



2022年7月1日

期货研究

金融衍生品研究所

中证 1000 衍生品系列专题四：

IM 基差展望及基于机器学习算法的历史模拟

✍ 虞堪 投资咨询从业资格号：Z0002804

✉ Yukan010359@gtjas.com

✍ 李宏磊（联系人） 从业资格号：F3078764

✉ Lihonglei023572@gtjas.com

报告导读：

- **IM 上市的初期，我们认为基差更多会受到情绪的影响。**（1）刚上市的阶段，品种的投资结构并不稳定，这可能会导致基差的波动相对比较大。（2）目前行情偏向中小盘，多头情绪或引导基差偏强。
- **稳定之后，IM 的基差大概率弱于 IC，预计在 IC 的贴水幅度上增加 2%-5%。**主要原因有以下三方面：（1）**融券利率：**融券利率和基差会有互相影响。中证 1000 对应 ETF 的融券成本往往高于中证 500ETF，而 1000ETF 的融券成本和贴水成本不会相差过大；（2）**产品预期收益：**产品预期收益率也会影响管理人使用 IM 的意愿。中证 1000 中性、增强等产品的预期收益率更高，管理人使用中证 1000 去做中性产品的意愿也会更强，也可以承受更高的对冲成本；（3）**替代性：**高相关性使得 IM 和 IC 的贴水不能相差过大。如果二者升贴水差异很大，可能出现互相替代的现象，或者套利的机会。
- **IM 上市之后对 IC 的基差产生影响，但 IC 的基差收敛和扩大均存在可能，取决于参与者的情况。**（1）对冲需求转移可能分担 IC 的对冲压力，但也可能刺激市场整体中性产品规模的上升；（2）大量雪球产品转移到中证 1000 上来，中证 500 的买入对冲需求会减少，IC 的基差可能走弱，但投资者对中证 1000 雪球产品的兴趣还有待观察。
- **由于机器学习算法在样本内的优异表现，我们尝试将机器学习算法应用于股指基差的解释上，并在中证 1000 期货上市之际，对 IM 的基差进行历史模拟。**本文主要尝试将 XGBoost 的回归算法应用于股指基差的解释和模拟。经过模拟我们发现，**近期 IM 基差模拟值持续低于 IC 基差，2022 年以来，IM 基差（模拟）平均低于 IC 基差 4.7%，**这与我们此前对于 IM 上市后基差的定性判断较为一致。

(正文)

1. IM 基差的基本面信息

上市的初期，我们认为基差更多会受到情绪的影响。

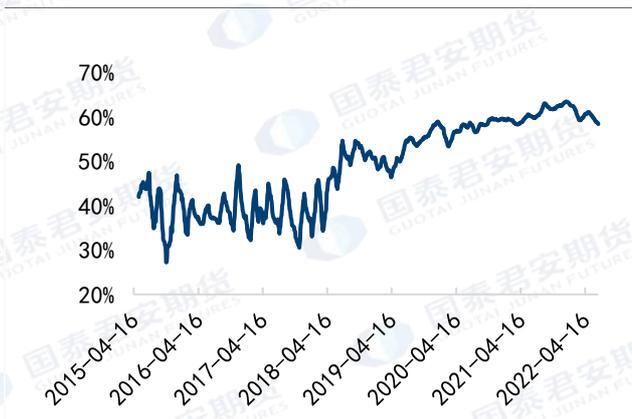
1. 刚上市的阶段，品种的投资者结构并不稳定，这可能会导致基差的波动相对比较大。我们参考 IF 和 IC 上市时的情况，观察前十大会员持仓占市场总持仓的情况，这可以部分代表机构的持仓情况，无论是 IF 还是 IC，持仓结构都有缓慢偏向机构化的特点，并且上市初期的占比比例波动明显更大。不过经过目前三个品种的运行，市场对股指期货的认知已经相当成熟，加上目前整个行情也相对平稳，IM 上市后自我调整和趋向稳定的时间窗口必然会缩短。
2. 目前行情偏向中小盘，多头情绪或引导基差偏强。由于近期中证 1000 的表现明显优于其他指数，投机情绪比较浓，投资者对于该指数后市的表现也相对乐观，贴水幅度可能不会太深，甚至会强于中证 500 股指期货。本轮市场的上涨很大程度上要归功于政策的红利，而且这次政策和以往有所不同，没有使用传统的经济政策去刺激，重心还是在一些转型升级方面的领域，比如说新能源、芯片等产业的中高端部位。从估值的角度，目前 A 股的 PE 处于过去 10 年 50%分位左右，而中证 1000 处于 20%历史分位数以下。因此后市中证 1000 可能更容易受到市场的追捧。

