i < \alpha_j$ ，也就是对于较旧的数据我们将设定较小的权重。权重之和必须为 1，即

$$\sum_{i=1}^m \alpha_i = 1$$

对于上述公式可以推广。首先，将权重赋予长期平均方差，上述公式则可变形，如下所示：

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^m \alpha_i u_{n-i}^2$$

其中， V_L 为长期方差率， γ 为 V_L 的权重，则

$$\gamma + \sum_{i=1}^m \alpha_i = 1$$

上述模型就是有 Engle 由 1982 年首次提出的 ARCH (m) 模型。由多个方差和一个长期平均方差构成，赋予越新的数据越高的权重。令 $\omega = \gamma V_L$ ，我们可以将公式写为

$$\sigma_n^2 = \omega + \sum_{i=1}^m \alpha_i u_{n-i}^2$$

之后几年的发展，在 1986 年，Bollerslev 提出了 GARCH (1, 1) 模型。GARCH (1, 1) 表达式为：

$$\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

其中 γ 为对应于 V_L 的权重， α 为对应于 u_{n-1}^2 的权重， β 为对应于 σ_{n-1}^2 的权重。因为权重之和为 1，有 $\alpha + \beta + \gamma = 1$ 。

GARCH (1, 1) 模型中的 (1, 1) 的含义是方差 σ^2 是由上一期的 u^2 观察值和上一期的方差率估计出来的。GARCH (p, q) 模型中， σ^2 是由 p 个 u^2 观察值和 q 个方差率估计出来的。而 GARCH (1, 1) 学术界大部分学者都十分青睐的 GARCH 模型的一种。令 $\omega = \gamma V_L$ ，我们也可以将 GARCH (1, 1) 模型写成

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

(2) EGARCH 模型

1991 年, Nelson 在 GARCH 模型的基础上改进了模型 EGARCH 模型。EGARCH 模型中出现正的和负的资产收益率的非对称性, 表明收益率的涨跌对金融资产的影响是不同的。EGARCH 模型从更加优化了 GARCH 方面的不足, 从一个新的角度描述了价格收益率的条件方差。模型的形式如下:

$$\ln h_t^2 = \vartheta_0 + \sum_{j=1}^q \vartheta_j * \left| \frac{\varepsilon_{t-j}}{\sqrt{h_{t-j}}} \right| + \sum_{i=1}^q \lambda_i * \left(\frac{\varepsilon_{t-j}}{\sqrt{h_{t-j}}} \right) + \sum_{i=1}^p \beta \ln h_{t-i}^2$$

该模型具有以下特点:

a. 它的条件方差是线性的, 不需要考虑 $\ln h_t^2$ 的值, 由于 h_t 不是负值, 因此能够允许回归系数出现负值。 h_t^2 对不同的正负效应的反应不同, 而且对于相同程度的负效应, 反应更加强烈。

b. ε_{t-j} 能够更好的解释冲击。

c. 刻画条件方差的不对称性特点。

综上所述 EGARCH 模型是应用的条件更多, 且可以对条件方差的不对称性进行刻画的模型。因此模型在参数使用时的灵活性要更高。即使使用了不对称分布函数, 模型依然合理。

(3) TGARCH 模型

传统的 GARCH 模型并没有考虑杠杆效应, Glosten 在 1993 对基础的 GARCH 模型进行了优化, 提出了 T-GARCH 模型。该模型考虑到了金融数据的非对称性。模型的形式如下:

$$\sigma_t^2 = w + \sum_{i=1}^p \alpha_i \sigma_{t-i}^2 z_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \delta I_{t-1} \sigma_{t-1}^2 Z_{t-1}^2$$

其中, I_{t-1} 为示性函数, 当 $z_t < 0$ 时, I_{t-1} 则取 1, 当 $z_t > 0$ 时, I_{t-1} 则取 0, 为非对称效应项。从上述公式中可看出只要 δ 不为 0, 说明则存在杠杆效应。杠杆效应意味着, 正负向冲击造成的市场的涨跌对方差的影响是不同的, 市场上涨对方差的影响是 α , 市场下跌对条件方差的影响是 $\alpha + \delta$ 。若 δ 取负值, 则杠杆效应削弱了资产价格的波动; 若 δ 取正值 则说明杠杆效应增强了资产价格的波动。T-GARCH 模型可以很好的反映金融资产收益率的有偏分布, 且描述出杠杆效应。

2.2.3 隐含波动率模型

隐含波动率是基于经典布莱克舒尔斯模型所衍生出来的一个概念。隐含波动率模型是使用 B-S-M 模型中,6 个变量分别为期权价格,行权日时标的资产价格,行权价,无风险收益率,到期时间以及波动率。其中,除了波动率以外的所有数据均可以在现实市场中获取,从而倒推出波动率,这种算法所得到的波动率叫做隐含波动率。隐含波动率的实质是利用现实可得的数据去对未来不可知的结果进行的一次预测。自从期权定价公式问世以来,很多学者都着重研究隐含波动率,认为隐含波动率在研究期权中占有非常重要的地位。公式如下所示:

$$c = SN(d_1) - Xe^{-r(T-t)}N(d_2)$$

$$d_1 = \frac{\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = d_1 - \sigma\sqrt{T-t}$$

已知欧式期权的市场价格 C、标的物价格 S、行权价 X、到期日 T、无风险利率 r,可以反推 T-t 时间内的沪深 300ETF 的隐含波动率 σ 。

2.2.4 高频数据模型

由于受以前计算机搜集数据和处理数据能力有限,学者们早期研究金融资产的波动率时,往往选取金融资产的日数据或者频率更低的数据,如周数据,月数据,并建立相应的波动率估计模型进行估计。

高频数据模型基于市场异质假说。短期交易者对资产短期的波动十分的敏感,他们对市场中不断出现的新信息加以处理和加工,又会出现对未来金融资产的新的预期,并根据新的预期不断调整自己的交易。而对于中长期交易者来说,比如基金经理,他们是周交易者或者月交易者。他们对于市场上新出现的信息处理较慢,只关心在中长期资产价格的表现。不会很迅速地调整自己的预期,因此中长期的交易者们的交易频率会较低。

根据市场异质假说,将市场中的交易者按照频率分为了三种不同类型,在 2004 年,Corisi 将三种不同频率的波动率加总在一起,发现通过普通最小二乘法可以很好的拟合出来标的资产的波动率,且能很好地反映出来金融资产的长记忆

性这一特点。近来，诸多学者都将研究重点放在了高频数据上。因此，越来越多的学者喜欢使用 HAR-RV 模型用来预测金融资产的波动性。HAR-RV 模型是使用日内高频数据（一般使用五分钟高频数据）的一阶自回归模型。公式表示如下：

$$RV_{t+H}^d = \beta_0 + \beta_D RV_t^d + \beta_w RV_t^w + \beta_m RV_t^m + \varepsilon_{t+H}$$

$$RV_t = \sum_{i=1}^M r_{t,i}^2$$

$$RV_t^w = \frac{1}{5} \sum_{i=1}^5 RV_{t-i}^d$$

$$RV_t^m = \frac{1}{22} \sum_{i=1}^{22} RV_{t-i}^d$$

其中， RV_t 是日已实现波动率， RV_t^w 是周已实现波动率， RV_t^m 月已实现波动率。HAR-RV 模型将市场上的交易者依据频率将其分为了三种不同类型（日，周，月）对应了不同的交易者类型，之后运用 OLS 进行参数估计。

2.2 期权交易策略相关理论

一般而言，期权交易策略可以分为单式交易策略和复式交易策略。单式交易策略是指持有或者做空一份期权。复式交易策略是指投资者可以做多或者做空多份的不同期权。而复式交易策略又可以分为同类型和不同类型。同类型是指交易者持有两份以上的相同期权。比如，交易者同时持有不同行权价格的看涨期权的多头，或者持有有一个看跌期权的多头再持有有一个看跌期权的空头。而不同类型是指交易者可以持有两种及以上的不同类型的期权。比如持有有一个看涨期权的多头，在持有两个看跌期权的空头。但是，不论是哪一种类型的复式交易策略，都离不开最原始的四种期权交易策略，看涨多头，看涨空头，看跌多头和看跌空头。根据积木理论，所有的复杂期权交易策略都是由这四种基本元素不同的排列组合而得到的，复杂策略的损益图也是由这四种类型经过不断地叠加而得到的。正是由于期权交易策略具有丰富的灵活性，所以交易者可以根据自己的需求叠加出令自己满意的策略，使得策略的损益图切合自身的需求。除非特别说明，否则以下策略中认购期权和认沽期权的标的资产都是相同的。期权有各种各样的套利策略，但是由于金融资产的价格是随机的，因此确定资产价格走势的方向以及创建在稳定价格方向的获利能力的期权策略更加困难。期权策略应仅基于波动率的大小，但收益将小于定向策略，但是显著的降低了风险。下面介绍几种常见的波动率期

权策略。

2.2.1 牛市价差组合

牛市策略由两份同一期限的，不同行权价格的认购期权多头所组成的。或者由两份同一期限的，不同行权价的看跌期权组成。牛市策略是预期未来标的资产价格是上涨的，所以也被称为牛市看涨价差组合。该策略的特点是风险有上限，亏损也是有下限的。若标的资产价格设为 S ，行权价为 X 且 $X_1 < X_2$ ，看涨期权价格 C 。当标的资产价格小于执行价格时（ $0 < S < X_1$ ）。损益情况统计如下表所示。

表 2.1 牛市价差组合盈亏表

| 标的物价格 | 策略盈亏 |
|-----------------|----------------------------------|
| $0 < S < X_1$ | 亏损为 $C_1 - C_2$ |
| $X_1 < S < X_2$ | 从亏损逐渐盈利（ $S - X_1 - C_1 + C_2$ ） |
| $S > X_2$ | 最大收益（ $X_2 - X_1 - C_1 + C_2$ ） |

牛市看涨期权你作为一种价差组合，需要初始的投资来在此阶段初期进行构建策略。与单式策略相比，牛市看涨期权降低了成本且扩大了盈利可能性，并在中长期中有较好的表现。此策略优点在于第一，预期未来标的资产价格会有小幅度的上涨，此时进行牛市价差组合比直接做多的策略相比，成本会低，自然地回报也会低。第二，投资者卖出认沽期权，预期未来市场价格上涨赚取收益，同时又买入了一份行权价格低于于之前认购期权的认沽期权的多头，进行风险的管理。第三，认为期权价格不合理，进行套利。

