

学校代码: 10272
学 号: 2020211076



上海财经大学

SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

MASTER DISSERTATION

论文题目	基于高频订单簿的短期股价预测研究
作者姓名	赵俊杰
院(系所)	金融学院
专 业	金融工程与量化投资
指导教师	谢斐
完成日期	2022/06/16

摘要

高频交易 (High Frequency Trading, HFT) 是指通过市场短暂的价格波动而获利的量化交易策略。对于高频交易而言,能否准确预测短期价格的变动是至关重要的。但大多数传统的股票研究主要集中于对未来股价中长期变动的预测上,而对于未来股价短期变动预测的研究还相对较少。

订单簿作为电子化市场的产物,其中包含了大量的数据与信息。订单簿是一系列价格与需求量集成的产物,订单簿最大的特点便是它可以反应市场参与者在短期内的买卖意愿。尤其是在市场参与者大多没有深入了解、运用订单簿的时候,市场参与者此时并不会刻意去掩饰他们的交易意愿,致使市场参与者的短期交易意愿会不遗余力地暴露在订单簿中。并且由于 level-1 的订单簿数据在实盘交易时是公开的,这就意味着只要市场参与者盯盘,他们就或多或少会注意到盘口数据。但重要的是,虽然市场参与者能获得盘口数据,但实际会分析利用盘口数据的市场参与者并不多,更多的市场参与者反倒会受到盘口数据的影响,做出极度短视、不理智的行为。因此,订单簿数据蕴含的信息量巨大且丰富,它不仅能反应市场参与者的交易意愿,还有助于推测市场参与者的交易行为。因此,本文认为订单簿数据是有助于未来短期股价的预测的。

本文以 2020 年 7 月至 2021 年 6 月的 A 股数据为例,以 level-2 盘口数据为核心数据,从整体与个体的角度出发研究了盘口数据的 2 个方面: 1、从整体出发,利用 Cont (2014) 中的 *OFI* (订单流不平衡) 指标来计算股票在 1tick 内的净流动,从而实现对未来短期股价变动的预测。并认为如果在 1tick 内出现较大规模的净流动,那么股价的变动很有可能会带有一定的持续性,从而影响短期内股价的走势。2、从个体出发,利用 Næs (2006) 中的订单簿斜率指标对多空双方各

自的强弱进行计量,从而实现对未来短期股价变动的预测。并认为买(卖)方斜率大时,市场投资者容易受到影响而改变其卖出(买入)的策略,从而导致股票在下一个 tick 更可能向上(下)变动。在研究过程中,本文发现由于原始指标是基于美股的市场环境提出的,因此这些指标在 A 股上的效果并不理想,因此本文结合 A 股实际情况,针对各类指标所存在的问题提出了针对性的意见并对指标进行了优化,验证结果表明这些修改方案是有效的。

通过对于 2 类指标的深入研究后发现 2 类指标有各自特有的特性,适用的环境也大相径庭,*OFI*指标适合于短期股价变动持续性的预测,因此*OFI*更适合高波动的市场,而订单簿斜率指标却恰恰相反,它更适合对于稳定的市场的股价进行预测。这 2 类指标是矛盾的,因为它们适合于不同的市场环境,但它们又是互补的,也因为它们适合于不同的市场环境。本文认为*OFI*是一个更适合于实战的指标,因为高频策略的成本很高,震荡的市场很容易出现入不敷出的情况,只有预测对短期的趋势才能真正实现获利。

除了理论上的研究外,本文在一定的假设前提下,根据新*OFI*指标提出了一个*OFI*高频策略模型,并用实际的数据对模型进行了回测,并将回测结果与以旧*OFI*指标构建的高频策略模型、同为高频模型的*VPIN*策略模型进行了对比。回测结果表明:相比于*VPIN*策略模型、旧*OFI*高频策略模型而言,新*OFI*高频策略模型在降低风险的同时获得了相对稳定的收益。更重要的是,相比于*VPIN*策略模型而言,新*OFI*高频策略模型对于市场的涨跌并不敏感,无论市场大涨还是大跌,模型都展现出了稳定的收益。但模型对于市场的波动率是敏感的,模型主要的 2 段回撤(2020.09-2020.10 以及 2021.03-2021.04)都是在市场整体波动率较低的时期发生的,但虽然模型的回撤期较长,但回撤幅度并不高。值得注意的是,市场在 2021.02-2021.03 时出现了极强的风格切换,并且由于选股的方法存在一定的滞后性而使得模型无法及时识别到该问题。但即便如此,模型在这段时间内

也很好的控制住了风险，这展现出了模型的强鲁棒性。

最后，本文也对模型的参数敏感性与成本进行了讨论，得出了以下结论：1、模型对于平滑窗口的选择、股票数量并不敏感，但对于阈值的选择有一定的敏感性。2、目前模型是基于盘口数据进行的，利用逐笔数据进行测试或许可以更准确地把握开仓平仓点，降低模型的交易成本，从而提高模型的收益表现。

关键字：订单流不平衡(*OFI*) 订单簿斜率 统计检验 *OFI*高频模型

Abstract

High frequency trading (HFT) refers to a quantitative trading strategy that makes profits through short-term price fluctuations in the market. For high-frequency trading, it is very important to accurately predict the short-term price changes. However, most traditional stock studies mainly focus on the prediction of medium and long-term changes in the future stock price, while there are relatively few studies on the prediction of short-term changes in the future stock price.

As the product of electronic market, order book contains a lot of data and information. Order book is the product of a series of price and demand integration. The biggest feature of order book is that it can reflect the buying and selling intention of market participants in the short term. At this time, most of the participants will not deliberately use the order book to cover up the market, especially when they do not have the will to go deep into the market. Moreover, since the order book data of level-1 is open during the firm trading, it means that as long as the market participants keep an eye on the order, they will more or less notice the opening data. However, it is important that although market participants can obtain the opening data, there are not many market participants who will actually analyze and use the opening data. On the contrary, more market participants will be affected by the opening data and make extremely short-sighted and irrational behavior. Therefore, the information contained in the order book data is huge and rich. It can not only reflect the trading intention of market participants, but also help to speculate the trading behavior of market

participants. Therefore, this paper believes that the order book data is helpful to the prediction of short-term stock price in the future.

Taking the A-share data from July 2020 to June 2021 as an example, and taking the level-2 data as the core data, this paper studies two aspects of inventory data from the perspective of the whole and the individual: 1. From the whole, the *OFI* (order flow imbalance) index in Cont (2014) is used to calculate the net flow of stocks in 1 tick, so as to predict the future short-term stock price changes. It is believed that if there is a large-scale net flow in 1 tick, the change of stock price is likely to have a certain sustainability, which will affect the trend of stock price in the short term. 2. Starting from the individual, the order book slope index in Næs (2006) is used to measure the strength of both long and short sides, so as to predict the short-term stock price change in the future. It also believes that when the buyer (seller) slope is large, market investors are easy to be affected and change their selling (buying) strategy, resulting in more upward (downward) changes in the stock in the next tick. In the research process, this paper found that because the original indicators are proposed based on the market environment of US stocks, the effect of these indicators on a shares is not ideal. Therefore, combined with the actual situation of a shares, this paper puts forward targeted opinions on the problems of various indicators and optimizes the indicators. The verification results show that these modification schemes are effective.

Through the in-depth study of the two types of indicators, it is found that the two types of indicators have their own unique characteristics and different applicable environments. *OFI* indicator is suitable for the prediction of the sustainability of short-term stock price changes. Therefore, *OFI* is more suitable for the high

volatility market, while the order book slope indicator is just the opposite. It is more suitable for the prediction of stock price in a stable market. These two kinds of indicators are contradictory, because they are suitable for different market environments, but they are complementary, and because they are suitable for different market environments. This paper believes that *OFI* is a more suitable index for actual combat, because the cost of high-frequency strategy is very high, and it is easy to make ends meet in a volatile market. Only by predicting the short-term trend can we really make profits.

In addition to the theoretical research, under certain assumptions, this paper proposes an *OFI* high-frequency strategy model according to the new *OFI* index, and back tests the model with the actual data, and compares the back test results with the high-frequency strategy model constructed with the old *OFI* index and the *VPIN* strategy model which is the same high-frequency model. The back test results show that compared with the *VPIN* strategy model and the old *OFI* high frequency strategy model, the new *OFI* high frequency strategy model not only reduces the risk, but also obtains relatively stable benefits. More importantly, compared with the *vpin* strategy model, the new *OFI* high frequency strategy model is not sensitive to the rise and fall of the market. No matter the market rises or falls, the model shows stable returns. However, the model is sensitive to the market volatility. The main two-stage pullback of the model (2020.09-2020.10 and 2021.03-2021.04) occurs in the period of low overall market volatility. However, although the pullback period of the model is long, the pullback range is not high. It is worth noting that there is a strong style switching in the market from February 2021 to March 2021, and the model cannot identify the problem in time due to the lag of the stock selection method. But

even so, the model has well controlled the risk during this period of time, which shows the strong robustness of the model.

Finally, this paper also discusses the parameter sensitivity and cost of the model, and draws the following conclusions: 1. The model is not sensitive to the selection of smooth window and number of shares, but it is sensitive to the selection of threshold. 2. At present, the model is based on the opening data. Using the test data one by one may more accurately grasp the opening and closing points, reduce the transaction cost of the model, and improve the income performance of the model.

Key words: order flow imbalance(*OFI*) Order Book slope statistical test *OFI*
high frequency model

目 录

第一章 绪论.....	1
第一节 研究背景与研究意义	1
一、研究背景.....	1
二、研究意义.....	5
第二节 研究内容与创新点.....	6
一、研究思路与方法.....	6
二、论文框架.....	7
三、论文创新点.....	8
第二章 文献综述	9
第一节 金融市场微观结构研究	9
第二节 订单流不平衡指标研究	11
第三节 订单簿斜率指标研究	12
第三章 基于订单簿的短期股价预测模型	14
第一节 数据选取与预处理.....	14
第二节 订单流不平衡指标.....	16
一、订单流不平衡指标定义.....	16
二、订单流不平衡指标的统计检验.....	18
三、订单流不平衡指标的缺陷与改进优化.....	20
四、OFI 的回归优化	23
五、OFI 的准确率	25
六、小结.....	26
第三节 订单簿斜率指标.....	26
一、订单簿斜率指标的定义与缺陷.....	26
二、新订单簿斜率指标的统计检验.....	29
第四节 OFI 与订单簿斜率指标的关系	30
第五节 本章小结.....	31
第四章 基于 OFI 的高频策略模型.....	32
第一节 模型假设.....	32
第二节 模型介绍.....	33
一、OFI 高频模型介绍	33
二、VPIN 策略模型介绍	36
第三节 模型结果及评估.....	37
一、模型评价指标.....	37
二、模型结果与评估.....	38

三、模型缺点	41
第四节 模型参数敏感性检验	41
一、阈值敏感性检验	42
二、指数平滑数据设置敏感性检验	43
三、股票标的池大小	44
第五节 模型表现成本分析	46
第六节 本章小结	47
第五章 总结与展望	49
一、总结	49
二、展望	49
参考文献	51

第一章 绪论

第一节 研究背景与研究意义

一、研究背景

量化交易是一种利用计算机技术从庞大的历史数据中海选能带来超额收益的多种“大概率”事件以制定策略，极大地减少投资者情绪波动影响的策略。随着计算机设备的功能与性能的飞速发展，证券交易的形式也完成了由线下交易到线上交易的转变，至此之后，越来越多的市场参与者开始逐步尝试利用量化交易来辅助交易，市场参与者的下单手段也在由手动下单迈向程序化下单。截至目前，量化交易已经屡见不鲜了。但与发达国家的股市不同，中国股市有着大量的散户投资者。长期以来，散户投资者都是 A 股的重要参与者。但由于每位投资者自身的能力与精力毕竟是有限的，散户投资者没有办法像机构投资者那样提供完备的设备，也没有强大的研究能力与时间精力，这也是阻碍国内量化交易发展的原因之一。但经历了 2020 年的基金大年、2021 年的量化大年之后，投资于机构的散户愈发增加，散户投资者的比例开始减少，机构投资者的占比开始提升，这也推动着量化交易的飞速发展。和常规的策略一样，量化交易中也有着高频策略与中低频策略，高频策略无论是在设备支持上还是技术支持上，都有较高的需求，因此更多的量化策略集中于中长期，但用的人多了，策略也就开始逐渐失效了。此时，高频交易领域愈发受到人们的关注。高频交易作为特殊的量化交易，是指从那些人们无法利用的极为短暂的市场变化中寻求获利的计算机化交易策略。早在几年前，高频交易在众多发达国家中就已经占据了半壁江山，如：美国

市场 60%以上的交易都是由算法与高频交易贡献的，日本市场 15%以上的交易也都是由算法与高频交易贡献。与之相比，高频交易在中国市场上仅仅才刚刚开始崭露头角，有着极大的发展空间，因此迫切需要针对中国市场的高频数据进行分析研究。

尽管高频交易在发达国家已经颇有成效，但值得注意的是，研究者对于高频交易的态度依然是褒贬不一的，部分研究者认为高频交易可以增加市场流动性，从而减少交易中的买卖价差，达到降低日内波动的目的；但另一部分研究者恰恰相反，他们认为高频交易者过快的交易速度会使得其他市场参与者处于极其劣势的地位，并且高频交易者可能会加剧新闻所产生的波动，如：当一则坏消息传开了，高频交易者会利用速度优势，快速大量的进行抛售，给市场带来极大的压力，从而影响其他市场参与者的行为，最终扩大新闻产生的波动。

相比于普通的量化交易，高频交易就是与时间赛跑，很可能机会只在短短的几秒钟，一旦耽搁几秒钟，可能一笔获利的交易就会扭盈为亏，因此高频交易只能通过计算机来实现交易，并且高频交易对于计算机的硬件需求与软件需求也极高。不难发现，很多从事高频交易的投资者都会配备性能极佳的设备，并且争取尽可能将服务器放置在靠近交易所的位置，只为了能减少些许的延迟。