图 1：前十大会员持仓占比情况（IF）



资料来源：Wind、国泰君安期货金融衍生品研究所

图 2：前十大会员持仓占比情况（IC）



资料来源：Wind、国泰君安期货金融衍生品研究所

表 1：本轮反弹以来（2022 年 4 月 27 日起）指数行情信息

	中证 1000	上证 50	沪深 300	中证 500
年内涨跌幅	26.3%	11.8%	15.1%	18.7%
年化波动率	25.1%	18.5%	19.1%	21.9%
平均换手率	215.3%	31.5%	52.9%	143.1%

资料来源：Wind、国泰君安期货金融衍生品研究所

稳定之后，IM 的基差大概率弱于 IC，预计在 IC 的贴水幅度上增加 2%-5%。

1. 融券利率：融券利率和基差会有互相影响。毕竟两者都是做空的工具，在一定程度上也可以相互替代，如果 1000ETF 的融券利率和贴水成本相差过大，可能存在套利机会。我们从融资融券的信息来看，由于中证 1000 的波动率更大，券源也更少，因此中证 1000 对应 ETF 的融券成本往往高于中证 500ETF，这样也会导致在期货市场中，使用 IM 对冲的成本比 IC 更高，IM 的基差更弱。

- 产品预期收益：产品预期收益率也会影响管理人使用 IM 的意愿。**中长期来看，股指期货的贴水是由需求决定的，造成 IC 长时间贴水的一个重要原因是中性产品对 IC 的使用，从整个市场来看，目前 IC 作为风险对冲工具的使用程度已经超过 IF，即使使用 IC 的成本更高，管理人还是更愿意从中证 500 的成分股中优选个股。而中证 1000 中性、增强等产品的预期收益率更高，根据不完全数据，挂钩中证 1000 的私募产品比挂钩 500 的产品收益高 10-20%，所以管理人使用中证 1000 去做中性产品的意愿也会更强，同时也可以承受更高的对冲成本，因为超额的收益部分可以抵消掉 IM 的贴水成本。当然如果贴水水平过深，中性产品的预期收益率下降，也会降低管理人使用 IM 这个风险管理工具的意愿，因此从这个角度看，IM 的基差大概率弱于 IC，但不会相差太大。
- 替代性：高相关性使得 IM 和 IC 的贴水不能相差过大。**但是本身 500 和 1000 的相关性是很高的，历史相关性超过 0.95，如果升贴水差异很大，可能出现互相替代的现象，或者套利的机会，因此他们之间的基差水平不会相差太多。

表 2：中证 1000 增强产品的超额收益

基金代码	基金名称	基金规模 (亿元)	近一个月超额%	近三个月超额%	近一年超额%
161039.SZ	富国中证 1000 指数增强 A	22.2	-0.6	4.3	5.2
014201.OF	天弘中证 1000 指数增强 A	5.7	1.8	8.5	0.1
005313.OF	万家中证 1000 指数增强 A	4.1	0.6	4.1	13.7
015495.OF	景顺长城中证 1000A	3.2	-4.4	7.7	0.1
015466.OF	太平中证 1000A	2.1	-1.4	8.9	0.1
006165.OF	建信中证 1000 指数增强 A	2.1	1.5	6.1	8.2
014831.OF	兴银中证 1000 指数增强 A	1.6	0.7	6.5	0.1
004194.OF	招商中证 1000 指数增强 A	1.1	-0.1	5.2	9.7
014125.OF	华夏中证 1000 指数增强 A	0.9	0.4	6.7	0.1
003646.OF	创金合信中证 1000 增强 A	0.5	1.6	8.1	7.6