2.2.2 熊市价差组合

熊市价差组合和牛市价差组合相类似，也可以由不同行权价的认购期权或者认沽期权所构成。只不过熊市策略的预期方向与牛市策略恰好相反。熊市价差组合的应用场景是熊市或者中性市场。它与牛市看涨期权类似，风险和收益都是有

限的。若 S 为标的资产价格, X 为行权价, 并且 $X_1 < X_2$, C 为看涨期权的期权费。熊市价差组合的损益情况如下表所列:

表 2.2 熊市价差组合盈亏表

| 标的物价格 | 策略盈亏 |
|-----------------|-----------------------------------|
| $0 < S < X_1$ | 亏损净权利金($P_2 - P_1$) |
| $X_1 < S < X_2$ | 从盈利逐渐亏损 ($X_1 - S - P_1 + P_2$) |
| $S > X_2$ | 最大亏损 $X_1 - X_2 - P_1 + P_2$ |

此策略优点在于第一, 预期未来标的资产价格会有小幅下跌, 此时进行熊市价差组合比直接做空的策略相比, 成本会低, 自然地回报也会低。第二, 投资者卖出认购期权, 预期未来市场价格下跌赚取收益, 同时又买入了一份行权价格高于之前认购期权的认购期权的多头, 进行风险的管理。第三, 认为期权价格不合理, 进行套利。

2.2.3 跨式组合策略

跨式组合策略是进行相同执行价格的认购认沽的投资策略组合。在现实股票市场中, 市场的行情可以分为大致的两类: 趋势行情和盘整行情。趋势行情就是意味着大盘有着明显的上升或者下降趋势, 基于此种行情下, 投资者可以选择牛市价差组合或者熊市价差组合进行获利。但是, 在现实中, 趋势行情并不多见, 更多的是盘整行情, 即大盘在一个范围内进行波动。在这种行情下, 盘整行情持续的时间和大盘未来走势方向都不明确, 基于此, 投资者往往会选用跨式期权组合进行投资套利。正因为如此, 跨式组合又称为规避方向性策略。

根据策略方向的不同, 跨式组合分为买进跨式和卖出跨式。买进跨式组合是指即同时持有相同行权价格, 相同期限的认购期权和认沽期权的多头。设 C 为认购期权的期权费, P 为认沽期权的期权费。投资者在起初进行买进跨式组合的构建时, 需要支付 $C+P$ 的初始成本, 这个成本也是投资者未来的亏损的下限。在当期日时, 标的资产的价格 S 位于 $K-(C+P)$ 和 $K+(C+P)$ 之间时, 策略会亏损, 但是亏损的幅度是小于初始的构建成本。若 $S > K+(C+P)$, 即标的资产进行了较大的

波动，且波动方向为上涨，则投资者可以在此情形下获利。同理，若 $S < K + (C + P)$ ，即标的资产进行了较大的波动，且波动方向为下跌方向，则投资者可以在此情形下获利。买进跨式组的损益见下表：

表 2.3 买进跨式组合损益表

| 标的资产价格 | 状态 | 损益 |
|-------------------------------|----|---------------------|
| $s_T \leq k - (c + p)$ | 盈利 | $k - (c + p) - s_T$ |
| $k - (c + p) \leq s_T \leq k$ | 亏损 | $s_T - k + (c + p)$ |
| $k \leq s_T \leq k + (c + p)$ | 亏损 | $K + (c + p) - s_T$ |
| $s_T \geq k + (c + p)$ | 盈利 | $s_T - k - (c + p)$ |

2.2.4 蝶式价差组合策略

蝶式策略是指由四份相同期限，不同行权价格的同种期权头寸组成而成的策略。设 x 为期权的行权价格，且 $x_1 < x_2 < x_3$ 。具体的组成种类有以下几种：

(1) 由四份认购期权组合而成的正向组合。该蝶式策略是由做空两份认购期权空头（行权价 x_2 ），做多两份不同行权价的认购期权，且行权价（行权价 $x_1 < x_2 < x_3$ ）。盈亏图如图 2.1 所示。

(2) 由四份认购期权组合而成的反向组合。该蝶式策略是由做多两份认购期权（行权价 x_2 ），做空两份不同行权价的认购期权，且行权价（行权价 $x_1 < x_2 < x_3$ ）。盈亏图如图 2.2 所示。

(3) 由四份认沽期权组合而成的正向组合。该蝶式策略是由做空两份认沽期权（行权价 x_2 ），做多两份不同行权价的认沽期权，且行权价（行权价 $x_1 < x_2 < x_3$ ）。盈亏图如图 2.1 所示。

(4) 由四份认沽期权组合而成的反向组合。该蝶式策略是由做多两份认沽期权（行权价 x_2 ），做空两份不同行权价的认沽期权，且行权价（行权价 $x_1 < x_2 < x_3$ ）。盈亏图如图 2.2 所示。

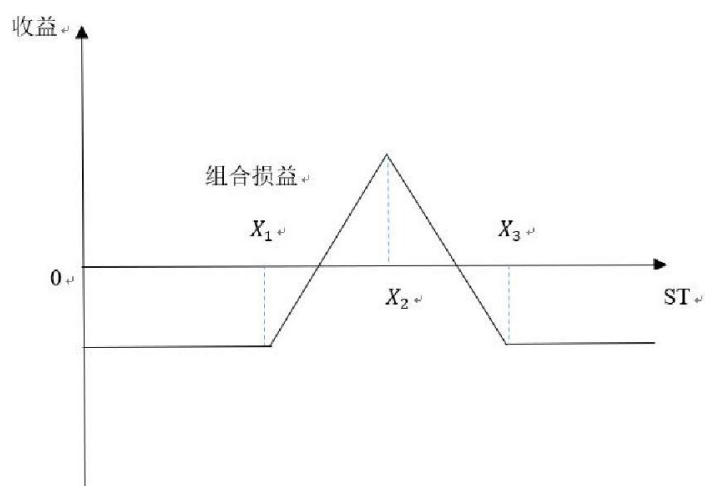


图 2.1 买入蝶式价差组合盈利图

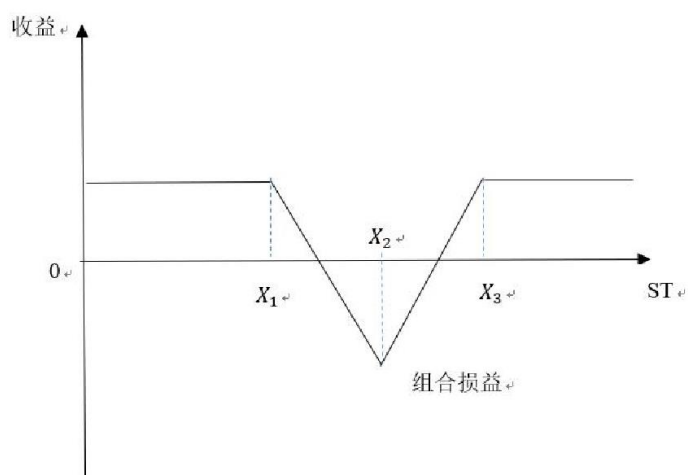


图 2.2 卖出蝶式价差组合盈利图

以买进蝶式价差组合为例，当预期标的金融资产的价格在行权价附近小范围的震荡，投资者可以使用该组合，运用较低的成本来锁定稳定的收益。因此，买进蝶式价差组合，虽然防止了由于标的资产的大幅波动而造成的巨大损失，但是在锁定了一个相对低的盈利机会。总而言之，如果临近行权日，标的资产的价格和执行价格偏离不大，则该策略组合可以获得正的收益。在金融市场中，我们一般假定投资者都是风险厌恶者，多一单位的风险，便要求大于一单位的风险补偿。换言之，只要风险低，收益率低一些也是可以接受的。综上所述，该期权套利策略适合市场中风险厌恶的投资者。

从上述的策略阐述中，我们可以看出，若未来市场是牛市或者熊市，又或者说未来的市场有明显的方向性趋势，投资者可以应用牛市策略或者熊市策略进行套利。当未来市场中性或者有一点偏空时，可以选择宽跨式策略或者蝶式策略，蝶式策略和宽跨式策略有着差不多的最高收益，但蝶式策略优异在于有亏损的下限。综上所述，当波动率变化不大时，蝶式策略比较优异，当波动率的变化方向明显时，可以使用牛市或者熊市策略。

2.3 蒙特卡洛模拟理论

Monte Carlo 是一种基于概率和统计的随机模拟技术，该技术将思想从数值计算转移到统计和概率，实际上，在许多实验案例中，Monte Carlo 都允许使用频率使它们更接近概率估计。Monte Carlo 的思想是将多次试验的频率作为某事件的概率。Monte Carlo 使用数学方法对对象（即实验的数字类似物）的几何和几何特征建模。

在风险中性世界，Monte Carlo 假定随机变量服从以下过程：

$$dS = \tilde{\mu}Sdt + \sigma Sdz$$

模拟价格 S 的路径，将期间分割为 N 个长度为 Δt 的小区间，并得到下式：

$$S(t + \Delta t) - S(t) = \tilde{\mu}S(t)\Delta t + \sigma S(t)\varepsilon\sqrt{\Delta t}$$

Monte Carlo 模拟基于大数定律。伪随机数的产生和大规模的计算量是使用 Monte Carlo 所必需的。这就是为什么蒙特卡洛方法在 1940 前就已经出现，但是一直无法被应用于学术研究的原因。Monte Carlo 广泛用于金融工程，生物医学等领域。

3. 沪深 300ETF 期权套利策略的构建

3.1 沪深 300ETF 期权现状

股票期权是金融市场中重要的风险管理工具，在稳定市场和提高效率方面发挥着重要作用。目前为止，上交所由两只场内的股票期权，分别是上证 50ETF 期权和沪深 300ETF 期权。截至 2020 年末，上交所 ETF 期权合约累计成交 9.82 亿张，上证 50ETF 期权合约全年累计成交 5.19 亿张，沪深 300ETF 期权合约累计成交 4.64 亿张，虽然沪深 300ETF 上市时间只有一年，但是其交易量和上证 50ETF 期权一样活跃。尤其在疫情期间，中国防疫措施到位，中国金融市场平稳运行，规模稳步增长，沪深 300ETF 更是受到了全球投资者的青睐，将沪深 300ETF 期权作为风险管理的良好工具。如下图所示：