信息技术的进步是高频交易发展的必要条件，技术的进步为其提供了硬件设备支持，而在技术发展的同时产生的产物为其提供了软件支持——它就是限价订单簿（下文称之为订单簿）。限价委托是指按客户确定的价格或履约时间的价格进行交易的订单。而限价订单簿，即是以不同价格记录限价委托单的系统，包含了市场参与者提供的最微观的交易信号。随着信息技术的进步，订单簿数据的颗粒度愈发细腻，从 level-1 数据到 level-2 数据到逐笔交易数据，这为高频分析研究提供了充足的数据支撑。同时，相比于美国而言，中国股市并没有隐藏单的概念，因此市场参与者的交易需求在订单簿中得到了最大的还原。

股票的研究分析一直是金融经济学领域内研究的热点话题之一。股票的研究角度可以分为3类：宏观研究、中观研究、微观研究。宏观研究与中观研究更加注重股票的基本面信息，其中涉及国家政策、公司基本面等，而微观研究更加注重股票价格变动本身。微观结构理论主要研究的角度有两大类：(1)、市场风险预测的研究。这一领域内很常见的研究即是对于波动率的预测。诚然，宏观研究也会对于股票的波动率进行研究分析，但宏观研究无法做到微观研究那么快速、那么细腻。(2)、市场价格发现机制的研究。该研究领域也是整个金融经济学领域内研究的热点问题之一。该领域内很常见的研究即是对于价格形成机制的研究，其中诞生了很多耳熟能详的模型，如：存货模型、信息模型等。无论是对于风险的控制还是股价的预测，微观结构理论都展现出了很重要的地位。

宏观研究与中观研究是股票研究中不可或缺的领域，但由于它们的研究方法与研究角度是相对稳定的，更加重要的反倒是不同的研究内容。相比之下，微观研究更注重的是股价变动本身。随着信息技术、大数据技术、人工智能技术的飞速发展，以及高频交易的迅速崛起，金融资产的微观结构与过去发生了翻天覆地的变化。技术上，越来越多的跨领域模型被不断提出，其中涉及到了数学模型、物理学模型、力学模型等；数据上，现有的数据有着过去的难以媲美的颗粒度。微观研究的研究无论是在研究方法、研究角度、研究内容上都发生了翻天覆地的变化，相比于宏观研究，微观研究迫切地需要不断地进行深入研究。

相较于过去使用的日频数据、分钟频数据而言，数据上最大的变化就是限价订单簿数据了，限价订单簿数据将数据的颗粒度提升到了1tick(3s)，这样的数据颗粒度是过去所不能想象的，因此限价订单簿数据已然成为了当下微观结构理论中的数据基石，协助微观结构理论由低频到高频的转型。

本文将要针对微观结构理论中极其重要的一大领域——市场价格发现机制的研究，进行深入研究。这一领域的研究角度与研究成果都是非常丰富多样的，

但值得注意的是,虽然研究的角度十分多样,但研究过程所基于的深层逻辑是类似的。其中,有一类深层逻辑认为股价之所以存在着差异是因为股票的不平衡所导致的。而在这一逻辑下衍生出了多个不同的研究角度,如:经典的信息模型认为股票的不平衡是知情交易者与非知情交易者博弈的结果,在该角度下衍生出了很多的研究成果,如:Easley(1996)中的 PIN (知情交易概率)、Easley(2012)中的 $VPIN$ (等交易量知情交易概率);还有一类模型认为股票的不平衡是多空买卖意愿差异导致的结果,这也正是本文所要介绍的角度,其中最经典的即是Cont(2014)中的 OFI 指标。2个角度的差异仅仅在于它们对于股票的不平衡的解释上,它们的深层逻辑是不变的。但正因为这种解释上的差异,导致同类型的指标的适用范围却大相径庭, $VPIN$ 需要时间刻度、交易量的选取不能太短、太小,过短的时间会使 $VPIN$ 无法很好衡量知情交易者的占比,但 OFI 需要较短的时间刻度,因为时间过长会使得买卖意愿的可持续性减弱。

每一类指标都有它的合理性、逻辑性,但出于更高频的指标可能有更好的预测效果上,本文将以频率更高的 OFI 作为核心研究角度、以多空双方买卖意愿作为底部逻辑进行短期股价变动预测问题的研究。同时,订单簿研究过程中有一类指标是从买卖双方各自的角度出发,研究订单簿对于市场参与者行为的影响,如:订单簿斜率指标。本文也将对于该类指标进行介绍,来完善对于限价订单簿数据的研究。

值得注意的是,中国股市目前采用T+1交易制度,这使得高频交易显得看似难以实现,但投资者完全可以通过在前一天提前进行建仓,从而实现当日的T+0交易,完成高频策略。高频交易中较为常见的策略有算法交易、做市策略等。无论是哪一种高频策略,最基础的数据永远都是订单簿数据,最核心的难点在于数据的分析以及是否能抢单成功。

二、研究意义

高频策略主要的 2 个难点在于未来短期时间内股价变动的预测以及抢单是否能成功。本文暂时不考虑抢单的问题，主要研究短期股价变动预测。虽然随机漫步理论认为股价的变动是随机的、不可预测的，但或许是由于机构投资者的集体买入或卖出、亦或许是散户投资者的短视行为致使的跟风，这使得股价在极短的时间内或许是可以预测的，因此本文将从各类角度出发分析订单簿数据，研究订单簿是否有助于提升短期股价变动的预测能力。

Cont (2014) 构建了订单流不平衡指标(下文称之为 OFI)，并利用标准普尔 500 的股票证实了 OFI 对于同期的中间价变动有较强的解释性。但 Chi (2021) 证实了 OFI 对于下一期的中间价变动的解释性不足并给出了解释，即 OFI 是一个很好的同期指标，但其没有什么预见性。导致这样的结果是因为 OFI 指标的计算方程所致的，因此本文将在利用 OFI 指标的设计思路的基础上，对于 OFI 进行优化，并将其直接运用于历史数据中，验证其效果。这将作为本文研究的第一个方向，也作为本文研究的核心方向。

Naes (2016) 构建了订单簿斜率指标，并利用挪威股市证实了订单簿斜率对于股价波动率有很好的解释力，但在 Naes (2016) 中，作者本人也提出了该指标所存在的问题。本文将尝试在该指标的构建思路之上，优化订单簿斜率指标，解决目前指标存在的问题，并将其用于短期股价变动的预测上，验证其效果。

本文认为订单簿中隐藏着大量的指示性信息，这些指示性信息是有助于短期股价的预测的，但传统的定义方式或许不太适用于当下的市场环境。比如：

OFI 衡量了一段时间内市场参与者的净买入意愿或净卖出意愿的强弱， OFI 为正(负)，意味着市场参与者更愿意在该段时间进行买入(卖出)， OFI 数值越大，意味着市场参与者的单边意愿越强烈，从而对当前时刻的股价产生影响。当 OFI

足够大时，市场的买卖意愿强烈且容易延续，此时在短时间内，市场的买卖意愿很有可能会存在延续，从而对下一段时刻的股价产生影响；当 OFI 在 0 附近时，这种延续性可能并不强烈。因此 OFI 对于短期内买入卖出意愿的强弱具有一定的指示性，即对短期内未来价格的走势具有一定的指示性（买入意愿越强，意愿越有可能延续，则价格越有可能上升，反之亦然）。

订单簿斜率考察了订单簿对于市场参与者给予的压力，比如：如果买方的订单簿斜率远大于卖方，就意味着愿意限价买入的人越多，因此卖方压力较大，很多的散户可能会因此而做出跟涨的操作，亦或是以更优的价格进行限价卖出的操作，从而导致短期内未来股价出现上升，反之亦然。

看似 2 个角度是类似的，但各个角度看问题的方向是不同的： OFI 是从订单簿整体出发，考察市场参与者的买卖意愿来判断价格走势；而订单簿斜率是从买卖双方分别出发，以压力的强弱判断价格走势。但无论如何，订单簿的数据可能可以为短期股价变动预测提供帮助。

本文意图提出订单簿中较为重要的 2 个研究角度，并对指标存在的缺陷进行提出与改进，便于为后续的理论研究提供一些思路。同时，本文会尝试利用订单簿指标构建高频策略，利用历史数据对策略进行验证，便于为后续的实战策略研究提供思路。

第二节 研究内容与创新点

一、研究思路与方法

本文从订单簿数据出发，试图从订单簿数据中挖掘出有利于短期股价变动预测的信息，但订单簿研究最大的难点就在于订单簿的数据量及其研究维度非常大，选择哪个方向来研究、用什么数据来研究都是首要问题。但虽说订单簿数

据量很大，但数据的主要构成要素无非就是 level-1 数据、level-2 数据、逐笔成交数据、逐笔委托数据，而这 4 类数据又可以划分为快照数据与逐笔数据。在实盘中，快照数据是散户投资者可以观察到的数据，并且这类数据在一定程度上对于散户投资者起到引导作用，可能会诱导散户的行为，即常说的主力诱多或主力诱空，因此考虑从快照数据出发进行本文的研究。但值得主要的是，只有逐笔数据能最真实的还原每个参与者的交易欲望，而快照数据是不能做到的，因为快照数据是将发布间隔内发生的所有交易进行汇总后得到的数据。即便如此，快照数据也已经能在很大程度上胜任这项工作。

确定完使用的数据后，通过阅读文献，确定了订单流不平衡指标以及订单簿斜率指标作为研究方向，因为这 2 种指标都能从不同的角度反应出整个市场参与者对于股票的态度，并可能可以协助预测未来短期股价的变动。

本文在研究每个指标的过程中，会从高市值、中市值、低市值的标的中各选 10 只最活跃的标的，分析指标对于不同市值的标的的效果是否有差异。

本文用来验证指标有效性的工具以线性回归为主，利用回归得到的 T 统计量以及 F 统计量来验证指标以及方程的显著性，并利用回测来验证指标在历史数据上作用的效果。

二、论文框架

本文行文框架如下：第一章主要介绍了研究背景、研究意义、研究思路与方法、论文框架以及论文创新点，引出本文研究的方向与目的。第二章主要介绍了与研究角度相关的研究成果。第三章将会对所用的数据进行介绍，并且重点介绍 2 个指标的定义、缺陷以及改进，并将其与未来一段时间的股价变动进行回归，从统计学角度给出结论——指标对于预测未来股价短期变动是否显著有效。第四章将会以第三章的研究成果为基础，并根据手续费问题增加一定的改进与约

束,形成一个高频策略,利用历史数据对策略进行回测,并将策略与 2 个类似策略进行对比。第五章将会对全文进行系统性的总结,并针对文章存在的问题以及可以优化的点进行探讨。

三、论文创新点

本文的创新点主要体现在 3 个方面:

1、改进了 *OFI* 指标, Cont (2014) 中所定义的 *OFI* 指标是合理的, 并且是显著有效的同期指标, 但并不是一个非常显著的预测指标, 本文针对 A 股特有的特点, 通过对定义进行修改, 提高了 *OFI* 指标对未来短期股价变动的预测能力。

2、改进了订单簿斜率指标, Naes (2016) 中提出的订单簿斜率指标存在着一定的缺陷, 本文提出了一种可行的思路, 有助于未来进一步的研究。

3、本文根据上述改进后指标, 基于订单簿 level-2 数据, 构建出了一个高频策略, 并在忽略了成交问题的前提下, 给出了策略的效果。并与原指标下的模型与经典的 *VPIN* 策略模型进行了对比, 发现模型有显著优秀的表现。

第二章 文献综述

本章将从金融市场的微观结构研究、订单流不平衡研究以及订单簿斜率研究 3 个角度对国内外的相关研究成果进行介绍。

第一节 金融市场微观结构研究

微观结构研究是股票研究中极其重要的一大分支，但由于微观结构所涉及到的数据十分细腻、模型十分复杂，导致微观结构的研究并没有宏观研究那么完备。微观结构的产生源于德穆塞兹 1968 年发表的《交易成本》，但真正开始被学者重视是源于 1987 年 10 月的股市暴跌。时至今日，微观结构的研究时间也并不算长，尤其对于起步较晚的中国来说，更是如此。目前，A 股的飞速发展迫切需求对于股票微观结构的研究。

微观结构研究领域内有两类模型最为经典也最被学者所熟知——存货模型与信息模型，后续研究过程很大一部分研究也是基于这 2 类经典模型进行的。Garman (1976) 中提出了早期最经典的存货模型之一的 Garman 模型，在泊松分布的假设下，作者提出了设定不同的买卖价格是做市商最佳的手段，也是做市商立于不败的原因，从而得出结论：市场的微观结构决定了股价的变动；O'Hara (1999) 基于报价驱动市场建立了较为完善的微观结构理论基础，并提出用存货模型对于股价的变动进行分析；Bagehot (1971) 在研究了买卖价差与信息之间的关系后，发现做市商会根据市场中的知情交易者的数量来设定价格，从而弥补做市商的风险。从而在文中首次提出了信息不对称在价格形成机制中的重要性；O'Hara (1987) 引入了信息不确定性的概念，并认为当交易量增大

时，做市商可以从交易中获得更多的信息，价差会因此减小；Easley(1996)以信息不对称作为指标构建的基础理论，以成交量、买卖价差等低频数据为基础，提出并构建了 PIN (知情交易概率)指标；Hasbrouck(2003)同O'Hara(1999)一样，也基于报价驱动市场建立了较为完善的微观结构理论基础，同时提出了用信息模型对于股价的变动进行分析。金融市场的微观结构的研究已经初见雏形，但不幸的是，这一切都因为市场的快速变化而发生了翻天覆地的变化。

随着技术的发展，股票实现了从报价驱动市场到订单驱动市场的快速转换，市场的特征发生了翻天覆地的转变。这不仅仅是交易模式的转变，更是交易主体的转变，市场上的交易者不再是做市商与投资者，而变成了投资者与投资者，数据也从低频的数据转变成颗粒度更加细腻的订单簿数据。这一切的变化都推动着微观结构研究的进一步改革进化。

在这样的环境下，越来越多的微观结构研究开始尝试从高频的角度出发进行研究。Cont(2013)以马尔可夫过程为理论导向，提出用限价订单簿计算股票的转移概率，从而实现最终的股价变动预测；Palguna(2012)利用限价订单簿数据计算出市场深度来优化异质性自回归模型，发现模型可以更好地反应订单的历史信息对于股价的潜在影响；Easley(2012)以高频数据为基础，升华了Easley(1996)中提出的 PIN 指标，提出了 $VPIN$ (等交易量知情交易概率指标)，并利用高频数据验证了其有效性；刘波(2009)从指令流与限价指令簿之间相互作用的角度研究了双向拍卖交易机制下的短期价格行为和交易量的动态形成过程及其统计性质；Cont(2014)中率先提出了 OFI 指标的概念，并以level-1数据为基础验证了订单流指标的有效性；O'Hara(2015)中作者强烈提议着眼于高频世界进行微观结构的研究，并为此提出了一系列需要进行研究探索的新主题；Chi, Zhao, Huang(2021)通过hwakes模型与订单簿毫秒级高频重构动态计算订单速率，通过分析模型参数和成交概率发现了中国股票市场微观结构的高