资料来源：Wind、国泰君安期货金融衍生品研究所

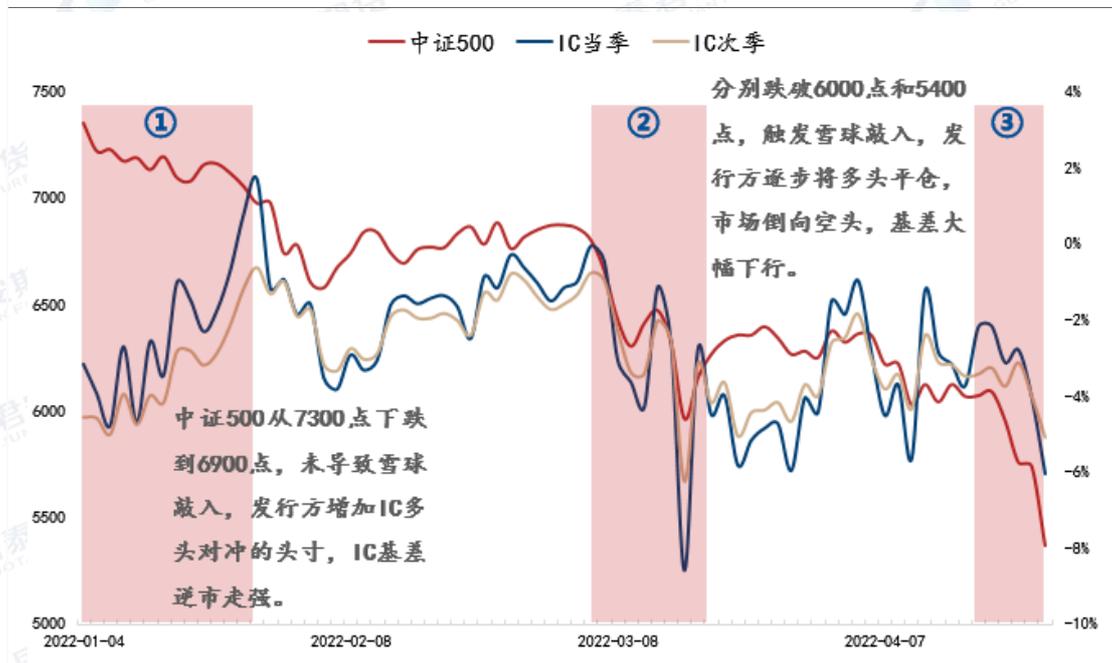
IM 对现有股指期货基差的影响

目前现有品种 IH、IF、IC 的投资者群体相对割裂，IM 上市之后我们认为只会对 IC 的基差产生影响。至于 IC 的基差会收敛还是扩大，我们认为都有可能，取决于届时的参与者情况。

- 对冲需求的转移。**在 IM 上市之前，有很多管理人采用 300+500 的组合去拟合中证 1000 的走势，对冲效果必然不理想，新品种上市之后，将会有不少对冲需求从 IC 端转移至 IM 端，但 IC 贴水并不一定会因为对冲需求的转移而收敛，因为从动态的角度来看，IM 的加入，可能会刺激市场整体中性产品规模的上升，届时 IC 贴水也有可能不降反升。
- 雪球挂钩产品的转移。**对于雪球产品，挂钩中证 1000 的雪球产品票息更高是毫无疑问的，据了解，大约比挂钩中证 500 的雪球票息高 3% 左右，如果大量雪球产品转移到中证 1000 上来，中证 500 的买入对冲需求会减少，IC 的基差会走弱。根据我们测算，雪球的 IC 多头对冲盘，可能占到市场 1/4 左右的头寸，对基差的影响非常大，可以参考上半年雪球敲入之后基差的变化（详情参考专题《雪球产品集中敲入的影响分析》）。但是这里需要注意的是，目前市场对中证 1000 的认知和接受程度不高，从挂钩的产品数量和规模就可以看出，虽然挂钩中证 1000 的雪球可以给出 3% 的更高票

息，但并不确定投资者是否愿意买单，市场人士的观点也没有统一，可能需要继续观察。

图 3：雪球产品敲入对 IC 基差产生较大影响



资料来源：Wind、国泰君安期货金融衍生品研究所（详情见专题《雪球产品集中敲入的影响分析》）

2. 用机器学习算法模拟 IM 基差

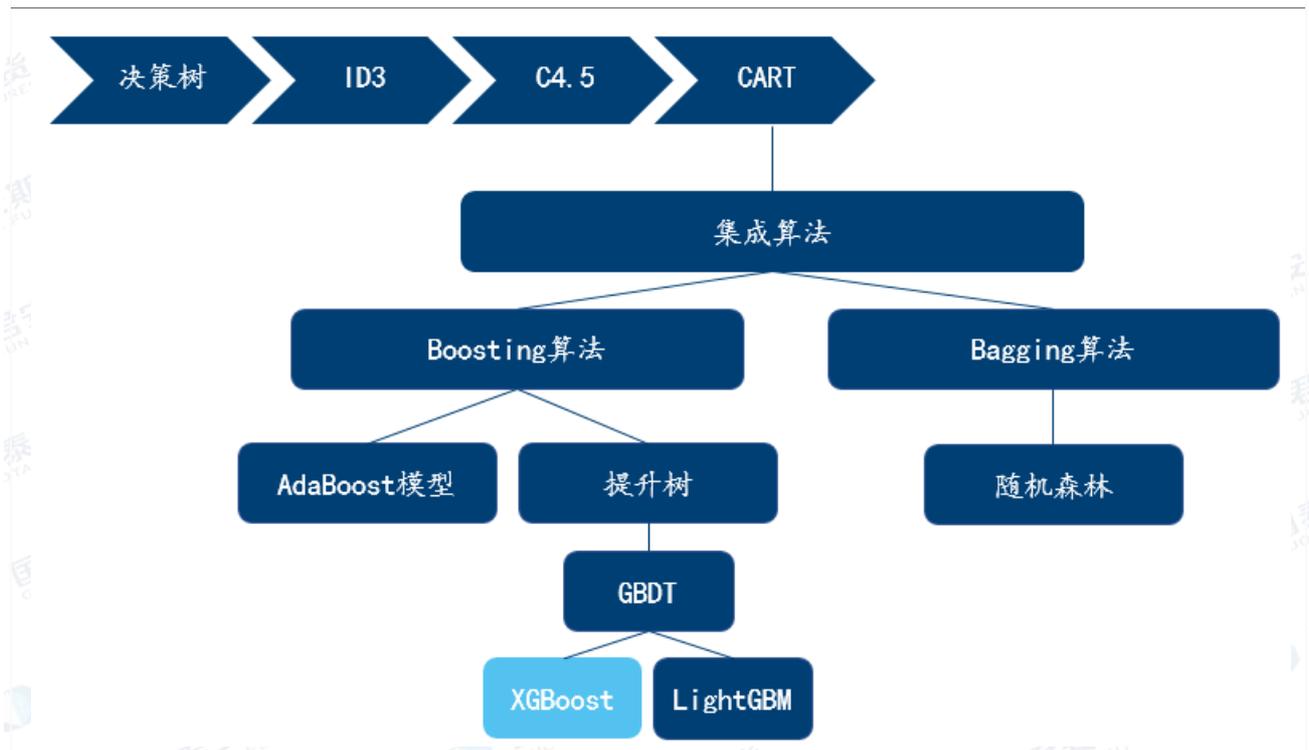
市场上越来越多的投资者开始将机器学习的方法应用于因子挖掘以及投资等等各个方面，由于机器学习算法在样本内的优异表现，我们尝试将机器学习算法应用于股指基差的解释上，并在中证 1000 期货上市之际，对 IM 的基差进行历史模拟。