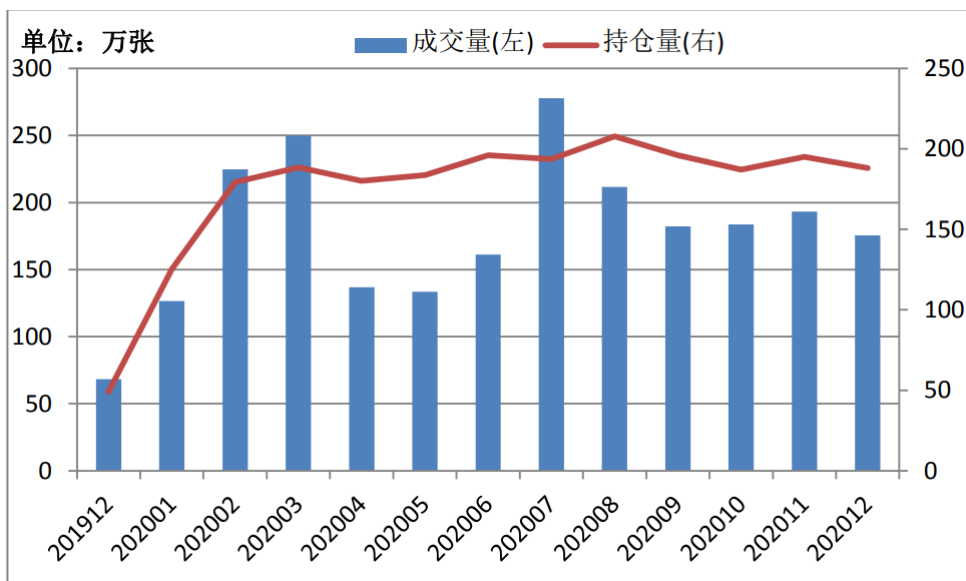


图 3.1 沪深 300ETF 期权日均成交量及日均持仓量情况

截至 2020 年末，期权投资者的账户总数为 48.97 万，2020 年内新增 7.64 万户。从期权交易量上来看，认购期权交易量中，机构投资者和个人投资者的交易量分别占比 54.01%和 45.99%。认沽期权交易量中，机构投资者和个人投资者的交易量分别占比 58.01%和 41.99%。在期权市场上，我国个人投资者的活跃程度和机构投资者分庭抗礼。从交易目的看，保险、增强收益、套利和方向性交易

四类交易行为占比分别为 9.95%、51.17%、18.54%、20.34%。其中，期权主要被应用的目的是增强收入。从投资者类别来分类，机构投资者主要以增强收益和套利交易为主要目的。个人投资者则主要以增强收益和方向性交易为主要目的。

3.2 套利策略的思路与方案设计

3.2.1 套利策略的思路

期权交易的核心是获取收益，获取收益的前提对未来的判断是否准确。从理论层面上来说，直接做认购期权多头和认沽期权多头，依靠着期权的杠杆性的特征能赚取高额的收益，但是前提是投资者对未来要有准确的判断。实际上，现实中很少有人能对未来进行准确的判断，哪怕专业的机构投资者也无法保证未来的走势。大多数投资者若在行情没有明显的趋势的情况下，比如牛市和熊市，投资者是无法判断当前的行情，甚至连资产价格的走势方向也判断不出来。因此我们需要一个直观的，量化的指标作为广大投资者对未来趋势判断的参考。波动率就是这么一个符合投资者预期的参考。投资者不需要关心未来价格走势和方向。只需要将估计的波动率与构建策略当日的标的 5 天的历史波动率作比较，选择一个合适的策略进行套利。这样的策略既可靠又大大降低了投资者的交易难度。本文在之前理论部分讲述了许多关于波动率的理论的介绍，波动率的预测难度要比预测标的资产价格容易得多。所以我们会首先预测波动率，然后由预测到的波动率进行蒙特卡洛模拟后得到行权日当天标的资产的价格，从而看出策略的效果。在上文中，同样介绍了我国投资者常用的投资策略，并对每个策略的应用场景和盈亏效果都做了详尽的陈述，本文会根据所估计出来的波动率与历史波动率相比较，选出最合适的策略并进行回测实验。

3.2.2 套利方案设计

基于波动率预测的期权套利模型的方案流程可以分为以下几个步骤：

第一步：数据的收集和预处理。本文需要收集 2018 年 1 月 9 日到 2021 年 1 月 1 日沪深 300ETF 的收盘价数据；2018 年 1 月 9 日至 2021 年 1 月 1 日沪深 300ETF 的 5 分钟高频数据；2020 年 1 月至 2020 年 12 月的沪深 300ETF 期权的数

据。数据来自 Wind 金融数据库。

第二步：建立波动率模型。本文会依次建立三种不同的波动率预测模型，分别为 GARCH 模型，隐含波动率模型以及高频数据模型。这章节会使用到 matlab 2018a 和 excel 等软件。

第三步：计算未来波动率。因为沪深 300ETF 期权的行权日为每个月的第四个星期三，预测日在行权日的前五个工作日，即会在每个月的第三个星期三收盘后建模，并且预测这五个交易日的波动率，并将得到的结果进行年化。根据不同模型预测的波动率和标的 20 日的历史波动率进行比较，选择合适的期权交易策略。

第四步：预测行权日的标的资产价格。本文通过 Monte Carlo 实现对标的资产价格的预测。

期权蝶式套利的具体流程图如下：

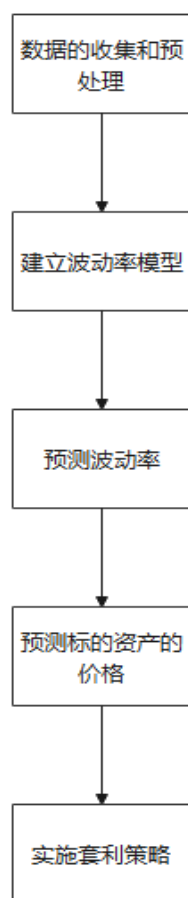


图 3.2 蝶式套利流程图

3.3 数据的收集和预处理

本文使用 Matlab 2018a 对进行沪深 300ETF 的相关处理进行处理。本文所需要的数据主要有沪深 300ETF 的每日收盘价和沪深 300ETF 期权的执行价，到期日，期权价等。这些数据都来自于 Wind 数据库。本文需要 2018 年 1 月 9 日至 2021 年 1 月 1 日沪深 300ETF 的收盘价。用这些数据去计算每日的对数收益率，用以建立 GARCH 族模型和 HAR-RV 模型；2020 年 1 月至 12 月的沪深 300ETF 期权的相关数据，计算 B-S-M 模型下的看涨期权的隐含波动率；收集 2018 年 1 月 1 日至 2021 年 1 月 1 日沪深 300ETF 价格的 5 分钟高频数据，计算沪深 300ETF 已实现波动率。不同的区间长度会影响预测的结果，本文根据文献综述中先前学者们的历史经验，选择 500 天为区间长度。

3.3.1 正态性检验

预测 2020 年 12 个月的沪深 300ETF 的波动率，需要样本数据从 2018 年 1 月 9 日到 2020 年 12 月 23 日沪深 300ETF 的收盘数据。在每个预测日，用预测日前 499 个交易日的数据为样本，计算每日对数收益率，并进行建模预测波动率，将得到的结果在进行年化处理。一年总共需要预测 12 次，在每个月的建模日进行预测。因为沪深 300ETF 的行权日定在每个月的第四个星期三，所以预测日应该在每个月的第三个星期三，并且在收盘之后进行建模预测。

整个预测区间是 2020 年 1 月 22 日到 2020 年 12 月 23 日，则本文的第一个预测日为 2020 年 1 月 22 日，选择过去 499 个交易日的历史数据，即 2018 年 1 月 9 日到 2020 年 1 月 22 日沪深 300ETF 收盘价数据对波动率进行建模和预测。

针对这 500 个数据，首先计算出这 500 个数据的 499 个对数收益率 $R_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$ ，并将其结果整理在 matlab 中，其描述性统计见下图：

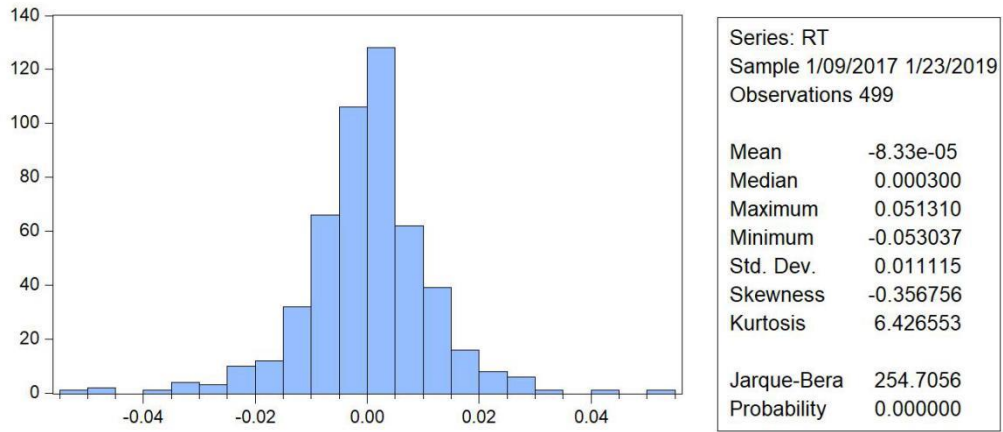


图 3.3 描述统计图

从上图可知：

(1) 沪深 300ETF 收盘价日收益率的峰度为 6.426553，其值远大于标准正态分布下的 3。证明沪深 300ETF 收盘价日收益率是尖峰肥尾。

(2) 沪深 300ETF 收盘价日收益率的偏度 $skewness = -0.356756$ ，形状有左拖尾，其不服从正态分布。

3.3.2 平稳性检验

在做时间序列模型时，首先要对金融时序数据做平稳性检验，则本文要对收益率样本数据进行 ADF 检验。首先做带有截距项和时间趋势项的 ADF 检验，得到下表：

表 3.1 ADF 检验结果表

| ADF (P 值) | 1%level | 5%level | 10%level |
|--------------|---------|---------|----------|
| -23.5264 (0) | -3.9765 | -3.4188 | -3.1319 |