频特征。

第二节 订单流不平衡指标研究

Bouchaud(2003)研究了巴黎股市中股票的股价与交易和报价数据之间的关系,并提出一种观点:价格的随机游走性质是由流动性提供者与流动性接受者竞争产生的结果,即价格的形成是由交易数据与报价数据共同决定的,只考虑交易数据无异于只考虑了交易形成机制中的一部分,而忽视了另一部分。Cont(2011)从实证以及理论的角度证实了订单流、流动性以及价格变动之间的相关性,并在Cont(2014)中率先提出了 OFI 指标的概念,并用最优出价、最优报价对 OFI 进行了计算。同时,作者在Cont(2014)中研究了交易不平衡指标(只考虑市价单)、 OFI (同时考虑市价单、限价单、撤单)和股价变动的关系,得到了与Bouchaud(2003)相似的结论:只考虑市价单对于股价的预测是不完备的,并认为这是因为只考虑市价单意味着将限价单与撤单作为模型中的噪音,只有当限价单与撤单的值足够小的时候,两个指标才会由相仿的效果。Cont(2014)利用标准普尔500成份股中的随机50只股票进行了 OFI 的显著性检验,回归后平均的 R^2 大约为65%,有98%的样本在5%的显著性水平下表现显著,验证了 OFI 对于同期价格变动有很强的解释性。在同文中的附录中,作者检验了利用其他出价、报价数据计算的 OFI 对于价格变动的解释性,得到了以下结论:二级出价、报价计算的 OFI^2 可以提升 R^2 大约6%,但其他档次的价格并不能显著提升 R^2 ,即一档、二档数据就可以很大程度上完成对于价格变动的解释。Shen(2015)中也验证出最优报价以及二级报价就已经可以解释价格变动中的很大一部分,同时表明提高频率有助于增加 OFI 的解释力度——在10s左右时可以获得和Cont(2014)中类似的结果(60%左右的 R^2)。Shen(2015)也将 OFI 指标直接运用于中金所的沪深300

股指期货中,获得了具有显著意义的正收益。但不幸的是,股票可能很难直接获得这样的效果。原因如下:股指期货的成本(0.0025%)远低于股票(至少0.1%的印花税);并且该测试的数据取自2014年的期货数据,当时的期货市场的性质与现在的股市的性质差异较大。Xu, Gould, Howison (2019)延续了Cont (2014)附录中的研究,研究了多级 OFI 对于股价变动的预测能力,文中解释了Cont (2014)附录中获得的结果, Xu, Gould, Howison (2019)认为之所以线性回归只得到一级、二级报价对于股价变动解释力强是因为不同级别报价计算出的 OFI 具有较强的多元自相关性,而岭回归可以在一定程度上克服该问题。在岭回归下,虽然并不是所有级别的 OFI 都呈现出显著性,但其显著性明显强于线性回归的结果。

与上述研究不同的是, Chi, Zhao, Huang (2021)将 OFI 用于A股上,但并未得出与Cont (2014)一致的结论,作者认为这是因为美股在70%的时间处于 OFI 均衡,因此 OFI 的失衡就很大程度上预示着价格的变动,而A股由于有撤单限制(撤单过于频繁会限制未来的交易)以及没有做市商的存在,导致长期处于不平衡状态,因此很难用 OFI 用来直接对股价的变动进行预测。

第三节 订单簿斜率指标研究

Lee, Mucklow and Ready (1993)对于最优出价、最优报价上的斜率以及订单深度进行了探讨,并对于收益公告日发布前后最优出价、最优报价上的斜率进行了讨论。Kalay (2004)从股票供需弹性的角度出发,发现股票在刚开盘时拥有更强的弹性,并认为这是因为刚开盘时市场参与者有着很强的信息不平衡,而这种不平衡会随着时间而降低。同时, Kalay (2004)发现买方的弹性要强于卖方,即买方对于价格的影响强于卖方,并认为这是因为做空限制所致。Næs (2006)设计了订单簿斜率,并得到了订单簿斜率、价格波动、金融分析师对资产价值的信念

分散有系统性的关系，即订单簿斜率越平稳，出价单与报价单就会越分散，则意味着金融分析师对资产价值的信念越分散，股票波动率就会越大。作者以挪威证券交易所中的一只蓝筹股与一只年轻的科技公司为例，说明了上述论点。但作者在文中也提到了订单簿斜率存在的缺陷，公式的第一项式与后面几项的单位是不相称的，这也意味着直接相加是不合理的，因此作者提出了多种优化方案：1、取对数(这也是文中所用到的)。2、用 Kim(2004)中提到的用最优报价除以报价队列数量总和。但即便如此，这一问题也没有得到彻底的解决，因此本文将对该问题提出一种改进思路。

第三章 基于订单簿的短期股价预测模型

第一节 数据选取与预处理

本文选取了交易跨度从 2020 年 7 月至 2021 年 6 月为期一年的 A 股作为本文的原始数据集。选取该交易时段的主要原因是因为这一年的 A 股具有一定的代表性：

从行情上来看，整个 A 股在这一年期间经历了连续的大涨、震荡、连续的大跌 3 个最重要的时期。以中证 800 为例，中证 800 在 2020 年 7 月上半月、2021 年 1 月初都出现了较高的涨幅，幅度达到 17%与 8%，而在 2 月底 3 月初出现了大幅度下挫，幅度为 15%，而在其他时期表现出了震荡的走势。从风格上来看，以 2021 年 2 月为分水岭，在这之前，机构抱团情况较为强烈，大盘股在该阶段显得格外强势，而在这之后，市场风格突变，机构的抱团意向减弱，大盘股不再显得那么强势，因此这一年时间展现出多种市场风格，是有代表性的。

在标的的选取上，在初始股票池的基础之上，将 A 股中半年内新上市的次新股以及半年内出现停牌等特殊事宜的股票予以剔除，将剩下的股票作为核心股票池，用于后续进一步的筛选。当然，核心股票池也可以由中证 800 代替，但是由于中证 800 每半年更新一次，频率较低，因此其准确性会随时间而降低。标的的具体选取过程如下：

首先，将核心股票池中的股票按市值进行排序，并取排名前 50 的股票作为大盘股标的池，排名中间 50 的股票作为中盘股标的池，排名后 50 的股票作为小盘股标的池。没有直接将核心股票池一分为三的主要原因是因为直接一分为三会导致标的池内的股票差异性过强。

其次, 参照(赵景东 2013)中活跃度筛选标的方法, 用每只股票前一个月的行情表现来计算出每只股票的活跃度, 并根据活跃度从每类标的池中再挑选出 10 个最活跃的股票, 如: 从大盘股标的池中选取 10 只最活跃的股票, 作为大盘股的验证标的池(下文所说的标的池, 即是指验证标的池)。并在每个月的月初, 对标的池进行滚动更新。活跃度识别的流程如下:

活跃度参考了活跃度(赵景东 2013)的定义, 是由股票前一个月的日平均交易量指数 VI 和日平均振幅 AT 来决定的:

$$VI = \sum_i^n \frac{Volume_i}{\frac{Total_share_i}{n}} \quad (3.1)$$

$$AT = \sum_i^n \frac{\frac{P_{Max} - P_{Min}}{P_{Open}}}{n} \quad (3.2)$$

其中, $Volume$ 是股票的成交量, $Total_share$ 是股票的总股数, P 是股票的价格, n 为一个月的交易天数。

在算出每个股票的 VI 与 AT 后, 分别计算出股票在对应标的池中的截面排名, 将截面排名相加后的结果作为股票的活跃度排名, 选出排名最高的十只股。

表 1.1 展示了各类标的池的描述性统计量。由于本文股票池的选择是滚动的, 展示所有结果会过长, 这里只选取第一个月标的池的描述性统计量进行展示:

表 2.1 标的池描述性统计量

	日平均交易次数			日平均买一价更新次数			日平均卖一价更新次数		
	大盘股	中盘股	小盘股	大盘股	中盘股	小盘股	大盘股	中盘股	小盘股
平均值	89842	22804	7199	2439	1104	520	2396	1072	519
最大值	190220	42120	10187	3008	1769	853	2931	1735	874
最小值	29723	6593	4154	1215	362	44	1199	359	44

标准差	50988	12672	1952	583	491	254	564	476	256
-----	-------	-------	------	-----	-----	-----	-----	-----	-----

第一列为日平均交易次数，第二列为日平均买一价更新次数，第三列为日平均卖一价更新次数。这三类指标的目的与 VI、AT 的目的是相似的，它们也是用来衡量股票活跃性的。但不同的是，VI、AT 是从股票一天整体的日活跃度来分析股票活跃性，而这 3 类指标是从更微观的角度分析股票短时间内的活跃性。从表中可以发现，即使选取的股票都是各自股票标的池中最活跃的 10 只，但它们之间的差异性依然是显著的。股票的局部活跃度明显与股票市值有较强的相关性，高市值的股票在局部变化上远比低市值股来的活跃。

第二节 订单流不平衡指标

一、订单流不平衡指标定义

正如文献综述中所提到的，*OFI* 指标是由 Cont (2014) 在研究了市价委托、限价委托、撤单与股价变化值之间的关系后提出的概念。在 Cont (2014) 中，作者在假设任何委托只发生在一档买入、一档卖出的前提下，提出了价格变动方程：

$$\Delta P_k^b = \delta \left[\frac{L_k^b - C_k^b - M_k^s}{D} \right] \quad (3.3)$$

$$\Delta P_k^s = -\delta \left[\frac{L_k^s - C_k^s - M_k^b}{D} \right] \quad (3.4)$$

$$\Delta P_k = \frac{\Delta P_k^b + \Delta P_k^s}{2\delta} = \frac{L_k^b - C_k^b - M_k^s - L_k^s + C_k^s + M_k^b}{D} + \varepsilon_k \quad (3.5)$$

其中， δ 表示股价的最小变动单位，上标 *s* 代表卖出，上标 *b* 代表买入。在时间区间 $[t_{k-1}, t_k]$ 内， $\Delta P_k^b = P_k^b - P_{k-1}^b$ 为期间股票买一价的变化值； L_k^b 、 L_k^s 为期间的限价单之和； C_k^b 、 C_k^s 为期间的撤单之和； M_k^b 、 M_k^s 为期间的市价单之和； D 代表期

间内的市场深度； ε_k 为误差项。

作者将式(3.5)的分子定义为 OFI_k ，即：

$$OFI_k = L_k^b - C_k^b - M_k^s - L_k^s + C_k^s + M_k^b \quad (3.6)$$

OFI 直观地衡量了 $[t_{k-1}, t_k]$ 内订单流的流动方向以及流动净规模，它最直接的反应出了买方与卖方的强弱。 OFI 为正，代表着买方强势，并且 OFI 的数值越大，意味着买方越强势，当 OFI 的数值大小超过市场深度时，价格将向上变化；反之， OFI 为负，代表着卖方强势，并且 OFI 的数值越大，意味着卖方越强势，当 OFI 的数值大小超过市场深度时，价格将向下变化。因此，作者最终得出股价的变动很大程度上可以由 OFI 来解释，并利用了标普 500 验证了这一观点。

由于逐笔交易数据的复杂性以及难获取性，作者提出了另一种基于快照数据计算 OFI (下文所提及的 OFI 都是指基于快照数据的 OFI)的方式：

$$e_n = q_n^b 1_{\{p_n^b \geq p_{n-1}^b\}} - q_{n-1}^b 1_{\{p_n^b \leq p_{n-1}^b\}} - q_n^s 1_{\{p_n^s \leq p_{n-1}^s\}} + q_{n-1}^s 1_{\{p_n^s \geq p_{n-1}^s\}} \quad (3.7)$$

$$OFI_k = \sum_{n \in (t_{k-1}, t_k]} e_n \quad (3.8)$$

其中， p^b 表示买一价， q^b 表示买一量， p^s 表示卖一价， q^s 表示卖一量，

$$1_{\{p_n^b \geq p_{n-1}^b\}} = \begin{cases} 1 & p_n^b \geq p_{n-1}^b \\ 0 & p_n^b < p_{n-1}^b \end{cases} \quad (3.9)$$

。 e_n 表示时刻 n 的 OFI ，当需要一个时间段内的 $OFI_k((t_{k-1}, t_k])$ 时，只需要将时间段内的每一个 OFI 相加即可。

以卖方为例， e_n 定义的含义是：在时刻 n ，如果卖一价上升了，即 $p_n^s > p_{n-1}^s$ ，则说明在这一段时间内，要么有市场参与者进行了卖出委托的撤单，此时卖方意愿变弱，要么有市场参与者进行了买入的市价委托，此时买方意愿变强，无论哪一种情况，至少说明买方强于卖方，此时对 OFI 产生正影响，影响的幅度为卖方的减少量，即 q_{n-1}^s ；如果卖一价保持不变，即 $p_n^s = p_{n-1}^s$ ，此时无法直接判断买卖

双方的意愿，因此通过 $q_{n-1}^s - q_n^s$ 的差值来判断买卖双方的强弱；如果卖一价下降了，即 $p_n^s < p_{n-1}^s$ ，则说明在这一段时间内，要么有市场参与者进行了卖出的市价委托，此时卖方意愿变强，要么有市场参与者进行了买入委托的撤单，此时买方意愿变弱，无论哪一种情况，至少说明买方弱于卖方，此时对 OFI 产生负影响，影响的幅度为卖方的增加量，即 q_n^s 。

二、订单流不平衡指标的统计检验

Cont 在 Cont (2014) 中就已经利用标普 500 数据验证了 OFI 在统计上是显著有效的同期指标，使用到的模型如下：

$$\Delta P_{k,i} = \alpha_i + \beta_i * OFI_{k,i} + \varepsilon_{k,i} \quad (3.10)$$

其中， k 代表时间， i 代表不同的股票标的， $\Delta P_k = P_k - P_{k-1}$ 代表股票中间价的变化值， $P_k = \frac{p_k^s + p_k^b}{2}$ ， $\varepsilon_{k,i}$ 为误差项。