2.1 决策树模型简介

为了解决离散状态下的分类问题，E. B. Hunt 等人于上世纪 60 年代提出了 Concept Learning System，是现代决策树的先驱，决策树思想广泛应用于各个领域，我们在后来的期权定价中亦能捕捉到决策树思想的影子。

决策树的思想可以简单概括为，按照特定的条件将样本划分为不同类别，相对于连续的线性模型，决策树更是一种离散型的分类判断。简单的决策树模型针对样本的每一个特征，进行逐层分类，最终实现分类和预测的效果。决策树本身比较容易理解，但单个决策树容易产生样本过拟合的问题，研究发现，通过将多个弱分类器组合成为一个强分类器的方法，可以提升稳健性，解决过拟合的问题，因此在此逻辑上发展出了 Bagging 和 Boosting 两个解决方案。

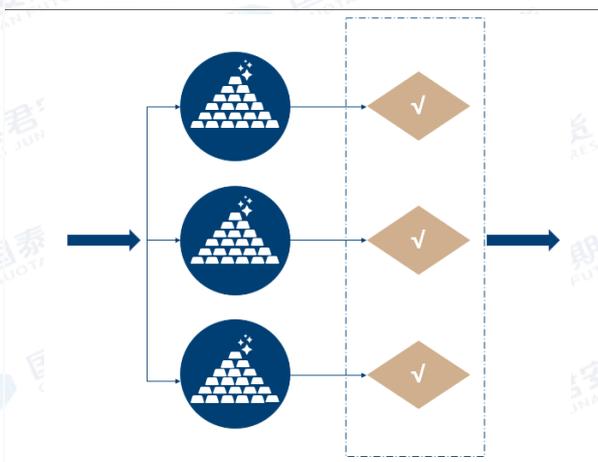
图 4：决策树算法的进化历程



资料来源：国泰君安期货金融衍生品研究所

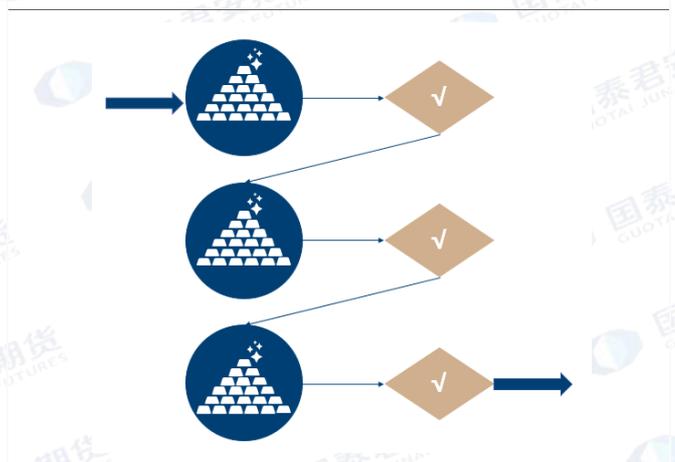
Bagging 和 Boosting 的区别在于，Bagging 算法对各个决策树进行并行训练，各个决策树的地位相同且互相独立，最终依据全部独立决策树的决策结果进行综合判断。而 Boosting 则是通过对单个决策树的残差进行渐进式预测，即每一次训练都是在训练前一次的残差，学习前一个决策树的错误，在此过程中不断减小偏差。