从表可看出，ADF 值为 -23.52640 远小于 1% 的临界值，且 p 值为 0。时间趋势项的系数显著，则拒绝原假设。这说明数据通过了平稳性检验。

3.3.3 自相关性检验

通过了平稳性检验后，需要对通过了平稳性检验的时序数据做自相关性检验，通过 ACF 和 PACF 来初步识别 p 与 q ，用以确定 ARMA 模型的 p, q 的数值。将平稳的数据通过 EViews 进行 ACF 和 PACF 检验，结果如下图所示。ACF 和 PACF 的显示拖尾震荡。通过赤池准则（AIC）确定 ARMA 的阶数问题，并进行建模。

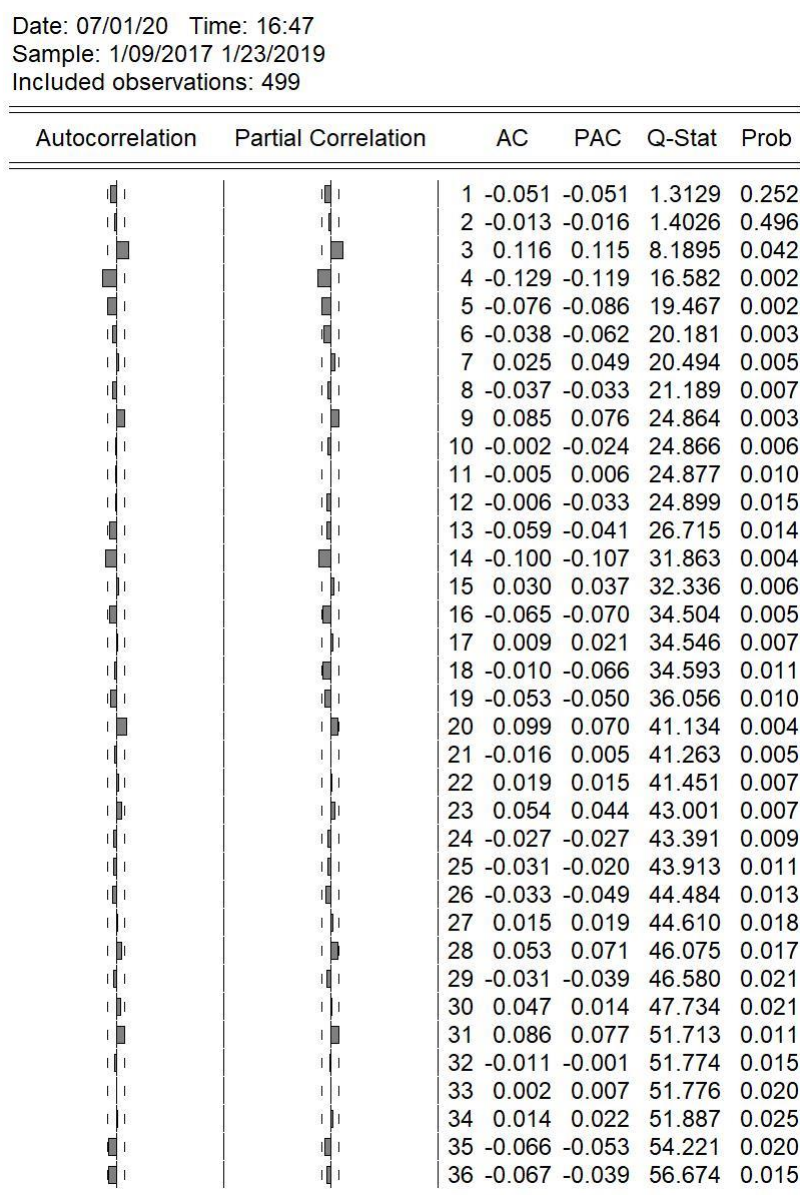


图3.4 ACF和PACF检验图

为了确定ARMA模型中 p, q 的阶数，使用软件构建不同 p, q 取值下的ARMA， p, q 的取值在为0, 1, 2, 3，通过对比不同 p, q 下，ARMA模型的AIC的大小，根据赤池

准则，选择AIC最小的模型，通过比较可得知，最后选取了ARMA（3,1）模型。

表3.2 ARMA (p, q) 模型的AIC值

| | AIC |
|--------------------|-------------------|
| ARMA (0, 1) | -2880.5839 |
| ARMA (0, 2) | -2878.8920 |
| ARMA (0, 3) | -2883.2582 |
| ARMA (1, 0) | -2880.6005 |
| ARMA (1, 1) | -2878.7951 |
| ARMA (1, 2) | -2894.1120 |
| ARMA (1, 3) | -2894.0882 |
| ARMA (2, 0) | -2878.7893 |
| ARMA (2, 1) | -2876.7893 |
| ARMA (2, 2) | -2874.7893 |
| ARMA (2, 3) | -2872.7893 |
| ARMA (3, 0) | -2884.4251 |
| ARMA (3, 1) | -2894.7678 |
| ARMA (3, 2) | -2892.7823 |
| ARMA (3, 3) | -2893.7392 |

3.3.4 ARCH 效应检验

在经过ARMA模型建立后，需要对时间序列数据残差进行建模，但是建模之前需要对残差序列进行ARCH效应检验。确定残差序列是否存在异方差。本文选择使用ARCH-LM进行ARCH效应检验。表3.3是ARCH-LM检验结果表，结果表明有异方差性，接下来就可以使用GARCH模型进行建模。

表3.3 ARCH-LM检验结果表

| | |
|-----------------|----------|
| Obs*R-squared | 6.987636 |
| Pro. Chi-Square | 0.0082 |

3.3.5 构建 GARCH 族模型计算波动率

在建立GARCH族模型时，本文会使用分别在正态分布下和t分布下的GARCH模

型, EGARCH模型和T-GARCH模型, 一共六种模型。根据AIC最小的原则, 计算出不同模型下的AIC值, 选取AIC最小的模型建立波动率模型, 并且预测波动率。在2020年1月进行第一次预测时, 选用这500个数据进行构建GARCH族类模型, 各个模型的AIC值统计如下表所示:

表3.4 GARCH 族模型的AIC值

| | GARCH-N | GARCH-T | EGARCH-N | EGARCH-T | T-GARCH-N | T-GARCH-T |
|-----|---------|----------------|----------|----------|-----------|-----------|
| AIC | -2933.7 | -2966.8 | -2929.9 | -2965.6 | -2932.2 | -2965.4 |

AIC的最小值在GARCH-T模型处取得。因此, 第一个月的波动率模型选择GARCH-T模型进行预测, 经过计算得到的均值和方差方程为:

$$r_t = 1.164e - 04 - 0.802r_{t-1} + a_t + 0.799a_{t-1} - 0.029a_{t-2} + 0.0677a_{t-3}$$

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = 1.9767e - 06 + 0.0403\sigma_{t-1}^2 + 0.9492a_{t-1}^2$$

在2020年1月至2020年12月之间, 将需要进行12次预测。进行动态预测, 即在每个月的建模日, 即在每个月的第三个星期三建立波动率模型。都进行一次上述的过程, 根据AIC准则, 选取最优的模型, 并得到均值方差模型, 最后预测未来的波动率。因此, 在每个月的建模日当天所选用的ARMA和GARCH模型会有所不同, 均是按照上述的做法根据AIC准则选择出当月的最优模型, 每个模型都会预测未来五日的波动率, 并将其转化为年化收益率。总共需要进行12次的建模和预测。

在2020年的12次建模预测中, 统计出T-GARCH总共使用了8次, GARCH-T模型4次, 合计12次。结果重复证明了, T-GARCH模型的拟合效果要优于其他模型, 在一月份中, 虽然使用了GARCH模型, 但是从表3.4中, 也可以发现GARCH-t模型和T-GARCH-t的AIC值十分相近。这一结果, 表明了500日的沪深300ETF的日收益率有着长记忆性和非对称性。再有, 在12次的预测中, 都是选用了t分布, 这一结果表明沪深300ETF的日收益率分布更接近于学生分布, 而不是正态分布, 这与之前的描述性统计所得出的结论相一致。

统计结果如下表所示:

表 3.5 GARCH 建模统计结果

| 500 天 | AR | MA | GARCH 族模型 | 波动率 |
|-------|----|----|-----------|--------|
| 1 月 | 3 | 1 | GARCH-T | 0.0154 |
| 2 月 | 2 | 1 | GARCH-T | 0.0131 |
| 3 月 | 2 | 1 | TGARCH-T | 0.0510 |
| 4 月 | 1 | 3 | GARCH-T | 0.0144 |
| 5 月 | 1 | 3 | TGARCH-T | 0.0115 |
| 6 月 | 1 | 3 | TGARCH-T | 0.0172 |
| 7 月 | 1 | 3 | TGARCH-T | 0.0328 |
| 8 月 | 2 | 3 | TGARCH-T | 0.0279 |
| 9 月 | 2 | 1 | GARCH-T | 0.0176 |
| 10 月 | 2 | 1 | TGARCH-T | 0.0120 |
| 11 月 | 2 | 2 | TGARCH-T | 0.0196 |
| 12 月 | 3 | 3 | TGARCH-T | 0.0370 |

由于之后我们需要年化后的波动率,但之前只是预测了五个交易日的波动率,因此我们要将其年化,具体公式为:

$$\sigma = \text{sqrt}(\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 + \sigma_5^2) * \text{sqrt}\left(\frac{250}{5}\right)$$

3.3.6 基于 B-S-M 模型计算隐含波动率

沪深 300ETF 期权为欧式期权,通过 Wind 收集到市场上有关期权的数据,根据 B-S-M 波动率公式,求期权的隐含波动率 IV。其中,无风险利率本文使用 shibor 利率。本文在每个月第三个星期三,计算当月的 IV。即,第一个月的期权数据选择 2020 年 1 月 15 日的期权数据。B-S-M 公式如下:

$$c = SN(d_1) - Xe^{-r(T-t)}N(d_2)$$

$$p = Xe^{-r(T-t)}N(-d_2) - SN(-d_1)$$

$$d_1 = \frac{\ln\left(\frac{s}{x}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = d_1 - \sigma\sqrt{T - t}$$

结果如下表所示：

表 3.6 隐含波动率 1-12 月统计结果

| 月份 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|-------|
| IV (%) | 10.93 | 15.72 | 35.83 | 17.84 | 15.11 | 12.84 | 31.59 | 23.17 | 17.63 | 17.07 | 16.4 | 14.91 |

