

刘文波

从业资格编号: F0286569

投资咨询编号: Z0010856

王锴

从业资格编号: F3063220

联系人

王锴

021-68982740

[wangkai@xzfutures.com](mailto:wangkai@xzfutures.com)

## · 策略介绍

**GARCH 模型展期** - 股指合约展期传统上以被动方式为主,即在到期前固定的某几个交易日进行分批次展期,以此来分散跨月价差的波动风险,有时也会配合日内的策略来减少市场冲击。然而即便被动型展期策略以不亏损为主要目标,在实盘交易中还是经常会出现跨月价差在到期前大幅波动造成亏损或难以覆盖交易成本的情况。由于跨月价差对应的两个标的同为期货合约,不会像基差随时间收敛。结合历史数据中所表现出的波动聚集性、尖峰厚尾的特点,时间序列中的 GARCH 模型能够较好地拟合价差短期波动的过程。通过预测并比较相邻两个交易日间的跨月价差,选择 1-2 个较优价差日进行展期。

**随机森林 Bagging 算法择时对冲** - 目前股指期货仅有 IF、IC 和 IH 三个品种。在实际投资过程,往往会遇到创业板、中小板和中盘成长股这样没有合约对应标的的问题。由于该条件下不是完美对冲,基差不收敛的风险较大。若仍然采用以期货合约在一固定时点对冲的方法,很有可能遇到期现两端相抵后出现净亏损的情况。基于 Bagging 算法的随机森林模型通过量化非期货对应标的板块的市场情绪以及行情量价指标,判断未来一段时间的下行风险。针对这些板块的择时对冲一方面可以部分抵消非完美对冲所带来的非系统性风险,另一方面可以在非持仓周期内灵活调配资金,保证流动性。

## · 实证研究

对测试集中所有换月时点的回测显示 GARCH 模型展期策略跑赢基准策略的概率能达到 70%以上。模型所选择的展期日实际价差大部分情况下均优于基准策略(到期前价差的平均值)。回测区间内累计收益率维持了非负的数值。

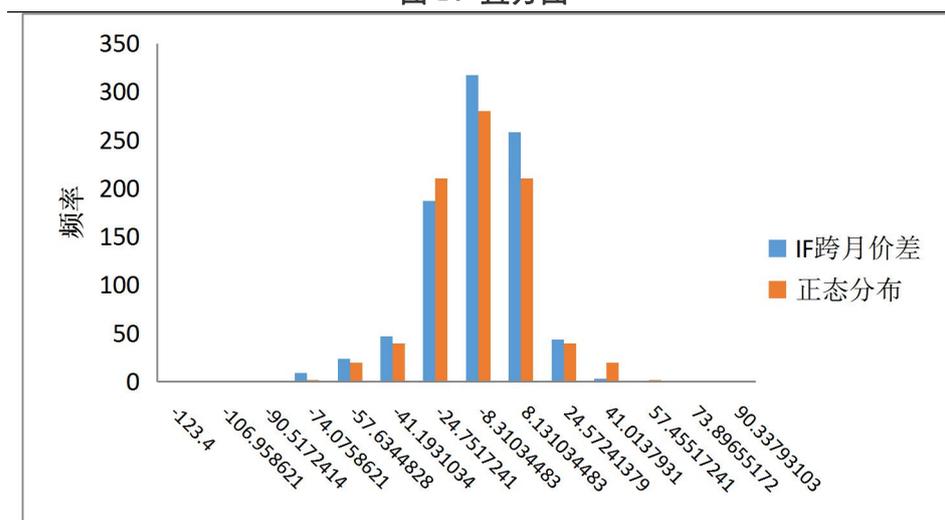
相比于在固定时间点对冲的方法,回测中基于 Bagging 算法的择时对冲策略能显著降低非完美对冲下的最大回撤,择时胜率也较高。期现两端相抵后在套保组合中仍能获得一定的净盈利。

# 1. 基于 ARIMA - GARCH 模型的 IF 合约展期策略

## 1.1 模型特点

事实上与很多金融数据一样，股指期货的跨月价差也呈现一个波动聚集性的特点，即在一段时间内一个大的波动之后可能跟随一个更大的波动，一个小的波动之后也可能跟随一个更小的波动。因此，跨月价差并不服从大部分模型中所假设的正态分布，而是具有尖峰厚尾的特征（图 1）。位于尾部较高的频数使得数据出现极值的概率比正态分布数据出现极值的概率要大。

图 1：直方图



数据来源：Wind，兴证期货研发部

如果采用期现套利中常用的趋势回归策略，很可能出现行情大幅波动的时候没有交易信号，而行情平稳的时候频繁产生信号。为了刻画变量的波动聚集特征，Bollerslev (1986) 在 ARCH 模型的基础上提出了广义条件异方差模型，即 GARCH 模型。这其中包含了一个均值方程和一个条件方差方程：

$$y_t = f(y_{t-1}, y_{t-2}, \dots | I_t) + \omega(B)r_t \quad (1)$$

$$r_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \dots + \alpha_j \sigma_{t-j}^2 + \dots + \beta_i r_{t-i}^2 \quad (2)$$