图 5：Bagging 类模型示意



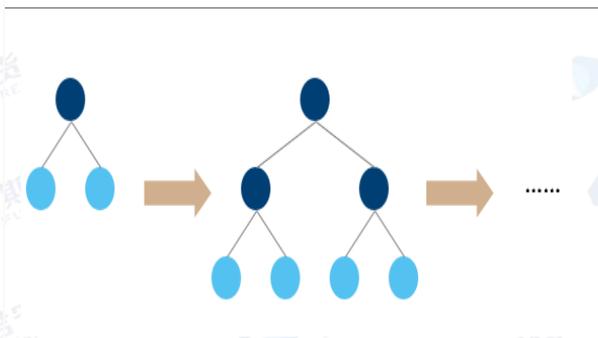
资料来源：国泰君安期货金融衍生品研究所

图 6：Boosting 类模型示意



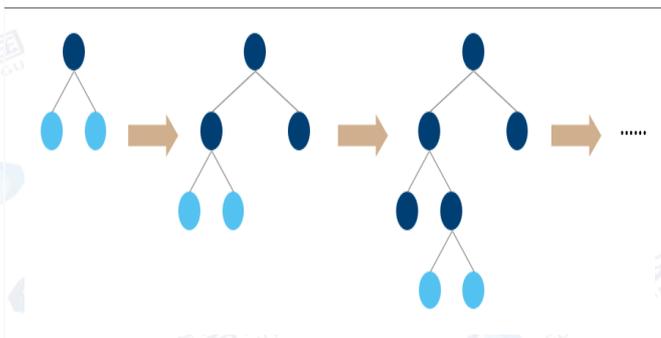
资料来源：国泰君安期货金融衍生品研究所

图 7: XGBoost 的层生长模式 (Level-wise)



资料来源: 国泰君安期货金融衍生品研究所

图 8: LightGBM 的叶子增长模式 (Leaf-wise)



资料来源: 国泰君安期货金融衍生品研究所

Bagging 模型中最为常见的就是随机森林模型, 而 Boosting 模型下比较流行的框架是 XGBoost 和 LightGBM 框架, 我们这里主要探究 XGBoost 的对股指基差的解释效果, XGBoost 在传统的 GBDT (Gradient Boosting Decision Tree) 的基础上, 添加了一个正则项, 可以有效控制模型整体的复杂度, 提升模型的稳健性, 其损失函数 (目标函数) 为:

$$Obj = \sum_{i=1}^n L(y_i - \hat{y}_i) + \sum_{t=1}^T \Omega(f_t)$$

$$\hat{y}_i = \sum_{t=1}^T f_t(x_i)$$

除应用于分类问题外, XGBoost 还可以对连续序列进行回归计算, 接下来我们尝试将 XGBoost 的回归算法应用在股指基差的解释上。

2.2 在股指基差解释方面的应用

解释变量方面, 由于本身的差值属性, 基差受市场风险情绪和氛围的影响较大, 因此我们从偏宏观的货币价格、市场风险、流动性、以及资金层面的市场情绪四个维度选取解释变量, 货币价格维度我们选取 R007 加权利率和 10 年期国债到期收益率两个变量; 风险因子维度, 考虑到波动率的聚类特性, 我们选取 GARCH (1, 1) 估计量、振幅和指数当日涨跌幅三个变量, 流动性维度我们选取指数成交量和换手率, 市场情绪维度, 考虑到数据长度等问题, 我们保留全单净主动买入额作为解释变量。由于上述解释变量均为指数现货的数据, 因此不存在数据不可得的问题。

我们尝试利用上述四个维度的解释变量对股指期货的历史基差进行解释, 由于股指期货当季合约基差相对稳定, 受合约到期日影响较小, 因此我们选取股指期货当季合约的年化基差率 (剔除分红) 作为被解释变量。

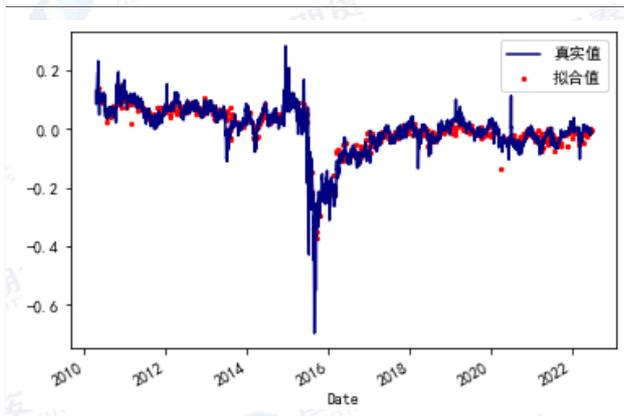
表 3：解释变量

一级变量	二级变量	指代简称
货币价格	R007 加权利率	R007
	10 年期国债到期收益率	TBILL
风险因子	GARCH(1, 1)	Garch
	振幅	Swing
	涨跌幅	Pct_chg
流动性因子	成交量	Volume
	换手率	Turn
市场情绪	全单净主动买入	Mf_amt