3.3.7 基于高频数据波动率模型计算波动率

在之前进行做 GARCH 模型时，已经计算出了沪深 300ETF 的对数收益率，在此基础上计算已实现波动率，计算公式如下：

$$RV = \sqrt{\frac{1}{n-1} * \sum_{i=1}^n (r_i - \bar{r})^2}$$

$$\bar{r} = \frac{1}{n} * \sum_{i=1}^n r_i$$

在闫会强，夏霄松等前学者的研究中，证实五分钟的高频数据所含信息最多，因此本文也使用沪深 300ETF 的五分钟高频数据模型。策略构建日为每个月的第三个星期三，因此在建模日时，使用当天以及前 499 个交易日的五分钟高频数据为样本。建立 HAR-RV 和 HAR-LNRV 模型，通过比较两个模型的 R^2 选择最优的预测模型。第一个建模日在 2020 年 1 月 15 日，搜集建模日前的日数据，周数据以及月数据进行建模。参数估计如表 3.7 所示。比较 HAR-RV 模型和 HAR-InRV 模型的 R^2 和 $ADJ - R^2$ 调整后的拟合优度，很显然 HAR-RV 模型的 R^2 值为 0.5425， $ADJ - R^2$ 值为 0.5424，都低于 HAR-InRV 模型的 R^2 值为 0.5494， $ADJ - R^2$ 值为 0.5493。两个方程都能通过 F 检验，综合来看，2020 年 1 月沪深 300ETF 的 5 分钟高频数据建立的模型应该是 HAR-LNRV 模型。

得到 2020 年 1 月的 HAR-InRV 模型如下所示：

$$RV_{t+1} = 1.094e - 05 + 8.279e - 01RV_t^d - 2.167e - 01RV_t^W + 8.333e - 02RV_t^M$$

表 3.7 HAR 模型对比的参数估计结果

| HAR-LNRV | Coefficient | Prob | HAR-RV | Coefficient | Prob |
|--------------------|-------------|--------|--------|-------------|----------|
| C | 1.094e-05 | 0.4709 | | -1.088e-05 | 0.37 |
| RV1 | 8.279e-01 | 2e-16 | | 8.201e-01 | 2e-16 |
| RV5 | -2.167e-01 | 2e-16 | | -2.206e-01 | 2e-16 |
| RV22 | 8.333e-02 | 0.0001 | | 1.052e-01 | 4.07e-10 |
| R-squared | 0.5494 | | | 0.5425 | |
| Adjusted R-squared | 0.5493 | | | 0.5424 | |
| F | 7092 | | | 6529 | |
| Prob(F-statistic) | 2.2e-16 | | | 2.2e-16 | |

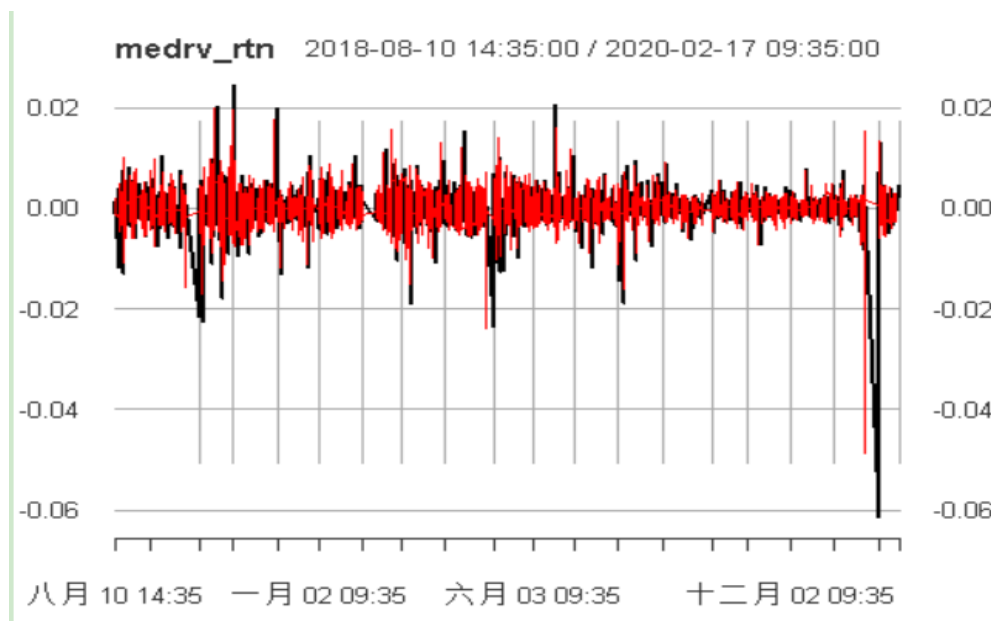


图 3.5 预测波动率与观察波动率图

表 3.8 HAR-LNRV 模型所预测波动率表

| | | | | | | |
|--------|-------|-------|-------|-------|-------|-------|
| 月份 | 1 | 2 | 3 | 4 | 5 | 6 |
| IV (%) | 10.42 | 14.81 | 32.73 | 15.74 | 12.21 | 14.74 |
| 月份 | 7 | 8 | 9 | 10 | 11 | 12 |
| IV (%) | 28.57 | 21.22 | 18.74 | 15.02 | 15.2 | 12.11 |

从 3.5 图可以看出, 预测波动率和观察波动率的走势基本一致, HAR-LNRV 的模型拟合效果较好, 这与 HAR-LNRV 的可决系数达到 0.54 相吻合。使用 HAR-LNRV 模型可以很好的预测未来波动率。2020 年 1 月到 12 月的预测结果如下表:

3.3.8 预测持有期标的资产价格

将上述计算得到的三种不同模型下的预测波动率与建模日当天的前五日以实现波动率汇总在下图, 如图所示, 结果发现相比于前五日历史波动率, 未来波动率不会发生大的变化, 说明未来五天市场中性, 适合使用正向买进蝶式套利策略。

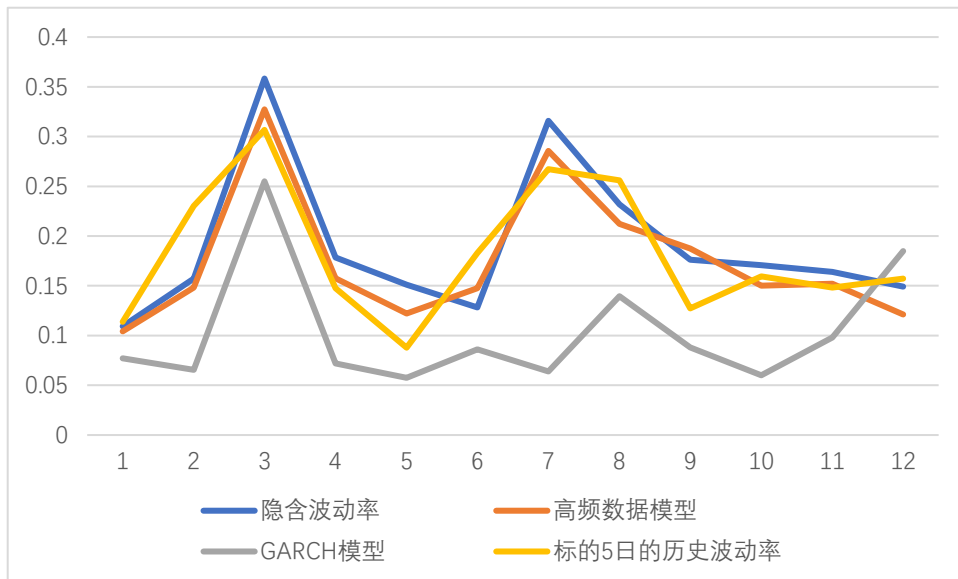


图 3.6 波动率拟合效果图

根据预测得到年化后的波动率, 使用 MATLAB 2018A 做 1000 次 Monte Carlo 产生正态随机数估计区间, 将其结果排序获得置信水平 90%的置信区间 $[s_1, s_3]$ 和标的资产价格的均值 s_2 。

3.4 期权策略选择

根据上图所知, 三种模型所预测出来的波动率基本和沪深 300ETF 的 5 日历史波动率相差不大, 依据前文对期权套利策略的描述中, 若预测波动率和历史的波动率相差不大, 说明未来五日很有可能是市场中性, 在平淡小幅震荡的行情中使用什么策略能获取可观收益。如果选择了卖出期权, 若未来波动率上升会造成

损失。若卖出期权，市场涨幅不大，哪怕方向正确也会导致亏损。所以对于短期的投资者而言，面对短期小幅震荡的行情，可以构建买进蝶式期权套利策略。只要预期行权日时，标的资产价格和行权价格误差不大，便可以获利。哪怕对未来波动率走势预期出现偏差，形成单边行情，损失也是有限的。投资者无需频繁跟随市场走势进行平仓止损，大大降低了投资者的操作难度。故投资者最优选择的策略为正向买进蝶式套利策略。可以构造预测价格区间上的固定收益。在持有期初进行建立空头头寸赚取权利金，在持有期末进行多头平仓，由于期权的时间价值不断减少，平仓的成本理应小于建仓赚取的權利金。如果对将来一段时间资产价格的变动范围预测的足够准确，波动率的预测效果还不错，这一思路是可行的。蝶式差价组合由三种不同行权价的期权组合而成，它与跨式组合的不同之处在于他的损失端是有限的，而两种跨式组合的损失段是无限的，这是蝶式差价的优势之处。

综合分析来看，本文选择构建蝶式差价组合策略来进行沪深 300ETF 期权的波动率套利研究，其盈亏分布如表 3.9 所示。

表 3.9 认购期权正向蝶式组合盈亏分布

| S_T 的范围 | c_1 盈亏 | 2 份 c_2 盈亏 | c_3 盈亏 | 总盈亏 |
|----------------------|-------------------|----------------------|-------------|--------------------------------|
| $S_T \leq X_1$ | $0 - c_1$ | $0 + 2c_2$ | $0 - c_3$ | $2c_2 - c_1 - c_3$ |
| $X_1 < S_T \leq X_2$ | $S_T - X_1 - c_1$ | $0 + 2c_2$ | $0 - c_3$ | $S_T - X_1 + 2c_2 - c_1 - c_3$ |
| $X_2 < S_T \leq X_3$ | $S_T - X_1 - c_1$ | $2X_2 - 2S_T + 2c_2$ | $0 - c_3$ | $X_3 - S_T + 2c_2 - c_1 - c_3$ |
| $S_T > X_3$ | $S_T - X_1 - c_1$ | $2X_2 - 2S_T + 2c_2$ | $S_T - X_3$ | $2c_2 - c_1 - c_3$ |