或许从式 (3.7) 中就能发现一些端倪。引起中间价变化的情况主要有四种：1、买一价上升+卖一价上升，此时 OFI 为 2 个正数相加， OFI 大概率为较大的正数；2、买一价下降+卖一价下降，此时 OFI 为 2 个负数相加， OFI 大概率为较大的负数；3、买一价上升+卖一价下降，此时 OFI 为负数与正数相加， OFI 可能在 0 附近；4、买一价下降+卖一价上升，这与 3 类似。很直观的能感受到：当 OFI 的数值较大时，中间价也会有明显的变动；而当 OFI 的数值靠近 0 时，中间价的变化也会偏低。因此， OFI 确实是一个非常优秀的当期指标，但往往我们需要的并不是一个同期指标，因为这类指标对投资并没有指示性，我们需要的是有预见性的指标。因此，接下来针对 OFI 是否对于 A 股具有预见性进行验证，使用到的模型如下：

$$\Delta P_{k,i} = \alpha_i + \beta_i * OFI_{k-1,i} + \varepsilon_{k,i} \quad (3.11)$$

表 2.2 展示了在显著性水平 α 设置为 1%的前提下模型回归结果的描述性统

计量。具体结果如下：

表 2.2 模型效果 ($\alpha = 1\%$)

	模型 P 值显著比例	变量 P 值显著比例	平均 R^2
大盘股	90.80%	90.80%	0.79%
中盘股	72.27%	72.27%	0.51%
小盘股	60.80%	60.80%	0.39%

第一列为模型 P 值显著比例，该值衡量了在样本期内模型显著的比例；第二列为变量 P 值显著比例，该值衡量了在样本期内变量显著的比例；第三列为平均 R^2 ，该值衡量了模型对于样本的解释力度。从表中数据可以看出，模型对于高市值的股票有更强的解释性，这是显著的。但从结论上来说， OFI 并未达到预期，甚至是远远不及预期，它并不是一个很好的预测指标。首先， R^2 是模型对于样本解释力度最直观的体现，但很可惜的是，平均 R^2 表现最优异的大盘股也只有 0.79%。再从模型显著性角度来看，由于是单变量，变量的显著性与模型的显著性是一致的。虽然大盘股的模型有 90%都是显著的，看似很美好，但对于一个高频指标来说，效果依然是欠佳的。指标的最终目的是作用于投资，而高频最大的特点在于高换手，这会引入高成本，如果 10 天里就有 1 天基本没有预测能力，就意味着这一天会亏钱亏成本，那这样的投资是很难获利的。

这可能是因为美股与 A 股的差异性所导致的原因。由于美股有做市商，因此美股在大多数时间内是处于订单流均衡状态，即 OFI 是接近于 0 的，此时的 OFI 可以更好地识别出价格的大幅度变化，而忽略掉小幅度横盘。这是因为当价格变化缓慢时，做市商会通过打对手单来平衡 OFI ，只有出现快速大幅度的变动时， OFI 才会出现较大的变动。但 A 股目前并没有成熟的做市商，因此 A 股长期处于订单流失衡状态，这也导致了 OFI 的解释性较低。

三、订单流不平衡指标的缺陷与改进优化

正如上文所提到的, *OFI* 的有效性与市场的特点是直接相关联的, 直接将 *OFI* 用于 A 股的效果不佳即是因为 *OFI* 的定义对于 A 股而言存在着一定的不合理性。以表 2.3 为例说明 *OFI* 存在的不合理性。表 2.3 展示了宁德时代在 2020.07.01 的买方 level-1 数据, 由于数据较长, 这里只展示出可以说明问题的数据, 具体数据如下:

表 2.3 宁德时代(300750)在 2020.07.01 的限价买入数据

	买价 1	买量 1	买价 4	买量 4	买价 5	买量 5	买价 6	买量 6
09:30:03	178.8	6607	178.7	4400	178.69	100	178.68	7400
09:30:06	178.68	2007	178.65	300	178.64	100	178.63	100

如表 2.3 所展示的, 订单簿数据即是由多个价格与价格对应的量构成的。正如本节中所介绍的, *OFI* 有一个非常重要的前提假设, 即是所有的市价单、限价单、撤单都发生在最优报价上, 即买一卖一上。但从表 2.3 所截取的数据来看, 这个假设可能存在着较大的歧义。根据 *OFI* 的定义, *OFI* 只考虑了买一价卖一价变化产生的贡献, 即这 3s 内买方对 *OFI* 造成的贡献为 -6607; 但从数据中可以发现, 仅仅在这 3s 内, 价格发生了惊人的变化, 09:30:06 时刻的买一价与 09:30:03 时刻的买一价之间的差距较大, 且新的买一价仅为上一时刻的买六价, 这就意味着在这 3s 内出现了大量的市价卖出单或限价买入单被大量撤单, 从而将价格快速的打压了下去, 但根据 *OFI* 目前的定义是无法识别这种大幅下挫的情况的。根据 *OFI* 最初始的定义 (*OFI* 是 3s 内全部的限价单、市价单、撤单的总和), *OFI* 应该变化的幅度应该至少是 -16600。这样的差距是显著的, 对于 A 股中低市值的不活跃股而言, 这个情况可能并不常见, 但对于 A 股中高市值的活跃股而言, 这是一种十分常见的情况, 而往往我们更青睐的正是这些高市值的活跃股, 因此需要

将该问题在最大程度上予以处理。

本文根据 OFI 的构建思路，在剔除了所有的市价单、限价单、撤单都发生在最优报价上的假设下，提出了新的 OFI （下文的 OFI 都是指本节中定义的新 OFI ）的定义，定义如下：

$$e_n^i = \Delta W^i(t_n) - \Delta V^i(t_n) \quad (3.12)$$

$$\Delta W^i(t_n) = \begin{cases} q_{n,1}^b + \dots + q_{n,k}^b & \text{if } p_{n,1}^b > p_{n-1,10}^b \\ q_{n,1}^b + \dots + q_{n,k}^b - q_{n-1,1}^b & \text{if } p_{n,1}^b > p_{n-1,1}^b \text{ and } p_{n,1}^b \leq p_{n-1,k}^b, 1 < k \leq 10 \\ q_{n,1}^b - q_{n-1,1}^b & \text{if } p_{n,1}^b = p_{n-1,1}^b \\ q_{n,1}^b - (q_{n-1,1}^b + \dots + q_{n-1,k}^b) & \text{if } p_{n,1}^b < p_{n-1,1}^b \text{ and } p_{n,1}^b \geq p_{n-1,k}^b, 1 < k \leq 10 \\ (-1) * (q_{n-1,1}^b + \dots + q_{n-1,k}^b) & \text{if } p_{n,1}^b < p_{n-1,10}^b \end{cases} \quad (3.13)$$

$$\Delta V^i(t_n) = \begin{cases} q_{n-1,1}^s + \dots + q_{n-1,k}^s & \text{if } p_{n,1}^s > p_{n-1,10}^s \\ q_{n-1,1}^s + \dots + q_{n-1,k}^s - q_{n,1}^s & \text{if } p_{n,1}^s > p_{n-1,1}^s \text{ and } p_{n,1}^s \leq p_{n-1,k}^s, 1 < k \leq 10 \\ q_{n,1}^s - q_{n-1,1}^s & \text{if } p_{n,1}^s = p_{n-1,1}^s \\ q_{n-1,1}^s - (q_{n,1}^s + \dots + q_{n,k}^s) & \text{if } p_{n,1}^s < p_{n-1,1}^s \text{ and } p_{n,1}^s \geq p_{n-1,k}^s, 1 < k \leq 10 \\ (-1) * (q_{n,1}^s + \dots + q_{n,k}^s) & \text{if } p_{n,1}^s < p_{n-1,10}^s \end{cases} \quad (3.14)$$

$$OFI_k^i = \sum_{n \in (t_{k-1}, t_k]} e_n^i \quad (3.15)$$

其中， n 表示时刻， i 表示股票， $p_{n,k}^b$ 表示时刻 n 第 k 档的买入价， $q_{n,k}^b$ 表示时刻 n 第 k 档的买入量， $p_{n,k}^s$ 表示时刻 n 第 k 档的卖出价， $q_{n,k}^s$ 表示时刻 n 第 k 档的卖出量， $\Delta W^i(t_n)$ 代表买方订单流的净变化， $\Delta V^i(t_n)$ 代表卖方订单流的净变化， e_n^i 表示时刻 n 的订单流净变化， OFI_k^i 表示股票 i 在时间区间 $(t_{k-1}, t_k]$ 内的净变化。

鉴于逐笔数据的复杂性以及难获取性，新 OFI 指标依然以订单簿数据作为构建基础。新 OFI 指标沿用了原 OFI 的构建思路，通过买卖双方订单流的变化来衡量买卖双方的强弱。以买方为例，新 OFI 的定义逻辑是：买一价在下一时刻可能发生的情况只有三种：价格不变、价格上涨、价格下降。价格不变时， OFI 即是

买一价上量的变化值；股价上涨时，这意味着买方强势，因此对 OFI 的贡献为正，贡献程度的计算不同于原 OFI ，这里通过将超过旧买一价的新买量以及旧买一价上的买量变化值求和作为 OFI 的贡献程度，这是因为，只要超过旧买一价的部分必然是在新的交易时段发出的委托，展现出了买方希望尽快买入的强劲势头；与之类似，股价下降时，此时意味着卖方强势，因此对 OFI 的贡献为负，贡献程度由超过新买一价的旧买量以及新买一价上的买量变化值求和决定。

新 OFI 主要考虑的依然是最优价格的变化所造成的影响，只是不单单只考虑最优价格之间的变化。没有考虑整个订单簿变化的原因有 2 点：首先，正如 Cont (2014) 所提到的，绝大多数的交易依然是在一档价格、二档价格处完成的，因此其他价格上的变化幅度并不大。其次， OFI 计算的是当前时刻对当前价格产生影响的净规模与方向，虽然限价单和撤单是会对价格变动有影响，但对价格产生根本影响的依然还是市价单。根据成交机制，市价单必然优先在最优价格上进行成交，因此只有其他价格成为最优价格，在这个价格上的限价单以及撤单才会对价格产生根本影响。但值得注意的是，这并不意味着其他价格的变化对价格预测没有意义，反而可能是提升预测能力重要的一环。虽然像之前所说，其他价格上的变化不是当期价格变化的造成原因，但它会对未来的价格产生影响，因此将其他价格上的变化纳入考虑或许是可以提高预测准确率的一环，也是未来有待研究的一环。

接下来将对新 OFI 的效果进行检验，验证这样的改进是否有效。表 2.4 展示了用新 OFI 指标回归结果的描述性统计量：

表 2.4 新 OFI 模型效果 ($\alpha = 1\%$)

	模型 P 值显著比例	变量 P 值显著比例	平均 R^2
大盘股	98.95%	98.95%	2.58%

中盘股	92.80%	92.80%	1.43%
小盘股	83.42%	83.42%	1.01%

结合表 2.2 和表 2.4，首先可以得到一个结论，*OFI*指标对高市值标的有更好的解释性，无论是新*OFI*还是原*OFI*。这是可以理解的，表 2.1 就展示了高市值标的有更强活跃性的特点，而*OFI*即是衡量市场多空双方强弱的指标，如果一个标的基本没什么人气，那么它的*OFI*也就会基本靠近于 0，那么它回归的结果也就自然而然表现不佳了。更重要的是，从表 2.4 可以明显发现，新*OFI*在表现上显著优于原*OFI*。首先，在解释力度上，新*OFI*的 R^2 达到了原*OFI*的 3 倍之多，说明了新*OFI*在预测未来短期价格变动上有更好的效果，但平均 R^2 也只有 3%不到的水平，对于这个问题将在下一部分进行解释。对于模型的显著性，新*OFI*使得各市值的股票的显著性都获得了提升，且提升力度是显著的。尤其对于大盘股，模型的解释力度平均可以达到 99%，这就意味着新*OFI*是显著有效的。

四、*OFI*的回归优化

表 2.2 和表 2.4 分别对原*OFI*以及新*OFI*进行了统计上的检验，虽然通过对定义进行了优化，模型的显著性比例得到了很大的提升，但是依然可以发现模型的 R^2 并不高，最好的表现也就才 2.58%，导致低 R^2 的原因有 2 点：

首先，过多的 0 值会弱化 R^2 的表现。例如：如果一组观测值都是 0，那么无论用什么指标，结果永远都只有 1 个——不显著。而这个问题在研究高频时异常常见，再次回顾表 2.1，表 2.1 的日平均买(卖)一价更新次数就代表着股票在一个交易日内价格发生变化的次数。如果假设一天内不涨停不跌停，那么会有大约 4800 个观测值，但从表 2.1 可以看到，表现最好的大盘股也只有大约一半的时候价格是变动的，50%左右的观测值其实都是 0，更不必说中盘股和小盘股了。这也是为什么市值大的标的的表现越好的原因。

其次,频率过高会弱化 R^2 的表现。股价在每一笔快照数据发生变化的概率就已经不高了,更不用说是在一笔快照数据获得的收益率了。因为即使股价上涨了、即使有很明显的上涨态势,但是 3s 的时间也很难反应完全。而往往 OFI 的高低在一定程度上是可以反应股票涨跌态势的,往往 OFI 的值越大,越预示着股价会发生较大较持续的变化,这不是在一个快照数据中就体现的出来的。表 2.5 中展示了这个问题。

表 2.5 宁德时代(300750)在 2020.07.01 的 OFI 表现

	新 OFI	收益率变动
09:32:12	15940	0.0530%
09:32:15	22600	0.0027%
09:32:18	20420	0.0892%
09:32:21	20780	0.0334%
09:32:24	22620	0.0279%
09:32:27	9200	-0.0835%

从表 2.5 展示的数据可以发现,在这 15s 内,股票的 OFI 是较高的,就如前文分析的那样,高 OFI 预示着股票有较强的上升态势,收益率也确实表现出了这趋势。但明显可以发现,收益率的变动幅度非常有限,基本都在 1%以内,而 OFI 的波动较大的,这也直接导致了回归的 R^2 表现不佳。