资料来源：国泰君安期货金融衍生品研究所

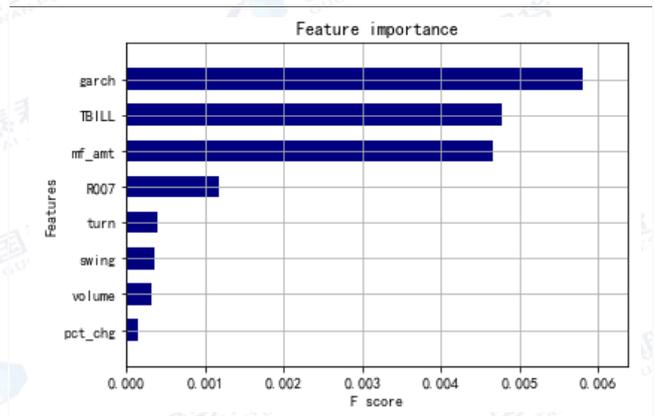
首先我们以 IF 为例，观察 XGBoost 在历史样本内的拟合效果。这里选取 IF 上市以来（即 2010 年 4 月 16 日至今）的全部数据作为样本，空值以前值填充，从中随机选取 80% 的样本作为训练集，其余 20% 的样本作为测试集。拟合结果表明用 IF 数据样本训练得到的模型对 IF 基差本身的拟合效果较好，误差平方和 (SSE) 为 0.55， R^2 为 0.85，同品种样本内效果较好，符合我们的预期。另外，波动率对 IF 基差的影响最大，资金价格和资金净买入额的影响次之，指数涨跌幅、成交量、振幅的影响相对较小。

图 9：IF 本身的拟合效果



资料来源：Wind、国泰君安期货金融衍生品研究所

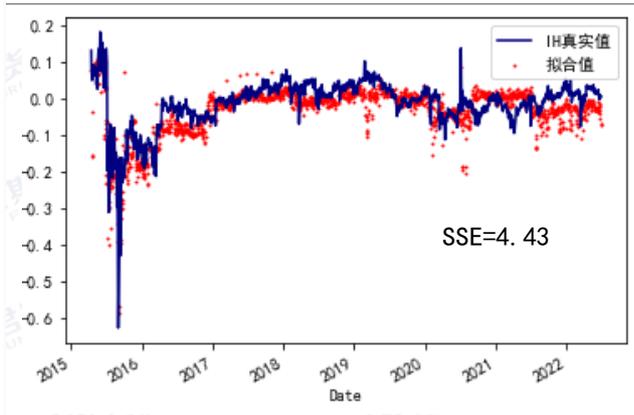
图 10：IF 模型的变量相对影响



资料来源：国泰君安期货金融衍生品研究所

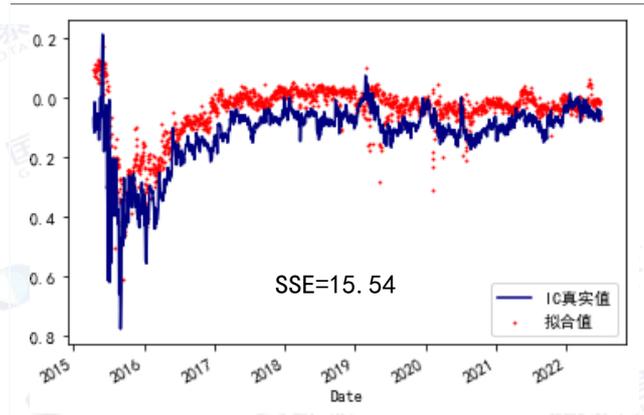
接下来，我们尝试以 IF 数据训练得到的模型，平移至上证 50 指数和中证 500 现货指数数据上，对 IH 和 IC 的基差进行拟合。我们发现，该模型对 IH 和 IC 的拟合值均与原值趋势相同，但对 IH 拟合效果更好，SSE 为 4.43，对 IC 拟合的 SSE 则高达 15.54。由于上证 50 和沪深 300 风格和走势相近，驱动因素类似，受到货币价格、风险情绪、流动性等方面的影响较为相似，因此以 IF 为样本训练的模型对 IH 拟合效果较好，但对风格差异较大的 IC 拟合效果较差。