我们在建模日根据获得的预测波动率对未来五天后，即行权日当天的标的资产——沪深 300ETF 的价格进行预测。策略构建在建模日开仓，做空两份行权(X_2)价最接近的预测到的标度资产价格 S 的认购期权，然后做多一份行权价 (X_1) 的认沽期权和做多一份行权价 (X_3) 的认沽期权，要求 $X_3 - X_2 = X_2 - X_1$ 。

4. 期权蝶式套利的实施方案

将蒙特卡洛模拟的结果排序获得置信水平 90%的置信区间 $[s_1, s_3]$ 和标的资产价格的均值 s_2 , 然后依据此建立期权蝶式组合套利策略, 做多一份行权价 X_1 接近 s_1 且大于 s_1 的认购期权, 做空两份行权价 X_2 接近均值 s_2 的认购期权, 做多一份行权价 X_3 接近 s_3 且小于 s_3 的认购期权, 要求尽可能满足 $X_3 - X_2 = X_2 - X_1$ 。即 X_2 为 X_1 和 X_3 之和的平均值。此时使收益 $\max(S_T, X) - \min(S_T, X) + 2c_2 - c_1 - c_3$ 最大即为最优期权套利组合。若未来标的资产价格在区间 $[s_1, s_3]$ 之外, 则会造成最大亏损, 最大亏损为初期的构造成本。在使用策略开仓和平仓时, 每一个交易日都可以选择两个时点, 即开盘和近收盘, 由此可见, 同一种策略排列组合总共有四种不同的开仓平仓时机组合。

4.1 基于不同波动率模型的期权蝶式组合回测收益

4.1.1 GARCH 族模型的期权套利策略

将 GARCH 模型预测的结果带入现实数据进行回测, 将回测结果统计在下表中:

表 4.1 预测日开盘开仓收益表

| 时间 | 行权价(X2) | 行权日开盘价 | 月收益 | 行权日近收盘价 | 月收益 |
|------------|---------|--------|---------|---------|---------|
| 2020/1/16 | 4.1 | 4.036 | 0.0407 | 4.065 | -0.4074 |
| 2020/2/20 | 4.0 | 4.011 | 0.0153 | 4.017 | -0.1531 |
| 2020/3/19 | 3.3 | 3.645 | -0.0123 | 3.658 | 0.1239 |
| 2020/4/16 | 3.5 | 3.730 | -0.1572 | 3.779 | 0.0571 |
| 2020/5/21 | 3.9 | 3.813 | -0.0249 | 3.795 | -0.2490 |
| 2020/6/18 | 3.1 | 4.084 | -0.0760 | 4.110 | 0.8361 |
| 2020/7/16 | 4.5 | 4.670 | -0.1343 | 4.730 | -0.0760 |
| 2020/8/20 | 4.3 | 4.756 | -0.1127 | 4.701 | -0.1343 |
| 2020/9/17 | 4.6 | 4.642 | -0.0470 | 4.652 | -0.1127 |
| 2020/10/22 | 4.6 | 4.700 | -0.0765 | 4.741 | -0.0470 |
| 2020/11/19 | 4.7 | 4.989 | -0.0676 | 4.913 | 0.1765 |
| 2020/12/17 | 4.8 | 4.974 | -0.0760 | 5.008 | -0.0676 |
| 收益额 | | | -0.6886 | | -0.0535 |
| 收益率 | | | -3.359% | | -0.82% |

表 4.1 表示在建模日开盘开仓的两种结果,表 4.2 表示在建模日收盘开仓的两种结果。结果发现 GARCH 模型的预测效果并不理想,四种情况下基本都是亏损状态下,最低亏损达到-0.82%,最高损失为 5.72%。

表 4.2 预测日近收盘开仓收益表

| 时间 | 行权价 | 行权日开盘价 | 收益 | 行权日收盘价 | 收益 |
|------------|-----|--------|---------|--------|---------|
| 2020/1/16 | 4.1 | 4.036 | 0.0692 | 4.065 | 0.0692 |
| 2020/2/20 | 4.0 | 4.011 | 0.0099 | 4.017 | 0.0099 |
| 2020/3/19 | 3.3 | 3.645 | -0.0054 | 3.658 | -0.0054 |
| 2020/4/16 | 3.5 | 3.730 | -0.0033 | 3.779 | -0.0033 |
| 2020/5/21 | 3.9 | 3.813 | 0.0841 | 3.795 | 0.0841 |
| 2020/6/18 | 3.1 | 4.084 | 0.0045 | 4.110 | 0.0045 |
| 2020/7/16 | 4.5 | 4.670 | 0.1760 | 4.730 | -0.1760 |
| 2020/8/20 | 4.3 | 4.756 | -0.1023 | 4.701 | -0.1023 |
| 2020/9/17 | 4.6 | 4.642 | -0.2586 | 4.652 | -0.2586 |
| 2020/10/22 | 4.6 | 4.700 | 0.0976 | 4.741 | 0.0976 |
| 2020/11/19 | 4.7 | 4.989 | 0.0117 | 4.913 | 0.0117 |
| 2020/12/17 | 4.8 | 4.974 | -0.1262 | 5.008 | -0.1262 |
| 年收益额 | | | -0.0426 | | -0.3948 |
| 年收益率 | | | -5.72% | | -2.612% |

4.1.2 B-S-M 模型下隐含波动率的期权套利策略

将隐含波动率模型预测的结果带入现实数据进行回测,将回测结果统计在下表中,表 4.3 表示在建模日开盘开仓的两种结果,表 4.4 表示在建模日收盘开仓的两种结果。结果发现 GARCH 模型的预测效果较好,且在开盘开仓时蝶式套利策略都有盈利,最高收益率为 12.039%,最大亏损为-1.91%。

表 4.3 预测日开盘开仓收益表

| 时间 | 行权价 (X2) | 行权日开盘价 | 月收益 | 行权日近收盘价 | 月收益 |
|------------|----------|--------|---------|---------|---------|
| 2020/1/16 | 4.1 | 4.036 | 0.0615 | 4.065 | 0.0375 |
| 2020/2/20 | 4.0 | 4.011 | 0.0159 | 4.017 | 0.0150 |
| 2020/3/19 | 3.5 | 3.645 | -0.0599 | 3.658 | 0.0024 |
| 2020/4/16 | 3.6 | 3.730 | -0.0279 | 3.779 | 0.0830 |
| 2020/5/21 | 3.9 | 3.813 | 0.0907 | 3.795 | -0.0560 |
| 2020/6/18 | 4.1 | 4.084 | 0.2044 | 4.110 | 0.0562 |
| 2020/7/16 | 4.5 | 4.670 | 0.0051 | 4.730 | -0.1005 |
| 2020/8/20 | 4.6 | 4.756 | 0.0409 | 4.701 | -0.0018 |
| 2020/9/17 | 4.6 | 4.642 | 0.0369 | 4.652 | 0.0547 |
| 2020/10/22 | 4.6 | 4.700 | 0.0935 | 4.741 | 0.0333 |
| 2020/11/19 | 4.9 | 4.989 | -0.0028 | 4.913 | 0.0030 |
| 2020/12/17 | 4.8 | 4.974 | -0.2044 | 5.008 | -0.0001 |
| 收益额 | | | 0.2541 | | 0.1269 |
| 收益率 | | | 12.039% | | 10.14% |

表 4.4 预测日近收盘开仓收益表

| 时间 | 行权价 (X2) | 行权日开盘价 | 收益 | 行权日收盘价 | 收益 |
|------------|----------|--------|---------|--------|---------|
| 2020/1/16 | 3.5 | 4.036 | 0.0610 | 4.065 | 0.0144 |
| 2020/2/20 | 3.5 | 4.011 | -0.0526 | 4.017 | 0.0081 |
| 2020/3/19 | 3.6 | 3.645 | -0.0063 | 3.658 | -0.0067 |
| 2020/4/16 | 3.9 | 3.730 | -0.0046 | 3.779 | -0.0070 |
| 2020/5/21 | 4.1 | 3.813 | -0.0011 | 3.795 | -0.0032 |
| 2020/6/18 | 4.5 | 4.084 | 0.0120 | 4.110 | 0.0085 |
| 2020/7/16 | 4.6 | 4.670 | -0.0002 | 4.730 | -0.0217 |
| 2020/8/20 | 4.6 | 4.756 | -0.0001 | 4.701 | 0.0213 |
| 2020/9/17 | 4.6 | 4.642 | -0.0601 | 4.652 | 0.0271 |
| 2020/10/22 | 4.9 | 4.700 | 0.0794 | 4.741 | 0.1106 |
| 2020/11/19 | 4.8 | 4.989 | -0.0055 | 4.913 | 0.0195 |
| 2020/12/17 | 4.1 | 4.974 | -0.0407 | 5.008 | -0.2011 |
| 年收益额 | | | -0.0191 | | 0.1037 |
| 年收益率 | | | -1.91% | | 6.284% |

4.1.3 高频数据模型下的期权套利策略

将高频数据模型预测的结果带入现实数据进行回测，将回测结果统计在下表中，表 4.5 表示在建模日开盘开仓的两种结果，表 4.6 表示在建模日收盘开仓的两种结果。结果发现高频数据模型的预测效果较好，且在近收盘平仓时蝶式套利策略都有盈利，最高收益率为 8.683%，最大亏损为-0.272%。

表 4.5 预测日开盘开仓收益表

| 时间 | 行权价 (X2) | 行权日开盘价 | 月收益 | 行权日近收盘价 | 月收益 |
|------------|----------|---------|---------|---------|---------|
| 2020/1/16 | 4.1 | 0.0636 | 0.0615 | 4.065 | 0.0226 |
| 2020/2/20 | 4.0 | 0.0074 | 0.0159 | 4.017 | 0.0105 |
| 2020/3/19 | 3.5 | -0.0001 | -0.0599 | 3.658 | 0.0362 |
| 2020/4/16 | 3.6 | -0.0063 | -0.0279 | 3.779 | 0.0588 |
| 2020/5/21 | 4.1 | 0.0126 | 0.0907 | 3.795 | 0.0019 |
| 2020/6/18 | 3.9 | -0.0154 | 0.2044 | 4.110 | -0.0290 |
| 2020/7/16 | 4.5 | -0.0251 | 0.0051 | 4.730 | 0.0061 |
| 2020/8/20 | 4.5 | 4.7560 | -0.0362 | 4.701 | 0.0271 |
| 2020/9/17 | 4.6 | 4.6420 | 0.0397 | 4.652 | 0.1112 |
| 2020/10/22 | 4.6 | 4.7000 | 0.0315 | 4.741 | -0.0238 |
| 2020/11/19 | 4.8 | 4.989 | -0.0035 | 4.913 | 0.0127 |
| 2020/12/17 | 4.8 | 4.974 | -0.0688 | 5.008 | 0.1813 |
| 收益额 | | | -0.0006 | | 0.1568 |
| 收益率 | | | -0.272% | | 8.683% |