因此,这里对回归的数据进行了一定的优化,不再直接使用计算后的数据,而是进一步对数据进行加工处理。处理策略如下:如表 2.5 所展示的,文中认为 $OFI > 0$ 意味着收益率会收涨,但此时并不以下一个 tick 的收益率作为观测值,而是将连续的正 OFI 聚合起来,将 $OFI > 0$ 的区间内的总收益率作为观测值与 OFI 进行回归。简单来说,以表 2.5 为例,就应该将最开始为正的 OFI 值

(09:32:12, 15940) 和 *OFI* 持续为正的区间收益率 ([09:32:12, 09:32:24], 0.206%) 作为一组回归的对应值。

根据上述方案, 进一步对 *OFI* 的回归进行优化, 表 2.6 展示了优化了观测值后的回归结果的描述性统计量:

表 2.6 优化回归后模型表现 ($\alpha = 1\%$)

	模型 P 值显著比例	变量 P 值显著比例	平均 R^2
大盘股	99.66%	99.66%	11.89%
中盘股	97.68%	97.68%	8.96%
小盘股	96.10%	96.10%	7.42%

单从回归后的结果可以看出, 将观测值优化后模型的效果是显著的。模型的解释力度从多方面都呈现出了显著的改善。大盘股模型的平均 R^2 超过了 10%, 一定程度上说明模型有较强的解释性。大盘股模型的显著比例已经非常逼近 100%, 并且中盘股和小盘股的表现得到了非常大的提升, 它们的模型显著比例都超过了 95%。这就是因为中小盘股的价格变动概率较低, 仅对变化的情况进行考虑的话, 即使针对中小盘股, 新 *OFI* 也展示出了明显的显著有效性。

五、*OFI* 的准确率

衡量一个指标有效性的标准除了统计学上的显著性检验外, 便是指标的准确率测试。表 2.7 展示了改进前后指标的平均胜率:

表 2.7 *OFI* 指标胜率

	原 <i>OFI</i> 的平均准确率	新 <i>OFI</i> 的平均准确率
大盘股	53.11%	60.60%
中盘股	53.33%	57.06%

小盘股	52.41%	56.50%
-----	--------	--------

从表 2.7 可以看出，原 OFI 指标的正确率基本维持在 53%左右，而新 OFI 指标的预测准确率都超过了 55%，甚至大盘股的准确率来到了 60%。一定程度上可以说明， OFI 的改进有切实有效的。

六、小结

本节给出了 OFI 指标的初始定义，并结合 A 股的实际情况挖掘出了 OFI 存在的问题，并针对问题给出了 OFI 的优化定义，并从统计学的角度验证了新 OFI 在预测未来股价变动这一问题上呈现出了更优秀的表现。同时，本节也解释了为什么回归的 R^2 一直处于较低的水平，并提出了改进方案。结果也证实了改进确实提高了模型的 R^2 以及模型的解释力。

第三节 订单簿斜率指标

上节系统性地讨论了 OFI 指标，从投资者的角度来看， OFI 是站在所有市场参与者的角度来分析市场多空强弱的；从订单簿数据的角度上来看， OFI 同时运用了多空双方的数据，是一个很明显的站在整体角度看问题的指标。从上述研究中，可能会联想到有没有什么指标是针对个体投资者行为呢、有没有什么指标可以将订单簿数据进行拆分分析呢、有没有什么指标可以从个体的角度看问题呢。订单簿斜率指标就是这样一个从更个体的角度看问题的指标，因此为了完善对于订单簿的研究，本节将简单针对订单簿斜率指标进行探讨。

一、订单簿斜率指标的定义与缺陷

正如文献综述中所提到的，Næs (2006) 最先设计了订单簿斜率，其定义如下：

$$SSlope = \frac{1}{N_A} \left\{ \frac{v_1^A}{\left| \frac{P_1^A}{P_0} - 1 \right|} + \sum_{i=1}^{N_A} \frac{\frac{v_{i+1}^A}{v_i^A} - 1}{\left| \frac{P_{i+1}^A}{P_i^A} - 1 \right|} \right\} \quad (3.16)$$

$$BSlope = \frac{1}{N_A} \left\{ \frac{v_1^B}{\left| \frac{P_1^B}{P_0} - 1 \right|} + \sum_{i=1}^{N_A} \frac{\frac{v_{i+1}^B}{v_i^B} - 1}{\left| \frac{P_{i+1}^B}{P_i^B} - 1 \right|} \right\} \quad (3.17)$$

其中, N_A 表示订单簿的数据深度, P_i^A 表示第 i 档的卖出价, v_i^A 表示第 i 档的累计卖出量 (即 $\leq i$ 档的其他档的卖出量的累计值), P_i^B 表示第 i 档的买入价, v_i^B 表示第 i 档的累计买入量, $P_0 = \frac{P_1^A + P_1^B}{2}$ 表示中间价。

式 (3.16) 与 (3.17) 的第一项测量了最优价格处的量对于斜率的影响, 后一项测量了其他价格处的量对于斜率的影响。指标的构建逻辑在于: 对于一个投资者而言, 当他发现订单簿的卖方斜率很大, 即卖方的量很厚很多时, 他很可能会意识到价格很难在短期内上升, 这时他的买入意愿就会下降, 甚至可能引起跟风卖出, 这在一定程度上就会引起价格向下走。也就是说, 卖方斜率越大, 未来价格更可能下降; 反之, 买方斜率越大, 未来价格更可能上升。不同价格对于投资行为的影响力度也是不同的, 很明显, 最优价格处的量对于当前时刻投资者行为的影响力最强烈, 并且这种影响力会随着价格对于最优价格的偏离而逐渐减小。指标中也体现出了这一特点。指标在构建时并不是直接用第 i 档的量作为分子, 而是将它除以了 i 档以前量的累计值, 这就意味着, 随着价格偏离最优价格, 该价格处的量对于订单簿斜率 (投资者行为) 的影响力也会逐渐减弱, 这是符合逻辑的。

虽然该指标的构建逻辑很清晰, 但不幸的是, 订单簿斜率指标的定义存在着一定的不合理性:

首先，作者在原文中也提到了指标存在的一个最大的问题在于式(3.16)与(3.17)的第一项与第二项的测量单位是不同的。第一项直接将需求量作为分子使用，而第二项用的则是需求量的增长率。产生这个问题的根本原因在于第一项无法找到参考值计算增长率。对于这个问题，不乏有学者提出改进的方案，如：作者在原文中提到了通过对量取对数缩小它的范围的方法，但实际这样做并没有解决问题本身，并且取对数从根本上影响了数据本身的属性；Kim(2004)对于该类问题提出的方案是标准化法：

$$(V_i^B)' = \frac{V_i^B}{\sum_{j=1}^{N_A} V_j^B} \quad (3.18)$$

这样的标准化法确实可以解决掉测量单位不同的问题，但值得注意的是，公式中的第二项的增长率是不会因为数字的缩放而改变的，但第一项却会直接受到数字缩放的影响，这意味着订单簿斜率指标会对选取的标准化方法很敏感。更值得注意的是，这样的标准化法对于预测未来价格变动是不利的。举个例子来说，如果买量是卖量的2倍，即每一档的买量都是相对应卖量的2倍，很明显，买方强于卖方，但因为标准化的原因，这样的特点无法被捕捉到。因此，如何优化这一问题将成为一个难点。

其次，Næs(2006)中发现买方斜率的影响力强于卖方斜率的影响力，这与过去的研究相仿，由于缺乏做空机制，或许买方确实要比卖方更重要，但光从该指标的回归表现很难给出这样的结论。观察指标的定义可以发现，指标的分母是用价格的偏离率来衡量的，这时候问题就出现了。卖价永远高于买家，因此即使在分子完全一致的前提下，买方的斜率也会因为价格偏低而被低估。

为了克服上述提到的指标的不合理性，本文基于Næs(2006)的构建思路，提出了一种新的订单簿斜率的定义，用于为后续研究时提供一个可行的思路，具体定义如下：

$$SSlope = \frac{1}{N_A} \left\{ \frac{v_1^A}{\left| \frac{P_1^A}{P_0^A} - 1 \right|} + \sum_{i=2}^{N_A} \frac{v_i^A - 1}{\left| \frac{P_i^A}{P_0^A} - 1 \right|} \right\} \quad (3.19)$$

$$BSlope = \frac{1}{N_A} \left\{ \frac{v_1^B}{\left| \frac{P_1^B}{P_0^B} - 1 \right|} + \sum_{i=2}^{N_A} \frac{v_i^B - 1}{\left| \frac{P_i^B}{P_0^B} - 1 \right|} \right\} \quad (3.20)$$

为了克服原有的问题，本文采取的手段是直接将问题进行剔除。对于分子的选取上，统一为每一档价格上的需求量，这样做消除了指标对于标准化过于敏感的问题以及测量单位不统一的问题，但这样做会忽略价格越偏离中间价，需求量的影响力越小的情况。在这里，通过对于分母的改变来考虑这个问题。价格越偏离中间价，分母就会越大，所产生的影响也会越小。

二、新订单簿斜率指标的统计检验

在验证新订单簿斜率指标的统计表现时使用到的模型如下：

$$\Delta P_{k,i} = \alpha_i + \beta_{i,1} * SSlope_{k,i} + \beta_{i,2} * BSlope_{k,i} + \varepsilon_{k,i} \quad (3.21)$$

其中，k 代表时间，i 代表不同的股票标的， $\Delta P_k = P_k - P_{k-1}$ 代表股票中间价的变化值， $P_k = \frac{p_k^s + p_k^b}{2}$ ， $\varepsilon_{k,i}$ 为误差项。

表 2.8 给出了指标回归结果的描述性统计量：

表 2.8 订单簿斜率指标的模型表现 ($\alpha = 1\%$)

	模型 P 值显著 比例	SSLOPE 变量 P 值显著比例	BSLOPE 变量 P 值显著比例	平均 R^2
大盘股	93.48%	92.87%	91.52%	6.56%
中盘股	84.34%	83.57%	81.59%	9.54%

小盘股	78.45%	77.89%	77.43%	9.45%
-----	--------	--------	--------	-------

从表 2.8 的结果可以发现, 订单簿斜率指标依旧对高市值标的的效果更显著。但与过去的结论有些差异的点可能在于卖方斜率的显著性要强于买方。值得注意的是, 模型的 R^2 对于中小市值的表现强于大市值的股票, 之所以产生这个情况是因为中小市值的标的虽然很容易被拉涨或拉跌, 但更多的时候其实是不活跃的, 就如第一节中所分析的那样, 中小市值的股票它日内的变化并不频繁, 这在回归时可能是有利的。

第四节 OFI与订单簿斜率指标的关系

正如第三节开篇说提到的, *OFI*与订单簿斜率指标代表了订单簿数据的 2 个不同的研究角度。它们之间既相似又不同。相似之处在于它们所用的数据都是订单簿数据, 且构造逻辑都在于思考订单对于投资者行为的影响, 区别之处在于 *OFI*是从整体(市场)的角度去分析问题, 而订单簿斜率指标关注点更加集中在个体(投资者), 它们互相弥补, 完善了订单簿的研究。

正因为 2 个角度的出发点不同, 它们所预测的方向也大相径庭。*OFI*对于强趋势有更强的预测性。这是因为股价的突然变化基本都是有预见性的, 且这种变化通常带有短暂的持续性, 而*OFI*恰好可以识别出这样的特点。相比之下, 订单簿斜率指标更适合用于多空双方均衡的状态下, 举个例子来说: 当买方量很大时, 订单簿斜率指标给出的建议必然会是价格会上升。但是实际上, 如果价格是突然降下来的, 那么极有可能出现的情况是卖方来不及下限价单、买方来不及撤单, 导致了买方看起来强于卖方, 但实际却恰恰相反, 而这样的错误预判, 会带来严重的后果。

或许 2 个指标的结合能获得更好的效果。但从实用性的角度来看, 对于一个

高频策略，成本是不可逃避的核心问题，而订单簿斜率有个很尴尬的问题。它适用于多空双方平衡的时期，但这样的时期价格的变化会很小，甚至可能难以覆盖成本，而在趋势阶段又可能会完全做出反向判断。相比之下，*OFI*由于可以更好的识别出每天的小趋势，因此它的获利空间远远大于订单簿斜率指标，这也意味着*OFI*比订单簿斜率指标更适合运用于实证。

第五节 本章小结

本章以订单簿数据为基础，从整体与个体的角度出发、市场与投资者的角度出发，研究了订单簿的 2 个不同角度。通过结合实际数据，发现了 2 个指标各自存在的风险，并结合 A 股实际情况给出了相应的改进建议，同时在理论上与统计上对每个指标进行了分析。

从结论上来说，对于 2 个指标的改进是有效的，改进后的指标在统计上都呈现出了更优秀的表现。同时在本章深入挖掘了 2 个指标的深层逻辑与适用情况，认为*OFI*可能更适用于实证。

第四章 基于OFI的高频策略模型

正如第三章所分析讨论的那样，OFI相比于订单簿斜率指标来说，可能更适用于高频策略模型，因为OFI是一个有助于识别出短期强趋势的指标。而高频交易可以将这些短期的强趋势转变为盈利。本章将从实证的角度出发，研究是否可以利用新OFI构造出一个有效的高频策略模型。

第一节 模型假设

由于高频模型的特殊性导致实战与理论相差较大，因此需要一定的假设予以辅助，本文的假设如下：

- 1、交易对象：以第三章第一节提到的标的选择方法为主，从A股中最高市值的300只股票中最活跃的30只股票；
- 2、交易时间：2020/07/01-2021/06/30；
- 3、交易费用：1‰的单边印花税与1‰的双边佣金；
- 4、仓位假设：模型假设前一天有足够的仓位用于下一天的所有交易；
- 5、交易假设：模型中将以对手价进行交易（即以卖一价买入，买一价卖出），并默认多少量都可以立即成交；
- 6、收益假设：模型假设初始投资额=当天使用股数*初始股价，收益计算时忽略持仓产生的收益或亏损。最终的模型收益率采取股价加权法。
- 7、不考虑任何的延迟情况，模型认为数据的获取、下单都没有延迟，即获得的是最新时刻的数据情况。

前三个假设为常规假设，为了确保模型的泛用性，股票数量由统计检验中的

十只提高到三十只。

第四个假设的目的在于解决仓位管理问题。由于目前 A 股采用的是 T+1 交易制度，这就意味着想要实现股票的 T0 交易，就只有 2 种方案：1、前一天持有足够的底仓。2、可以进行融券交易。但实际上，股票的融券有诸多限制，有些股票是没有券源的，因此假设前一天持有足够的底仓用于下一天的交易。