图 11: IF 模型对 IH 的拟合效果



资料来源: Wind、国泰君安期货金融衍生品研究所

图 12: IF 模型对 IC 的拟合效果



资料来源: Wind、国泰君安期货金融衍生品研究所

由于中证 1000 指数和中证 500 指数风格相近, 均属中小盘偏成长风格的指数, 历史相关性较高, 因此, 基于以上对 300 和 50 的尝试, 我们有足够的理由认为, 以中证 500 指数的数据训练得到的模型, 能够较好地模拟中证 1000 股指期货的历史基差。

图 13: 中证 1000 成分股行业分布



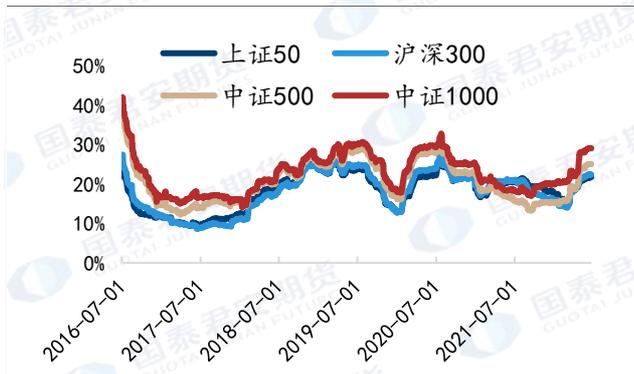
资料来源: Wind、国泰君安期货金融衍生品研究所

图 14: 中证 500 成分股行业分布



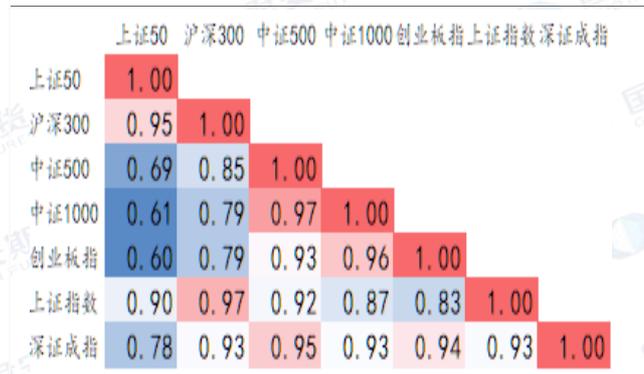
资料来源: Wind、国泰君安期货金融衍生品研究所

图 15: 中证 1000 和中证 500 波动率相近



资料来源: Wind、国泰君安期货金融衍生品研究所

图 16: 中证 1000 和中证 500 相关性高达 0.97

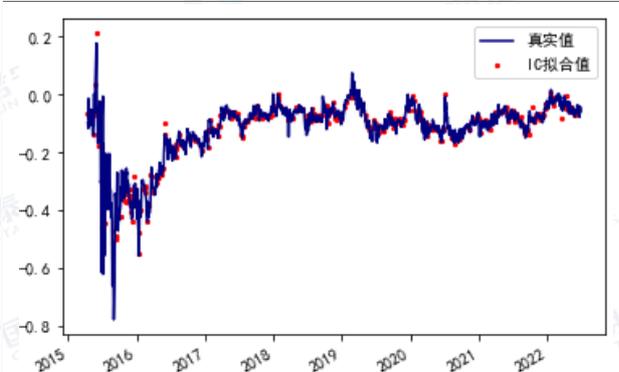


资料来源: Wind、国泰君安期货金融衍生品研究所

2.3 中证 1000 基差的历史模拟

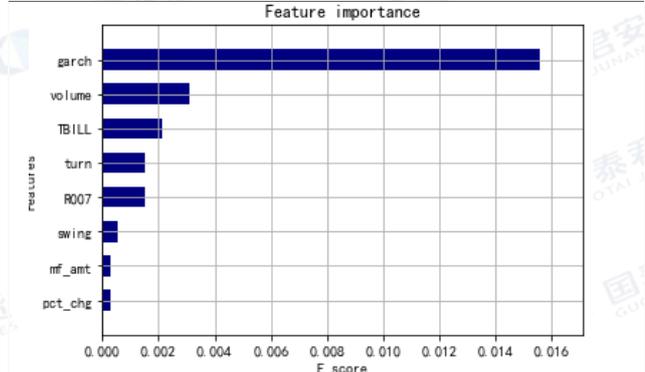
这里选取中证 500 股指期货上市以来（2015 年 4 月 16 日）的全部数据作为样本，空值以前值填充，全部数据作为模型的训练集，参数与前文模型一致，模型结果显示，IC 的基差受到波动率影响更明显，garch 变量的相对影响远高于其他任一变量。