表 4.6 预测日近收盘开仓收益表

| 时间 | 行权价 | 行权日开盘价 | 收益 | 行权日收盘价 | 收益 |
|-----------|-----|--------|---------|--------|---------|
| 2020/1/16 | 4.1 | 4.065 | 0.0333 | 4.065 | 0.0563 |
| 2020/2/20 | 4.0 | 4.017 | 0.0029 | 4.017 | 0.0570 |
| 2020/3/19 | 3.5 | 3.658 | -0.0030 | 3.658 | -0.0090 |
| 2020/4/16 | 3.6 | 3.779 | -0.0059 | 3.779 | 0.1059 |
| 2020/5/21 | 3.8 | 3.795 | 0.0045 | 3.795 | 0.0045 |
| 2020/6/18 | 3.9 | 4.110 | -0.0002 | 4.110 | -0.0002 |
| 2020/7/16 | 4.5 | 4.730 | -0.0032 | 4.730 | -0.0132 |
| 2020/8/20 | 4.5 | 4.701 | -0.0039 | 4.701 | -0.0229 |
| 2020/9/17 | 4.6 | 4.652 | 0.0302 | 4.652 | 0.0540 |

续表 4.6 预测日近收盘开仓收益表

| | | | | | |
|------------|-----|-------|---------|-------|---------|
| 2020/10/22 | 4.6 | 4.741 | -0.0028 | 4.741 | -0.0028 |
| 2020/11/19 | 4.8 | 4.913 | 0.0094 | 4.913 | -0.0020 |
| 2020/12/17 | 4.8 | 5.008 | -0.0588 | 5.008 | -0.0002 |
| 年收益额 | | | 0.0023 | | 0.1227 |
| 年收益率 | | | 1.899% | | 6.661% |

4.2 基于不同波动率模型的回测结果对比分析

将三种模型的最后预测结果统计下表 4.7 中,可以分析得出 GARCH 族模型的正向买进蝶式套利策略的平均回报为-3.13%,隐含波动率模型的正向买进蝶式套利策略的平均回报为-6.64%,HAR 族模型的正向买进蝶式套利策略的平均回报为 3.28%。结果上看,使用隐含波动率模型的预测模型所做的套利策略效果最好。且在隐含波动率下,建模日开盘时建仓,行权日开盘时平仓的收益率最高,达到 12.04%。

表 4.7 不同波动率模型盈亏结果汇总表

| 模型 | 开盘开仓 | 收益额 | 收益率 | 收盘开仓 | 收益额 | 收益率 |
|-----------|---------|---------|---------|------|-----------|--------|
| GARCH 族 | 开盘平仓 | -0.6886 | -3.35% | 开盘平仓 | -0.0426 | -5.72% |
| 隐含波动率模型 | | 0.2541 | 12.04% | | -0.0190 | -1.91% |
| HAR-RV 模型 | | -0.0062 | -0.27% | | -0.0024 | -1.89% |
| GARCH 族 | 收盘平仓 | -0.0535 | -0.82% | 收盘平仓 | -0.3948 | -2.61% |
| 隐含波动率模型 | | 0.1270 | 10.14% | | 0.1037 | 6.28% |
| HAR-RV 模型 | | 0.1068 | 8.68% | | 0.1227 | 6.60% |
| | GARCH 族 | | 隐含波动率模型 | | HAR-RV 模型 | |
| 策略平均回报率 | -3.13% | | 6.64% | | 3.28% | |

5. 结论与建议

5.1 结论

由于沪深 300ETF 期权于 2019 年年底才上市，上市至今只有一年多的时间，所可获取的数据有限，因此只能做一年的回测。回测数据是从 2020 年 1 月到 2020 年 12 月，虽然我国股市在这一年中因疫情的影响经历了上半年的急速下跌，大盘指数一路跌到 2646.81，下半年因疫情防控得当，股市开始了回升，大盘点数从下半年开始稳健回升并在在 3300 点上下进行波动，从回测结果来看，策略最高累计收益率高达 12.04%。说明策略是具有比较可观的可行性，但在长期是否有稳定性，需要进一步的持续跟踪。

本文根据沪深 300ETF 期权上市以来的标的和期权数据，运用波动率模型预测未来波动率，之后与五日的历史波动率相比较，发现其与历史波动率的走势基本吻合，相差不大，因此在此基础上建立了正向买进蝶式套利策略，最大收益为 $\max(S_T, X) - \min(S_T, X) + 2c_2 - c_1 - c_3$ 。回测比较四种开平仓时机下的期权组合收益额和收益率，最后选择回测收益率结果最好的波动率模型和开平仓时机。经过本文研究得出以下结论：

第一，GARCH 模型中，TGARCH 模型的效果要优于 GARCH 模型和 EGARCH 模型。传统的 GARCH 模型并没有考虑杠杆效应，T-GARCH 考虑到了金融数据的非对称性。从模型的结果可以分析得出，我国股市存在着明显的杠杆效应。正向消息对股市上涨的刺激和负向消息对股市下跌的刺激是不一致的，且负向消息的影响程度大于正向消息的影响程度。

第二，在进行短期波动率预测中，所有的波动率预测模型中，隐含波动率模型的预测效果最好。隐含波动率模型优于其他两种模型，说明沪深 300ETF 期权交易较活跃，市场价格的运行平稳，定价合理，规模稳步增长，经济功能日益凸显。在短期预测中，市场越成熟，交易量越活跃的期权，其隐含波动率包含的信息越多。

第三，通过对不同开平仓时机的回测结果对比发现，三种波动率预测模型都是在开盘开仓，近收盘平仓的情况下三种波动率都有不俗的收益率。每种开平仓时机不同，得到的收益额收益率存在差距。证明开平仓时机在一定程度上会影响

期权底部蝶式组合的收益。为了实现收益最大化，应该选择在建仓日开盘开仓，行权日近收盘平仓。就总收益率比较来说，隐含波动率收益率最高 12.04%。

综上所述，在理论上，市场上无交易成本的情况下，使用正向买进蝶式套利组合对沪深 300ETF 期权进行套利的效果良好，可行性高。

5.2 建议

本文的结论使用了真实的交易数据，运用 GARCH 族类模型，隐含波动率模型以及高频数据模型结合蒙特卡洛模拟预测出来的结果进行正向蝶式套利得到的结论对其在我国期权交易市场的发展提出如下建议：

5.2.1 应鼓励期权量化投资产品的创新

应鼓励期权量化投资产品的创新。当今世界是个信息化的时代，是一个互联网和大数据的时代，就在金融领域而言，是一个量化投资的时代。这一切都归功于计算机这些年的飞速发展和信息获取的便利。以前无法获取的数据，在如今大数据的帮助下，各种信息收集方法变得越来越方便快捷，并且以前无法访问的数据也变得可以获取。金融学术研究的基础是数据的可获得性，真实性和时效性。信息化，大数据的时代很好的满足了金融研究人员的需求。

我国期权市场发展较晚，目前仍处于萌芽发展阶段，许多研究正刚刚开始，许多技术和交易策略未得到市场长时间的验证。为了更有效和充分地反映出不同市场因素对投资者行为的影响，金融机构向投资者提供培训及教育平台，更可以向投资者提供有针对性的投资策略。此外，金融公司还应采用引进人才的方式寻求发展，注重对金融科技复合背景人才的培养，更加重视金融和技术领域的专业人才。我国目前正处于转型阶段，金融市场的良好运行和稳定发展是必不可少的因素。通过建立完善的法律制度来弥补与外国资本市场的差异，推动金融市场的高质量发展。政府应在制定法规保障金融市场良好稳定运行的同时，也应放松政策，给金融产品的创新一个合适的温床，减少政治方面的阻力，推动金融体系走向成熟。

5.2.2 量化期权交易策略的市场渗透率有望提高

从市场流动性和支持性政策两方面来看，我国期权市场处于“萌芽”阶段。因为期权有着杠杆性的特征，大部分投资者没有应用期权最初的初衷，即风险管理。市场中大部分期权交易者都在做增强收入，靠着期权的低成本高收益这个特点。但是高收益往往也带给投资者高风险，尤其像我国期权市场中个人投资者达到了接近一半的比例。投资者不能正确地认识到风险进行盲目的投资，不仅为个人带来了很大的风险，也会对我国的期权市场造成不良的影响。金融监管机构应提高市场参与者的素质，这样才能使我国期权市场稳健地走向成熟。但是市场参与者的教育需要时间。在逐步教育市场参与者并提高其素质的过程中，交易策略会逐渐变得理性。这样既可以使市场规范运行，又可以降低市场参与者的风险同时满足市场参与者的投资需求。要实现投资者的素质的提高，期权交易策略的普及是必不可少的。具有扎实的理论基础的商业策略可以使投资者更理性地看到市场。这些交易策略和定量产品的深度渗透可以使投资者的日常习惯越来越规范化，并有可能有效地抑制投机性交易。多种多样的期权交易策略在维持市场稳定和促进市场成熟方面发挥着非常积极的作用，同时丰富他们的投资决策并实现他们的投资目标。期权投资者的教育应理论与实际相结合，结合实际的盘面形态，学习多种事例，将理论知识与现实相结合，针对不同的行情有着相对应的交易策略和实施细则。其次，投资者的教育培训应不局限于培训这一块，模拟比赛以及短视频教学都是良好的宣传工具。因此，期权交易策略的向投资者的渗透是整个期权市场和资本市场成熟和完善的重要组成部分。

许多品种的定量分析和实时控制为所有投资者消除了不可逾越的技术障碍。得益于互联网和各种移动客户，量化投资策略产品也将逐步标准化，并将在用户的日常投资活动中越来越多地使用。与成熟的海外期权市场相比，中国市场还有很长的路要走。总体而言，在不久的将来，期权市场将在大量市场参与者和学者的共同努力有望进入新的篇章。

5.2.3 完善风险预警体系，保障市场平稳运行

2020 年以来，由于新冠疫情的影响，造成市场波动很大，交易所应制定全面

的市场监管与风控预案，这样能确保了股票期权市场运行平稳，定价合理，风险控制措施有效。优化现有的交易监管程序，加大对程序交易账户的监管力度。继续完善市场间的联动监管，实施全面的联动监管。利用大数据等手段对投资者交易进行交叉分析，及时识别异常交易行为，有效防范了市场风险。