第五个假设的目的是为了忽略无法成交的问题。高频策略模型的实盘与研究结果之所以会相差甚多的核心原因就在于实盘可能无法抢单成功。即使可以按希望的价格交易，也可能无法将自己想要交易的数量全部交易完。在研究的过程中，暂时先忽略了这一个问题，这可能是另一个领域内的问题了。

第六个假设是为了便于收益率的计算。实际情况是，在一天结束后才能知道当天交易了多少手，初始仓位有剩余是一个相当常见的现象。但如果真要考虑这个问题，那么计算出的收益率很大程度会受到初始仓位的数量的影响，是很难计算的，因此这里先忽略该问题。同时，持有仓位会存在着隔夜的风险，这同样会产生收益，但这并不是模型所带来的收益，并且该收益可以通过利用融券来进行规避，因此在计算收益时忽略掉了该部分的收益。同时，由于模型将对一揽子股票进行交易来降低非系统性风险，因此需要规定组合收益率的计算方式。本文采用股价加权法，主要为了避免高价股主导模型表现的问题出现。股价加权即意味着低价股会配有更高的权重，甚至这个权重是一个非整数。如果权重是非整数，理论上可以通过乘以最大公因数使得所有股票交易的数量为整数。并且根据第二条假设，认为所有量都能按对手价立即成交。

第二节 模型介绍

一、OFI高频模型介绍

模型的基本框架如下：

图 3.1 高频策略模型

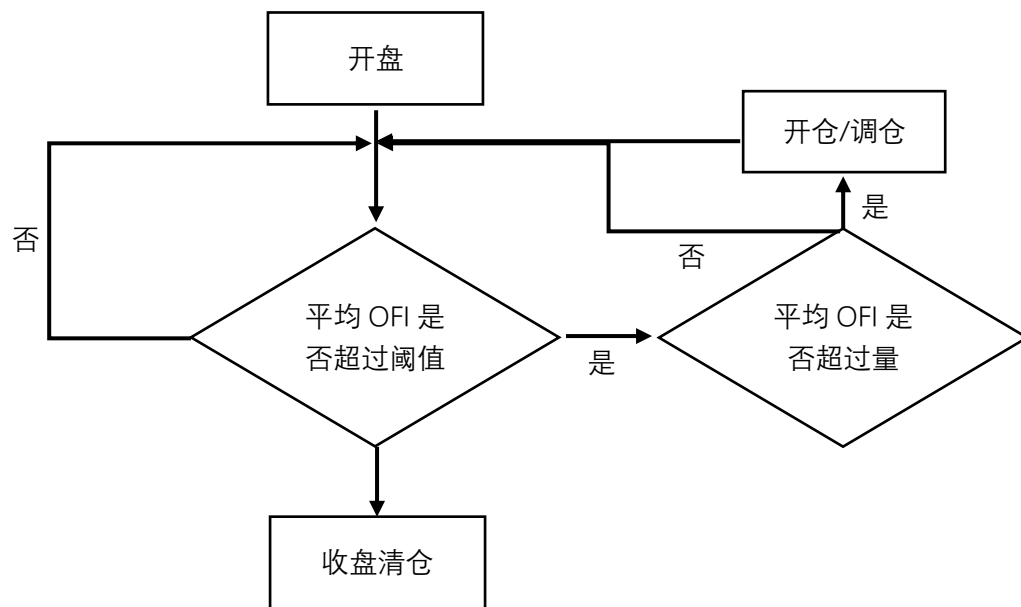


图 3.1 展示了高频策略模型的基本框架，模型基本上只有 2 个关于 *OFI* 的判断：*OFI* 的数值是否超过阈值以及 *OFI* 的数值是否超过对手方的量。接下来将对模型进行更细腻の説明。

首先，需要说明的是，模型的频率设定为 1tick，即每笔快照发出后就计算一次信号，基本 3s 一次。

值得注意的是，模型中并未直接使用 *OFI* 指标，而是使用了平均 *OFI*，即用指数移动平均法对 *OFI* 进行平滑后的结果。之所以这样处理是因为一个逻辑：我们最终的目标是识别长趋势，但 3s 内的走势的方向是难以预测的，出现突然的由正转负也是极有可能的，因此采用多笔数据加权的方法来最大程度减少这一问题。之所以选用指数移动平均，而不是等权，是因为当前时刻的 *OFI* 对于市场的影响是最大的，而这一特点在等权中是无法得以体现的。第一个判断标准是平滑后 *OFI* 的数值是否超过阈值，之所以需要这样的判断标准的原因有个：1、高频

的成本过高(对手价+1%的印花税),如果每出一次信号就交易的话,很容易出现入不敷出的现象,因此需要压缩信号,提取最高效的信号点。2、正如第三章分析的那样,OFI较高暗示着股价变动可持续,当OFI在0附近时,股价更可能处于横盘时期,因此应该选择高OFI作为买入信号,但这又会出现一个问题:阈值如何选取?本文利用过去5日的二十分位数的平均值作为阈值使用,即选取前5%的值作为交易信号。当然了,这样的选取标准不一定是最好的,但在一定程度上是合逻辑的。

如果OFI的数值超过了阈值,则会进入下一个判断:OFI的数值是否超过最优报价上的量。这一判断基于的逻辑是:引起价格上升的根本原因是需求超过了供给,OFI衡量的即是需求,最优报价的量即是供给。因此只有当OFI超过了最优报价的量,价格才会上升。

如果OFI同时满足这2个标准,则认为股价将发生持续性变化,这时无仓位则开仓,有仓位则不动,直到出现下一个信号或者收盘前再进行平仓。

大致的模型逻辑如上述所说,模型主要有3个超参数,分别是:指数平滑时选取的数据大小、阈值、股票标的池大小。这里选取的参数值为:指数平滑的数据大小设置为5,即过去15s的结果;阈值选了过去5日的较大的前5%的数据的平均值;股票标的池大小选取为30只。

对于这3个超参数,模型对于指数平滑的数据大小并不敏感。原因在于:对于指数平滑而言,过去值的影响力度会越来越小,最终趋近于0,因此该值对于结果的影响并不大。但模型对于阈值的选取极其敏感,这是可以预见的。正如刚刚所说的,高频的成本极高,手续费就达到了1%,外加对手价之间的成本。信号频率的提升是可以提升预测准确率的,但代价是极其高昂的成本,如何权衡成本与准确率确实是一大难点,下文会专门对这一问题进行讨论。

二、VPIN策略模型介绍

本文所采取的基准对照模型为VPIN策略模型。VPIN将每天的交易量根据确定的值 V 分成 N 个篮子，再计算每个篮子中的订单流不平衡，最终汇聚而成的指标既是VPIN。

选择VPIN策略模型的原因如下：1、正如第一章所提到的，VPIN与OFI的底层研究角度是一致的，都是基于股价的订单流不平衡这一概念提出的。2、虽然2类指标的底层研究角度相同，但它们的研究频率有极其大的差异：VPIN需要一定的成交量来分析市场种知情交易者的占比，因此频率相对较低，在模型中采用了过去一天的滚动值来计算每一个VPIN；相比之下OFI的频率非常高，模型中采用了过去15s的滚动值来计算每一个OFI。因此这样的2类模型也可以直观地展示频率高低对于模型准确性的意义，VPIN的具体计算公式如下：

$$VPIN = \frac{\sum |V^B - V^S|}{N * V} \quad (4.1)$$

$$V^B = \sum V_i * N \left(\frac{\Delta p}{\sigma_p} \right) \quad (4.2)$$

$$V^S = \sum V_i * \left(1 - N \left(\frac{\Delta p}{\sigma_{\Delta p}} \right) \right) \quad (4.3)$$

其中 N 表示计算VPIN所用的篮子个数， V 表示每一个篮子中的总交易量，由于VPIN要求每个篮子中的总交易量一致，因此 V 是常数。 V^B 表示买方的量， V^S 表示卖方的量， $|V^B - V^S|$ 衡量了订单流的不平衡性， Δp 是每个篮子中股价的变动， $\sigma_{\Delta p}$ 表示了股价变动的标准差， $N(*)$ 表示正态分布。

本文中所运用的VPIN策略模型参数设置如下： V 用过去5天的滚动成交量的均值除以50得到，即将每天的成交量分成50份左右； $\sigma_{\Delta p}$ 利用过去5天的股价变动值计算得到。

模型的基本框架如下：

图 3.2 *VPIN*策略模型

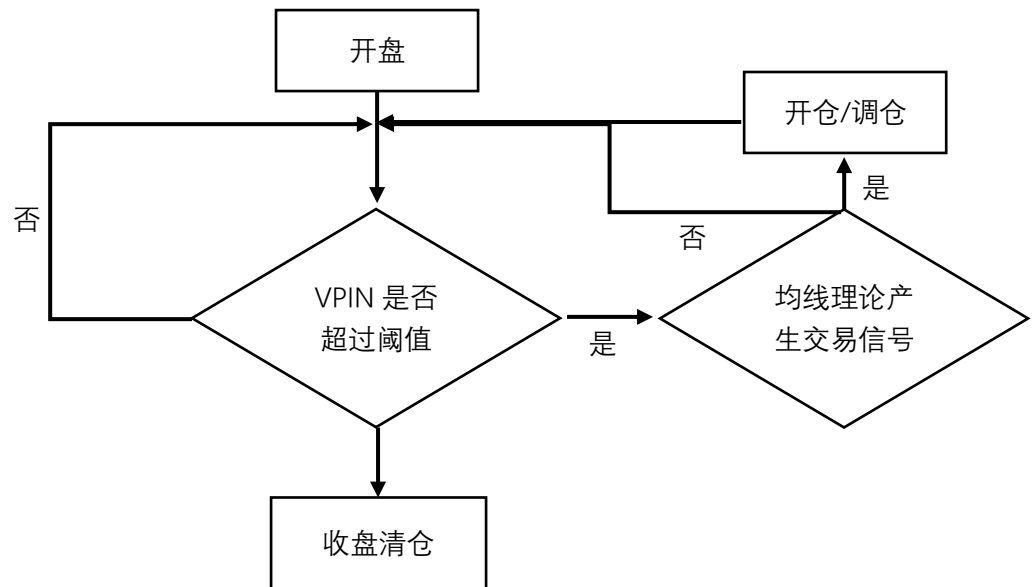


图 3.2 展示了*VPIN*策略模型的主要框架，基本逻辑在于：每次计算完*VPIN*后，对*VPIN*进行判断，如果*VPIN*较高，则意味着市场的知情交易者较少、非知情交易者较多，此时非知情交易者会随着时间的推移而获得更多的信息，从而使股价产生动量效应，此时的*VPIN*也会逐渐减小，当低于阈值时，动量效应随即消失。而对于动量效应的走向利用均线理论进行判断，当过去 5 个时间单位(每个时间单位即是*VPIN*中每个篮子的时间单位)的均价大于过去 20 个时间单位(每个时间单位即是*VPIN*中每个篮子的时间单位)的均价时，则认为后市股价会趋势向上；反之，则认为后市股价会趋势向下。

第三节 模型结果及评估

一、模型评价指标

表 3.3 评价指标

评价指标	计算方式	指标意义
年化收益率	回测期间收益率*一年的交易日数量/回测区间长度	收益指标。年化收益最直观地衡量了一个模型的收益表现
波动率	收益率序列的标准差	风险指标。衡量收益的波动程度，反应模型的稳定性。但由于标准差不考虑方向，因此会产生收益曲线在 0 附近的波动率低的问题。
最大回撤	回测期间产生的最大亏损	风险指标。衡量了任何时间使用模型可能出现的最大损失。弥补了波动率存在的缺陷。
夏普比率	$(\text{年化收益}-\text{无风险利率})/\text{波动率}$	同时考虑风险与收益的指标。衡量了每承担一风险所能获得的收益。
卡玛比率	$(\text{年化收益}-\text{无风险利率})/\text{最大回撤}$	同时考虑风险与收益的指标。衡量了使用模型可能承担的最大损失与模型收益之间的关系。

二、模型结果与评估

图 3.4 与图 3.5 对于新OFI模型、旧OFI模型与VPIN模型的净值走势与最大回撤进行了展示，并在表 3.6 中对模型的总体表现的描述性统计量予以展示，表现如下：

图 3.4 模型净值

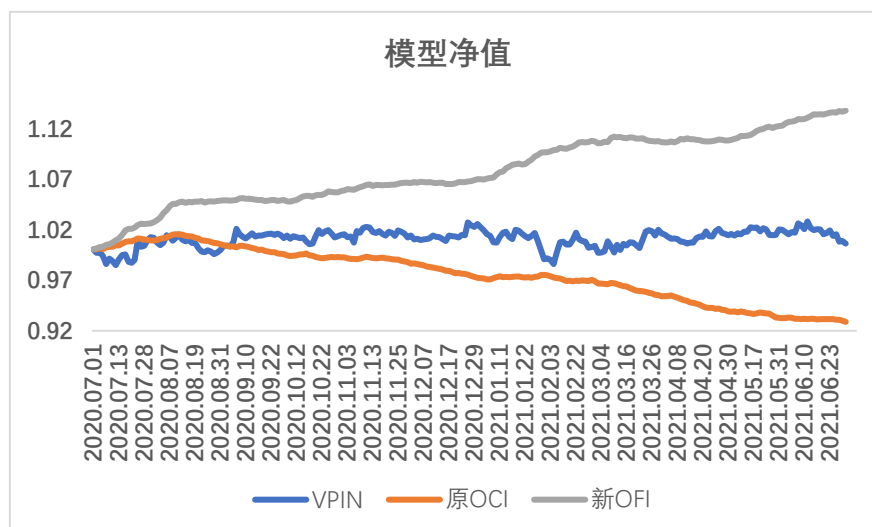


图 3.5 模型最大回撤

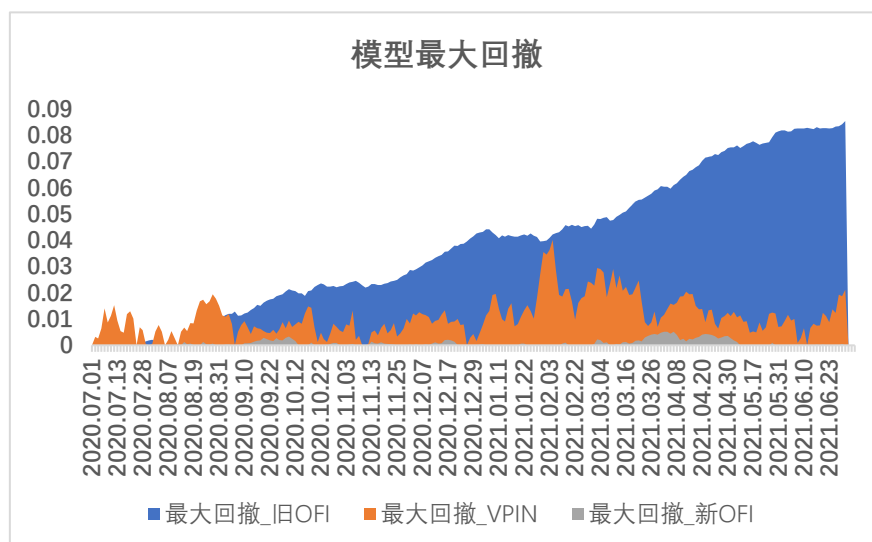


表 3.6 OFI模型的描述性统计量

	年化收益率	波动率	最大回撤	夏普比例	卡玛比例
OFI模型	13.79%	3.38%	0.52%	4.08	26.73
VPIN模型	0.64%	0.84%	4%	0.76	0.16
旧OFI模型	-7.11%	3.08%	8.54%	—	—