图 17: IC 本身的拟合效果



资料来源: Wind、国泰君安期货金融衍生品研究所

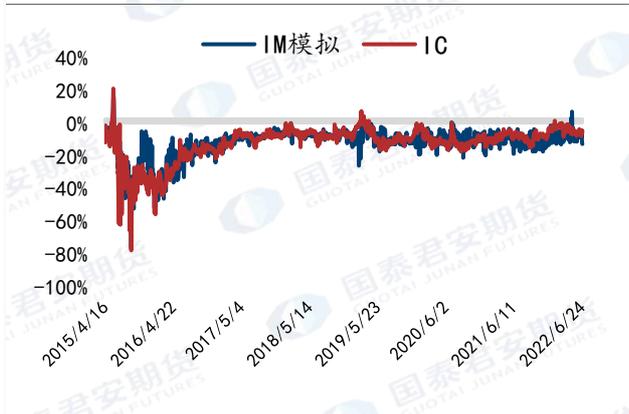
图 18: IC 模型的变量相对影响



资料来源: 国泰君安期货金融衍生品研究所

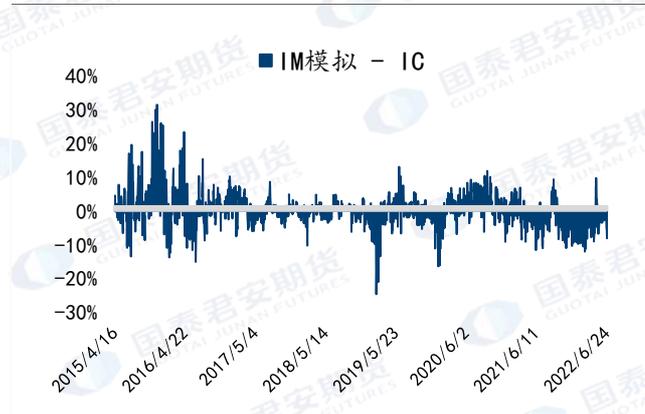
最后我们将解释变量数据替换为中证 1000 指数的数据，得到的中证 1000 基差模拟值如图所示，可以发现，近期 IM 基差模拟值持续低于 IC 基差，2022 年以来，IM 基差（模拟）平均低于 IC 基差 4.7%，这与我们对于 IM 上市后基差的定性判断较为一致。

图 19: IM 基差历史模拟值整体低于 IC 基差



资料来源: Wind、国泰君安期货金融衍生品研究所

图 20: IM 基差模拟值与 IC 基差的差值



资料来源: 国泰君安期货金融衍生品研究所

表 4: IM(模拟)-IC 差值统计属性

年份	Mean	Std	Min	25%	50%	75%	Max
2015	4.9%	8.4%	-13.4%	-0.1%	3.0%	9.4%	31.4%
2016	0.6%	5.8%	-15.1%	-2.4%	0.3%	3.8%	23.3%
2017	0.1%	2.8%	-7.5%	-1.6%	0.0%	1.7%	8.6%
2018	-0.3%	2.0%	-10.3%	-1.7%	-0.1%	1.1%	4.8%
2019	-1.4%	4.6%	-24.6%	-3.2%	-1.0%	1.2%	13.0%
2020	1.4%	4.3%	-16.4%	-0.6%	1.5%	3.9%	11.8%

2021	-2.0%	3.7%	-11.6%	-4.2%	-2.0%	0.2%	9.3%
2022	-4.7%	3.7%	-12.1%	-7.3%	-4.8%	-2.3%	9.7%
ALL	-0.04%	5.2%	-24.6%	-2.6%	-0.2%	2.1%	31.4%

资料来源：国泰君安期货金融衍生品研究所

本公司具有中国证监会核准的期货投资咨询业务资格

本内容的观点和信息仅供国泰君安期货的专业投资者参考。本内容难以设置访问权限，若给您造成不便，敬请谅解。若您并非国泰君安期货客户中的专业投资者，请勿阅读、订阅或接收任何相关信息。本内容不构成具体业务或产品的推介，亦不应被视为相应金融衍生品的投资建议。请您根据自身的风险承受能力自行作出投资决定并自主承担投资风险，不应凭借本内容进行具体操作。

分析师声明

作者具有中国期货业协会授予的期货投资咨询执业资格或相当的专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响，特此声明。

免责声明

本报告的信息来源于已公开的资料，本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的期货标的的价格可升可跌，过往表现不应作为日后的表现依据。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本报告中所指的研究服务可能不适合个别客户，不构成客户私人咨询建议。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。

市场有风险，投资需谨慎。投资者不应将本报告作为作出投资决策的唯一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向专业人士咨询并谨慎决策。

本报告版权仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用、刊发，需注明出处为“国泰君安期货研究所”，且不得对本报告进行有悖原意的引用、删节和修改。

国泰君安期货产业服务研究所

上海市静安区新闻路 669 号博华大厦 30 楼 电话：021-33038635 传真：021-33038762

国泰君安期货金融衍生品研究所

上海市静安区新闻路 669 号博华大厦 30 楼 电话：021-33038982 传真：021-33038937

国泰君安期货客户服务电话 95521