参考文献

- [1] Ahoniemi Katja, Markku Lanne. Overnight stock returns and realized volatility[J]. International Journal of Forecasting,2007,29:592-604.
- [2] Akgiray V. Conditional heteroskedasticity in time series of stock returns: evidence and forecasts[J],Journal of Business,1989,62,55-80.
- [3] Andersen T.G. and T. Bollerslev. Answering the skeptics: Yes standard volatility models do provide accurate forecasts[J], International EconomicReview, 1998,39,885-900.
- [4] Black F. and M. Scholes. The pricing of options and corporate liabilities [J],JournalofPoliticalEconomy,1973.81,637-672.
- [5] Blair B. S. H. and S. J. Taylor, Forecasting S&P 100 Volatility: The incremental information content of implied volatilities and High Frequency ReUirms [J],JournalofEconometrics,2001,105.5-26.
- [6] Bollerslev T. Generalized autoregressive conditional heteroscedasticity[J].Journal of Econometrics,1986,31,307-327.
- [7] BollerslevT, R. Y. Chou and K.P .Kroner, ARCH modeling in finance:a review of theory and empirical evidence[J],Journal of econo-metrics,1992,52,5-59.
- [8] Canina L. and S. Figlewski The Informational Content of Implied Volatility[J] Review of Financial Studies,1993 6(3):659~681.
- [9] Christensen B. J. and N. R. Prabhala. The relation between implied and realized volatility[J],Journal of Financial economics,1998,50,125-150.
- [10]Day T. and C. Lewis, Stock market volatility and the information content of stock index options[J],JournalofEconometrics,1992,52,267-2.
- [11]Day T. and C. Lewis. Forecasting futures market volatility[J],Journal of derivatives,1993,1,33-50.
- [12]Engle R. F. Autoregressive conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation[J], Econometrica. 1982,50(4):987-1007.
- [13]Franses P. H. and D. Van Dijk, Forecasting stock market volatility using (non-

- linear) GARCH models[J], *Journal of Forecasting*,1995,15,229-235.
- [14]Glosten L. R. R. Jagannathan and D. E. Runkle, On the relation between the expected value and the volatility of the nominal excess return on stocks[J], *Journal of Finance*,1993,48,1779-1801.
- [15]Gemmill G. The Forecasting Performance of Stock Options on the London Traded Options Market[J] *Journal of Business Finance & Accounting*. 1986,13(4):535~546.
- [16]John C. H. Options,futures and other derivatives[M].New York: Pearson Education,2008.
- [17]Lamoureux C. G. and W. D. Lastrapes Forecasting Stock Return Variance: Toward an Understanding of Stochastic Implied Volatilities[J] *Review of Financial Studies*.1993,6(2):293~326.
- [18]Luo J. W. and Chen L. N. Multivariate realized volatility forecasts of financial markets based on TVS-MHAR model[J].*Systems Engineering Theory & Practice*,2018,38(7):1677-1689.
- [19]Patton J. ,Sheppard K. Good volatility ,bad volatility: Signed jumps and the persistence of volatility[J].*Review of Economics and Statistics*,2015,97(3):693-697.
- [20]Poon S. and C. W. J. Granger. “Forecasting Volatility in Financial Markets: A Review.” [J],*Journal of Economic Literature* 2003,41(2), 478-539.
- [21]Szakmary A. The Predictive Power of Implied Volatility: Evidence from 35 Futures Markets[J] *Journal of Banking & Finance*.2003,27(11):2151~2175.
- [22]Martens M. and J. Zein, Predicting financial volatility: High-Frequency Time-Series Forecasts v is implied Volatility[J], *Journal of Futures Markets* .2004,24(11):1005-1028.
- [23]Nelson D. B. Conditional heteroskedasticity in asset returns: a new approach[J],*Econometrica*,1991,59,347-370.
- [24]陈芳平, 李松涛. 股指期货推出对股指波动性影响的实证研究[J]. *证券投资*, 2006, 136(2): 45-47.
- [25]成城. 基于 GARCH 模型的上证指数波动率特征分析[D], 山东大学, 2014.
- [26]陈浪南, 杨科. 中国股市高频波动率的特征、预测模型以及预测精度比较[J].

- 系统工程理论与实践, 2013, 33(2):296-307.
- [27] 惠晓峰, 柳鸿生, 胡伟, 何丹青. 基于时间序列 GARCH 模型的人民币汇率预测[J], 金融研究, 2003(05):99-105.
- [28] 黄海南、钟伟. GARCH 模型波动率预测评价[J], 中国管理科, 2007(06):13-19.
- [29] 洪晶晶, 吴鹏跃. 基于 GARCH 模型的金融市场波动性分析与预测[J]. 价值工程, 2016, 35(01):46-47.
- [30] 黄慧. 中国股市波动特征的实证研究—基于 GARCH 族模型[J], 科技创业月刊, 2015(23).
- [31] 黄慧舟, 郑振龙. 无模型隐含波动率模型及其包含的信息: 基于恒生指数期权的经验分析[J], 系统工程理论与实践, 2009(29).
- [32] 李航, 何枫. 沪深 300 波动率预测模型研究: 基于中国股票和期货市场高频数据分析[J]. 广义虚拟经济研究, 2017, 8(04):88-96.
- [33] 李亚静, 朱宏泉, 彭育威. 基于 GARCH 模型族的中国股市波动性预测[J]. 数学的实践与认识, 2003(11):65-71.
- [34] 骆珣, 吴建红. 基于波动率和 VaR 的油价风险计量[J]. 经济管理, 2009, 31(06):106-109.
- [35] 楼迎军, 基于 EGARCH 模型的我国股市 " 杠杆效应 " 研究. [J] 经济论坛, (3):49-53, 2003.
- [36] 刘广应, 蔡则祥, 张新生. 波动率度量方法的比较分析——基于 LHAR-RV-EVT 风险管理[J]. 南京审计学院学报, 2013, 10(06):43-56.
- [37] 龚旭, 文凤华, 黄创霞, 杨晓光. HAR-RV-EMD-J 模型及其对金融资产波动率的预测研究[J]. 管理评论, 2017, 29(01):19-32.
- [38] 龚旭, 曹杰, 文凤华, 杨晓光. 基于杠杆效应和结构突变的 HAR 族模型及其对股市波动率的预测研究[J]. 系统工程理论与实践, 2020, 40(05):1113-1133.
- [39] 康建林, 朱开永, 周圣武, 韩苗. GARCH 模型在中国股票波动观测中的应用[J]. 赣南师范学院学报, 2005(3):29-32.
- [40] 莫海菁. GARCH、GJR-GARCH 和 EGARCH 模型预测能力实证研究[J]. 经济视角(下), 2011(03):60-63.
- [41] 宋逢明, 江婕. 波动率度量模型研究的回顾及展望[J]. 财经论丛(浙江财经大学)

院学报), 2005 (06) :4-9.

[42] 孙德山, 颜妍, 陈芳琪. 基于 GARCH 模型的深证综合指数收益率波动性研究[J]. 辽宁师范大学学报(自然科学版), 2015, 38 (04) :446-451.

[43] 王天一, 赵晓军, 黄卓. 利用高频数据预测沪深 300 指数波动率——基于 Realized GARCH 模型的实证研究[J]. 世界经济文汇, 2014 (05) :17-30.

[44] 文凤华, 刘晓群, 唐海如, 杨晓光. 基于 LHAR-RV-V 模型的中国股市波动性研究[J]. 管理科学学报, 2012, 15 (06) :59-67.

[45] 闫会强, 夏霄松, 金浩. HAR 族模型对波动率的预测精度比较及其 SPA 检验——基于沪深 300 指数高频数据[J]. 经济论坛, 2017 (11) :75-84.

[46] 杨晓辉, 王裕彬. 基于 GARCH 模型的波动率与隐含波动率的实证分析——以上证 50ETF 期权为例[J]. 金融理论与实践, 2019, (5) : 80-85.

[47] 郑振龙, 黄慧舟. 波动率预测:GARCH 模型与隐含波动率[J]. 数量经济技术经济研究, 2010, 27 (01) :140-150.

[48] 赵华, 蔡建文. 基于 MRS-GARCH 模型的中国股市波动率估计与预测[J]. 数理统计与管理, 2011, 30 (05) :912-921.

[49] 周茂华, 刘骏民, 许平祥. 基于 GARCH 族模型的黄金市场的风险度量与预测研究[J]. 国际金融研究, 2011 (05) :87-96.

[50] 赵伟雄, 崔海蓉, 何建敏. GARCH 类模型波动率预测效果评价——以沪铜期货为例[J]. 西安电子科技大学学报(社会科学版), 2010, 20 (04) :27-32.

[51] 张波, 蒋远营. 基于中国股票高频交易数据的随机波动建模与应用[J]. 统计研究, 2017, 34 (3) :107-117.

[52] 朱万锐, 唐大为. 波动性的高频指标对收益分布预测能力的影响分析——基于上证综指的高频数据[J]. 中国市场, 2017 (18) :63-64.

[53] 张瑶. 波动率模型下沪深 300ETF 期权定价实证研究[D]. 山东大学, 2020.

后 记

在兰州财经大学的三年研究生生活即将告一段落，这三年的经历使人成长了不少，一个青涩懵懂的少年变得更加成熟。走进兰财的校园里，清风轻抚枝头绿，落日流金起漪涟，如缕清风吹过枝头绿荫，带来阵阵花香，如同春雨洗过肌肤清新的空气，使人神清气爽。

成长的过程离不开身边每一个老师的帮助，尤其要感谢我的导师—陈芳平教授，不仅在学术上倾尽心血的教授，在人生的选择上也会为我指亮前方的道路。在论文的选题，初稿，定稿每个阶段中，导师都尽心尽力，全心全意；在对未来的选择中，面对自己的困惑，导师也会耐心地告诉他的建议和理解。

每个人人生的旅途上都会遇见不同的人，感谢我的室友张延丰、朱坤炎以及同学们，在研究生的三年里，一起生活，一起学习，人生路途上遇到的每一个人都会形成一段难以忘却的羁绊。以后的人生还会经历风风雨雨，起起伏伏，愿我们如同一叶叶小舟向前行驶，在理想的海洋扬帆，在追求理想的道路上用智慧的箴言书写更多的人生。