从结论上来说，新OFI模型（下文称为OFI模型）的表现是优秀的。相较于以旧OFI构建的高频策略模型而言，OFI模型无论是在收益上还是在最大回撤上都

展现出了显著优秀的表现。*OFI*模型在控制住风险的同时(最大回撤显著优于旧*OFI*模型,波动率虽然不如旧*OFI*模型,但这是因为波动率的计算方式问题而导致的),还将原*OFI*模型的负收益转为了显著的正收益。*OFI*模型的改进是有效的。相较于稍微低频的*VPIN*模型而言,*OFI*模型也在控制住风险的前提下,使得收益获得了显著的提升。从表 3.6 可以看出,虽然*OFI*模型的波动率高于*VPIN*模型,但这是由于波动率存在的计算问题所导致的,在收益率以及最大回撤上,*OFI*模型都展现了压倒性的优势。从图 3.4 和表 3.5 可以看出,*OFI*模型有着极强的鲁棒性,无论是针对市场上升期的 2020.07,还是市场下跌期的 2021.02,模型都展现出了高度稳定的走势;而对于其他的平稳期,模型虽然也展现出了很强的鲁棒性,但可以明显发现,在市场波动较小的时候,模型的收益增速减慢了、风险损失增大了,但整体趋势依然是平稳向上的。因此,模型有很强的鲁棒性。

根据*OFI*指标的构建逻辑可以认识到,模型的最终目的是赚短期趋势的钱,因此股价对上涨还是下跌并不敏感,而是对波动率更敏感,即:高波动会带来高回报,低波动就会带有一定的风险。

模型的结果也证实了该逻辑。模型产生的最久的 2 段回撤都发生于行情震荡的时期,分别在 2020 年 9-10 月以及 2021 年 3-4 月,但虽然模型产生了较久的回撤,但模型的回撤幅度并不大,最大回撤仅 0.52%。之所以模型的回撤期会较长,很可能是因为当时市场波动率较低导致模型盈利较少,因此需要更长的时间来弥补可能存在的亏损。

值得注意的是,在 2021 年 2 月底以及 3 月初时,A 股的市场风格产生了巨大的转变,机构抱团风气减弱,大盘股弱势,市场上高波动标的出现了反转。这或许是导致模型在 2021 年 3-4 月产生较长回撤的原因之一,因为本文选取股票池的方法是按上一个月股票的表现进行选择的,因此模型对于市场风格的切换

存在着一定的延迟性。但值得注意的是，虽然模型在这段时间出现了较长的回撤期，但回撤幅度并不大，仅 0.52%。从这一层面来看，模型虽然无法对于市场风格的切换作出及时有效的反应，但也可以在很大程度上降低同时期产生的风险。

三、模型缺点

OFI模型在收益与回撤上都展示出了优异的表现，但值得注意的是，这样的结果是在理论上的最优结果，实际上很难达到，这是因为模型存在着不可忽视的 3 大问题：

1、模型存在的最大问题在于：OFI模型是一个高频模型，因此对于买卖点是十分敏感的。实战中并无法确保能及时获得股价信息，也无法确保下单能没有延续，反而很大可能会存在着一定的延迟，而这种延迟可能会为策略产生巨大的成本，而在策略回测过程中并未考虑延迟的问题。

2、模型的容量较小。由于模型需要在信号出现时快速进行介入，因此模型的容量必然会小，否则很难立即成交。

3、模型在计算收益时，默认所有仓位都用于操作。但实际情况是，仓位盈余或仓位不足占了绝大多数，这也会产生大量的成本。

因此，OFI模型虽然展示出了优秀的结果，但这还是一个理论上的最优解，想要将模型应用于实战中，依然有很长的路要走。

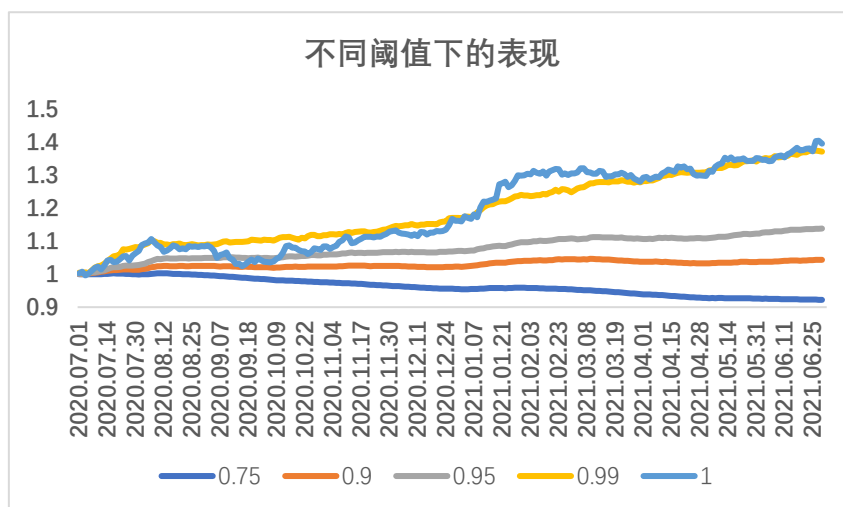
第四节 模型参数敏感性检验

如第二节所分析的那样，模型所需要设置的超参数有 3 个：指数平滑时选取的数据大小、阈值、股票标的池的大小。接下来将对这 3 个超参数分别进行敏感性检验。

一、阈值敏感性检验

阈值可以说是模型中最核心最重要的一个参数了，因为阈值不仅会影响模型的准确性与总收益还会影响模型的总成本，尤其当阈值设置过小时，过高的成本必然会使策略的收益入不敷出。图 3.7 展示了不同阈值下模型的表现：

图 3.7 不同阈值下模型的表现



从结果上来看，阈值越大模型的表现越好。这一点是完全符合逻辑的，模型的目的是识别短期趋势，*OFI*越大暗示着股价出现短期趋势的可能越高、预测的准确率也越高，因此阈值越大可以筛选出更好的买卖时机。

虽然阈值越大模型的表现越好，但也有需要注意的几点：

首先，可以明显看出，当阈值选取过去 5 日的前 25%数据的平均值时，模型的收益是负的，而当参数值增大(选取的数据排名更靠前)时，模型出现了显著的正收益。这就是因为：高阈值能更精准地挑选出好的买卖时机，并在挑出好时机的同时降低交易手术，因此成本也得到了最大限度的控制。

其次，值得注意的是，由于使用的是过去 5 日的平均值，因此即使取 100%，模型的收益也不会为 0，图中即展示了参数设置为 100%时模型的表现。可以看

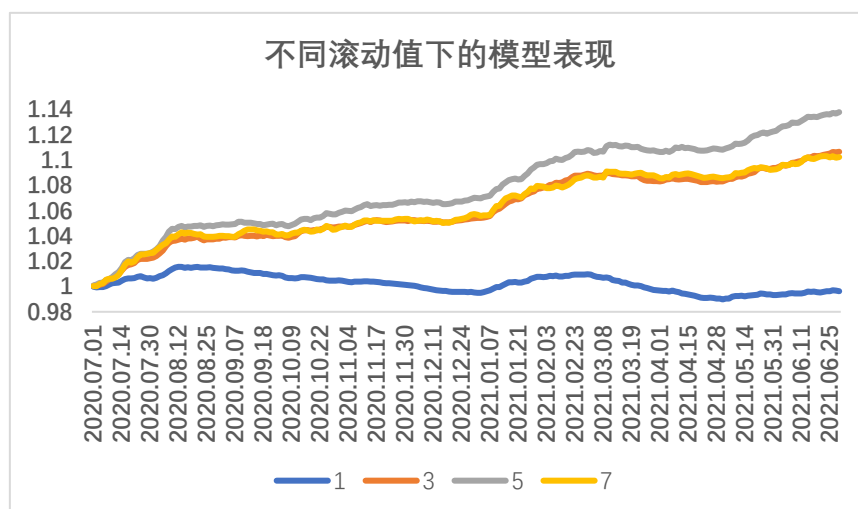
出，在取 100%时模型的波动率较大，但即便如此，它在收益上的表现也优于其他参数表现，是符合阈值越大模型的表现越好的特点的。之所以为出现这样的高波动，就是因为阈值选取的高，导致了交易信号过少，从而一天很容易出现不交易的情况，因此产生了较大的波动。但值得注意的是，高阈值表现好的前提时市场平稳，如果市场出现了巨大的风格切换，使得选择了波动性低的股票，那么即使是高阈值，也会产生亏损，正如 21 年 3-4 月的情况。当然了，可以通过更加频繁的选取股票池来克服风格切换的方法。由于本文通过过去 1 个月的表现进行股票池的选取，因此无法及时识别风格切换的问题。

从图 3.7 可以看出，阈值选取为 0.9 到 1 之间的数都是可行的，数字选取的越大表现越好，但这个数字也是有上限的，当阈值超过一个界限，信号将不再产生。当然了，这个阈值可以通过利用历史数据训练得到。

二、指数平滑数据设置敏感性检验

将OFI指数平滑的目的在于，一次突然的由正转负或由负转正并不一定能说明问题，如果这种负趋势或正趋势是持续的，那才能说明股票将会出现持续的趋势，接下来对指数平滑的数据数量参数进行敏感性检验。图 3.8 展示了不同数量下模型的表现：

图 3.8 不同滚动值下模型的表现



从结果上来看，以过去 5 个 tick 数据进行指数平滑的效果最好。但其实参数设置为 3、5、7 时模型的表现差异并不太大，这是因为模型采取的是指数平滑，因此离的越远，效果越小，最终会趋于稳定。但模型之所以会有差异的原因有 2 点：虽然说影响不大，但过去的数据还是会有影响的，因此模型之前还是存在差异。另外就在于指数平滑的值越大，就意味着出信号的时机越晚，因为它需要更多的时间计算信号。但即便如此，这样的模型差异也是可以接受的。但值得注意的是，当不采取指数平滑时，即参数为 1 时，模型的表现并不理想。原因在于，当不平滑时，信号完全由当前时刻发出，但是在 1 个 tick 内出现值的正负反转是常见的，并不可持续，这会导致模型频繁的进行开平仓，会带来巨大的成本，因此模型的表现并不理想。

从图 3.8 可以看出，模型对于指数平滑参数的设置的敏感性并不高，指数平滑值设置在 3-7，甚至更高都是合理可行的。当然了，这一参数也可以通过历史数据训练得到。

三、股票标的池大小

图 3.9 和图 3.10 展示了模型在不同股票数量下的模型结果：

图 3.9 不同股票数量下模型的表现

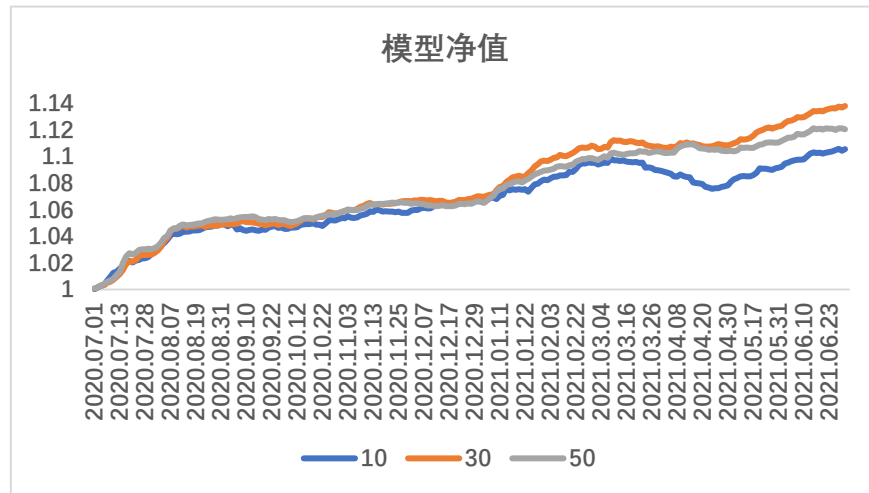
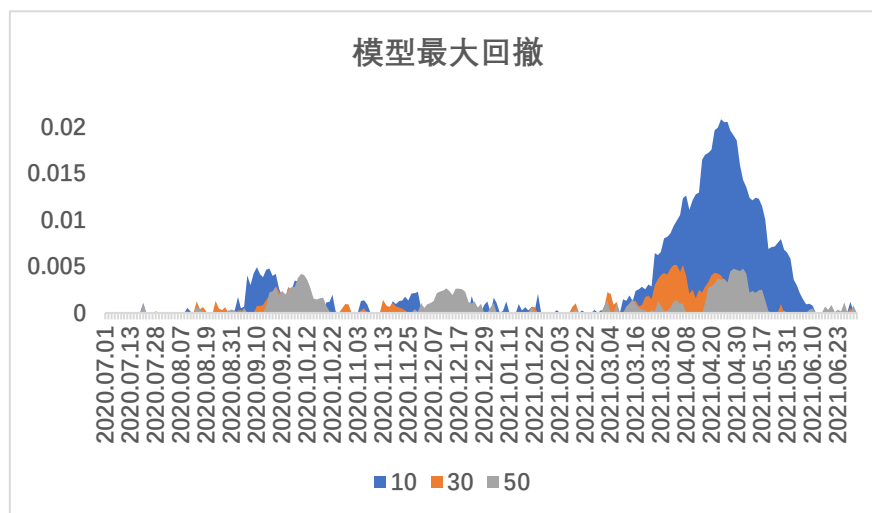


图 3.10 不同股票数量下模型的最大回撤



上述图片展示了以第二章的筛选策略分别选出 10(从 100 只股票中选取)、30(从 300 只股票中选取)、50(从 500 只股票中选取)只最活跃的股票作为标的池的模型结果。从模型净值与模型回撤图可以看出，模型对于不同股票数量的表现是类似的，但也存在着差异。之所以选择多只股票是为了平滑风险，所以扩大标的池有助于平稳收益曲线，正如 10 只标的的曲线与 30 只标的的曲线、50 只标的的曲线，10 只标的的曲线明显有着更高的回撤。但值得注意的是，并不是股票数量越多越好。扩大标的池的目的是分散风险，但当标的池数量达到一定水

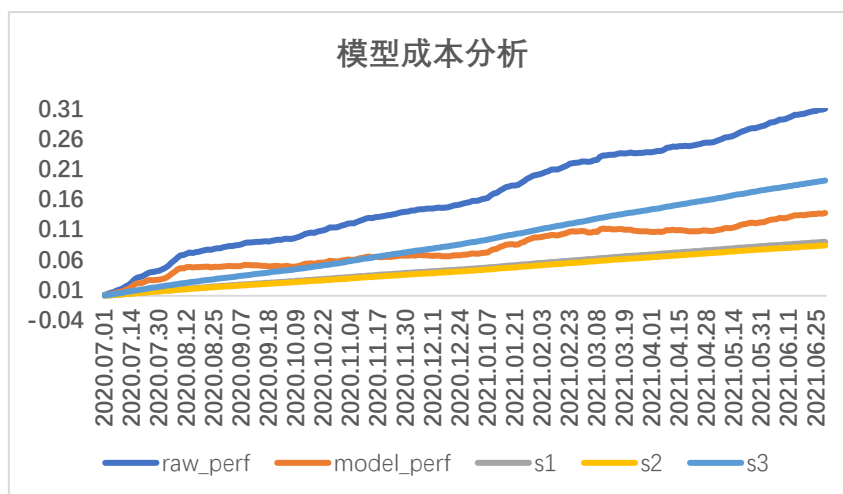
平后，增加股票数量所带来的风险的分散是有限的，反而增加了选到表现不佳的标的（即历史数据显示很活跃，但实际并不活跃的标的），因此 50 只标的的曲线虽然与 30 只标的的曲线走势基本一致，但要略差的原因。

总体来说，股票数量的选取对于模型的结果有影响，但并不会从根本上改变模型的结果，更多的是平滑模型的表现。股票数量的选取就是对分散风险以及股票选择正确率之间的一种平衡，因此可以通过历史数据来测试出表现最好的参数。

第五节 模型表现成本分析

图 3.11 对模型各类累计成本进行了展示，具体表现如下：

图 3.11 模型的成本分析



其中， raw_perf 表示不计任何成本的策略表现，买卖都按中间价进行的结果； $model_perf$ 是扣除所有成本后模型的表现； $s1$ 代表开仓时以对手价买入花费的价差成本； $s2$ 代表开仓时以对手价卖出的价差成本； $s3$ 代表在反转信号出现时平仓的价格与信号出现前平仓的价差成本。其中 $raw_perf = model_perf + s1 + s2$ 。

模型的主要成本可以分为 3 类：1、开仓时对手价的价差。2、平仓时对手价的价差。3、信号反转和最后一个信号的价差。首先，前 2 类成本是浅显易懂的，这 2 类成本是为了快速成交而支付的成本。当然了，高频交易对于平仓的要求是很高的。正如本文所说的OFI指标，当该指标出现反转信号的时候，很有可能价格也会出现反转，不以对手价成交的损失可能会比以对手价成交的成本更高，这一类成本是必要的。但是另外 2 类成本值策略值得优化的地方。

首先，针对开仓成本而言，打对手价、本方最优价、中间价都是可行的，哪个价格最好是不得而知的，这就需要在成本(打对手价而产生的费用)与收益(因为买入而产生的收益)间权衡。

最重要的成本是第三类成本。从累计成本来看，第三类成本为策略贡献了最高的成本，贡献幅度为 19%，而模型的收益也仅有 13%，可想而知这是一笔多大的费用。之所以这类成本会很高并不是偶然，而是必然的。由于模型使用OFI作为信号的产生标准，当出现平仓信号时，这预示着价格将要出现反转。而且由于OFI是个很强势的同期指标，OFI出现信号的那一刻，价格是会大幅度变动的，但这个问题并不是不可解决的。之所以会承担这样的高成本的根本原因是信号出现的不及时所致的。OFI指标最初即是一个利用逐笔数据进行计算的指标，出于简便的目的才将其降维成订单簿数据计算，如果利用逐笔数据进行交易，那么策略就可以更快的识别到买卖点，虽说这不可能将成本完全消除，但是可以对成本在最大程度上进行压缩，大幅提高收益。

第六节 本章小结

本章根据订单簿数据研究的结果构建了一个OFI高频策略模型，并进行了实证，将其与旧OFI高频策略模型、VPIN策略模型的结果进行了直接对比。实证结

果表明OFI高频策略模型在假设的前提下获得了远超前于旧OFI高频策略模型与VPIN策略模型的模型表现，表现出了低风险，收益稳定的特点。同时，本章也对模型与实战之间的差异进行了说明。同时，本章对于模型参数的敏感性进行了检验，认为模型对指数平滑参数、股票数量并不是特别敏感，但对阈值的选取敏感。但虽然模型对阈值是敏感，但在一个范围内的阈值都呈现出了稳定的正收益。最后，本章也从成本的角度出发，认为模型依然有很大的改进之处，如若改进成本，模型表现将更进一步。但值得注意的是，高频模型的回测结果与实战之间是存在着较大的差异的，若想将策略运用于实现，依然有很多细节需要深入考虑。

第五章 总结与展望

一、总结

Cont (2014) 等人在对美股的 level-1 数据进行研究后提出了一个衡量多空双方不平衡性的全新指标——*OFI* 指标。本文将该指标的构建逻辑作为理论基础, 根据 A 股与美股之间存在的差异性提出了一个新的 *OFI* 指标, 并将 *OFI* 运用于 2020.07-2021.06 为期 1 年的 A 股市场上进行了实证分析。结果表明, *OFI* 在非常高频的时间内对于未来短期股价的变动是有显著的预测性的。并且, *OFI* 值越大意味着趋势的可持续性越强, *OFI* 值越小意味着震荡的可能性越高。

根据与 *VPIN* 策略模型、旧 *OFI* 高频策略模型的历史数据回测结果表现的对比, *OFI* 高频策略模型是一个能在控制住风险的前提下获得稳定收益的模型, 即 *OFI* 是一个切切实实可以用于实战的指标。本文也指出了 *OFI* 策略模型与实战之间的差异, 想要真正将 *OFI* 运用于实战中, 依然有很长的路需要走。

目前, 我国对于 A 股高频的微观结构研究并不丰富, 尤其是对于多空双方不平衡性的衡量这一主题的研究并不多, 本文的研究对于未来对该领域的深入研究是有一定的帮助的。

二、展望

本文在研究的过程中, 发现有许多衍生出来的内容是值得后续进行进一步研究改进的:

首先, 本文在研究 *OFI* 与订单簿斜率指标时发现, *OFI* 更加擅长于识别持续的短期趋势, 而订单簿斜率更加擅长于识别平稳状态下的价格变动, 然而市场的

主要状态就是平稳与趋势，因此可能 2 个指标的有机结合可以带来意想不到的效果。但直接简单的叠加并不是一个可取的方案，因为 2 个指标的测量单位并不相同，因此需要思考一些方案将它们合理的进行有机结合。

其次，正如文末所提到的，*OFI*指标最初始的定义是基于逐笔数据的，用逐笔数据计算的*OFI*会更加细腻，对短期价格走势的判断更加准确，从而实现对于买卖交易点的优化，加强模型的表现。

最后，正如文中所提到的，*OFI*高频策略模型的一些假设与实战存在着巨大的差异，通过优化这些假设或许可以加强*OFI*模型用于实战的可能。

参考文献

- [1] 陈收,李双飞,黎传国,2010,《订单差、交易量变化对股票价格的冲击》,《管理科学学报》,13(9),68-75.
- [2] 池文涛,2020,中国股票市场限价单成交概率与微观结构高频分析[D],上海财经大学.
- [3] 刘波,2009,基于连续双向拍卖交易机制的金融市场微观结构研究[D],电子科技大学.
- [4] 麻文字,王春峰,房振明,2013,《A股市场订单流不平衡的价格冲击效应研究》,《武汉理工大学学报:信息与管理工程版》,35(6),937-941.
- [5] 孙煦初,2018,订单簿特征指标与市场波动率——基于我国证券市场的实证研究[D],南京大学.
- [6] 王庆学,2020,基于多级订单流不平衡的超短期价格趋势预测[D],山东大学.
- [7] 赵景东,朱洪亮,李心丹,2018,《基于限价订单簿高频动态演化的价格冲击及日内模式研究》,《证券市场导报》,4,52-60.
- [8] Ahn H J., Bae K H., Chan K., 2002, "Limit Orders, Depth, and Volatility: Evidence from the Stock Exchange of Hong Kong", *Journal of Finance*, 56, 767-788.
- [9] Anderson EW., Ghysels E., Juergens JL., 2005, "Do Heterogenous beliefs matter for asset pricing?", *Review of Financial Studies*, 18(3), 875-924.
- [10] Arora, Haritika, 2017, "Stylized facts and trends of high frequency data in financial markets", *Asian Journal of Research in Business Economics and Management*, 7(7), 303-317.
- [11] Akyildirim E., Sensoy A., Gulay G., Corbet S., Salari H N., 2021, "Big data analytics, order imbalance and the predictability of stock returns", *Journal of Multinational Financial Management*, 62, 100717.
- [12] Bagehot, 1971, "The only game in town", *Financial Analysts*

Journal, 27(2),12-14.

[13] Bouchaud J P., Gefen Y., Potters M., Wyart M., 2003, “Fluctuations and response in financial markets: the subtle nature of ‘random’ price changes”, 2004, *Quantitative Finance*, 4, 176.

[14] Bouchaud J P., Bonart J., Donier J., Gould MD., 2018, “Trades, Quotes and Prices: Financial scope.”, Cambridge University Press.

[15] Banerjee AK., Pradhan HK., 2020, “Order Flows, Investor Sentiments and Feedback Trade in Index Futures Market”, *Journal of Quantitative*, 1-16.

[16] Cont R., 2011, “Statistical Modeling of High Frequency Financial Data: Facts, Models and Challenges”, *IEEE Signal Processing*, 28, 16–25.

[17] Cont R., Kukanov A., Stoikov S F., 2014, “The Price Impact of Order Book Events”, *Journal of Financial Econometrics*, 12(1), 47—88.

[18] Cenesizoglu T., Zhou X Z., Dionne G., 2014, “Effects of the Limit Order Book on Price Dynamics”, *Effects of the Limit Order Book on Price Dynamics*, 2523643.

[19] Easley D., Kiefer N M., O’Hara M., Paperman J B., 1996, “Liquidity, Information, and Infrequently Traded Stocks.”, *The Journal of Finance*, 51(4).

[20] Easley D., Prado M M L D., O’Hara M., 2012, “Flow Toxicity and Liquidity in a High-frequency World”, *Review of Financial Studies*, 25(5).

[21] Garman MB., 1976, “Market microstructure”, *Journal of Financial Economics*, 3, 257-275.

[22] Hautsch H., Huang R., 2012, “The Market Impact of a Limit Order”, *Journal of Economic Dynamics and Control*, 36(4), 501-522.

[23] Huang H C., Su Y C., Liu Y C., 2014, “The performance of imbalance-based trading strategy on tender offer announcement day”, *Investment management and financial innovations*, 11(2), 38-46.

[24] Jain PK., Jain P., McInish TH., “Does high-frequency trading increase systemic risk?”, *Journal of Financial Markets*, 31, 1-24.

- [25] KL Chan, WM Fong,2000, “Trade Size, Order Imbalance, and the VolatilityVolume Relation”. *Journal of Financial Economics*,57, 247–273.
- [26] Kalay A., Sade O., Wohl A., 2004, “Measuring stock illiquidity: An investigation of the demand and supply schedules at the TASE”,*Journal of Financial Economics*, 74(3), 461-486.
- [27] Kim J W., Lee J., Morck R., 2004, “Heterogeneous Investors and their Changing Demand and Supply Schedules for Individual Common Stocks”, *SSRN Electronic Journal*,2004,10410.
- [28] Lee CMC., Ready MJ., 1991, “Inferring trade direction from intraday data.”, *The Journal of Finance*,46(2),733-746.
- [29] Næs R., Skjeltorp J A., 2006, “Order book characteristics and the volume–volatility relation: Empirical evidence from a limit order market”, *Journal of Financial Markets*, 9, 408–432.
- [30] O’Hara M., 2015, “High frequency market microstructure”, *Journal of Financial Economics* ,116(2),257-270.
- [31] Q. Wang, B. Teng, Q. Hao, Y. Shi,2021, “High-frequency Statistical Arbitrage Strategy Based on Stationarized Order Flow Imbalance”, *Procedia Computer Science*,187,518-523
- [32] Ravi R., Sha Y.,2014, “Autocorrelated order-imbalance and price momentum in the stock market”, *International Journal of Economics and Finance*,6(10),39-54.
- [33] Rannou Y., 2017, “Liquidity, information, strategic trading in an electronic order book: New insights from the European carbon markets”, *Research in International Business and Finance*, 39, 779-808.
- [34] Shen D., 2015, “Order imbalance based strategy in high frequency trading”, oxford university.
- [35] Tian X., Duong H N., Kalev P S., 2019, “Information content of the limit order book for crude oil futures price volatility”, *Energy Economics*, 81,584-597.
- [36] Xu K., Gould Martin D., Howison Sam D., 2019, “Multi-Level

Order-Flow Imbalance in a Limit Order Book”, *Market Microstructure and Liquidity*, 4(03n04), 1950011.