其中条件方差项的值，(2) 式残差项的方差依赖于常数项、前 p 个时期的误差平方的移动平均以及本身前一年的条件方差。如果正态分布下的假设为  $N(\mu, \sigma^2)$ ，那么对应到异方差条件下的假设就是  $f(\mu, \sigma(t)^2)$ 。GARCH 模型很好地解决了 ARMA 方程所面临的异方差问题。

## 1.2 策略构建

这里以 IF 合约 2016 年-2018 年 12 当月-次月价差作为训练集，从稳定性、协整性以及条件异方差性三个方面来验证其是否适用 GARCH 模型。

**表 1: 时间序列检验**

	P 值	结论
价差单位根检验	0.062	不平稳
当月价格	0.43	不平稳
当月价格一阶差分	<10e^8	一阶单整
次月价格	0.56	不平稳
次月价格一阶差分	<10e^8	一阶单整
回归残差协整性	0.012	协整
ARCH 效应检验	0.0082	异方差性

数据来源：Wind，兴证期货研发部

时间序列检验显示次月-当月价差并不平稳，而当月与次月合约价格都属于一阶单整。通过对其线性拟合后的残差项进行 ADF 检验，可得出在 5% 的显著性水平下当月与次月合约价格间存在协整关系。利用 Breusch-Pagan 方法对残差项的 ARCH 效应检验，可得在滞后阶数为 3 的情况下，有显著的异方差性。因此，IF 合约的跨月价差序列满足了适用 GARCH 模型的前提条件。

对于跨月价差这样的时间序列，我们会首先采用 ARIMA 模型对其进行拟合，并且根据赤池信息量准则(AIC)以及贝叶斯信息量准则(BIC)来选择 MA(q)与 AR(p)的自相关阶数。因此 (1) 式在这里被替换为：

$$\psi(B)Y_t = \theta(B)\xi_t$$

ARIMA 方程中的残差项就是 GARCH 模型所需检验的目标。很多实践研究已经发现 GARCH(1,1), 即 1 阶滞后的误差平方项的移动平均和 1 阶滞后的条件方差项的方程可以消除大部分金融数据中存在的波动聚集影响。我们在训练集中的交叉验证也反映了这一结果。

**表 2: GARCH 阶数检验**

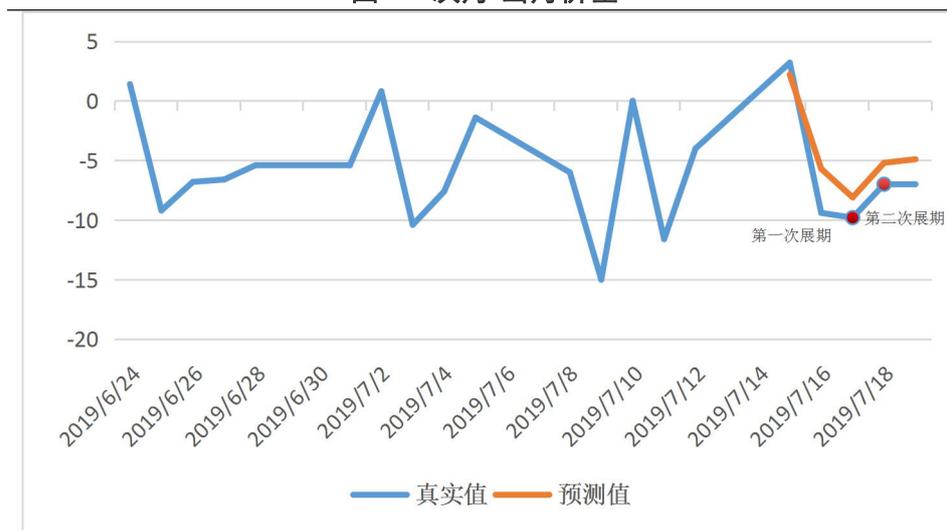
GARCH(p,q)	赤池信息准则 AIC	施瓦茨准则 SC
(1,1)	-15.23	-5.11
(1,2)	-8.20	-1.32
(2,1)	-10.42	-0.27
(2,2)	-5.82	1.38

数据来源：Wind，兴证期货研发部

通过比较不同 p,q 阶数下的 GARCH 模型，我们发现 p=1,q=1 时，参考指标达到最优。将 GARCH 模型预测的残差项代入到 ARIAM 方程中可以得到模型所预测的跨月价差。

具体操作策略上，考虑到远月合约的交易量通常会在换月前 5-7 天开始变得活跃，我们会在 IF 当月合约到期前 5 天开始预测下一交易日的开盘价差。如果预测价差缩小，当日以总量 1/2 的比例展期部分合约；如果价差扩大，则保留到下一交易日再判断。需要说明的是，到期日当天和前一天的展期总是以剩余全部的比例进行。

图 2：次月-当月价差



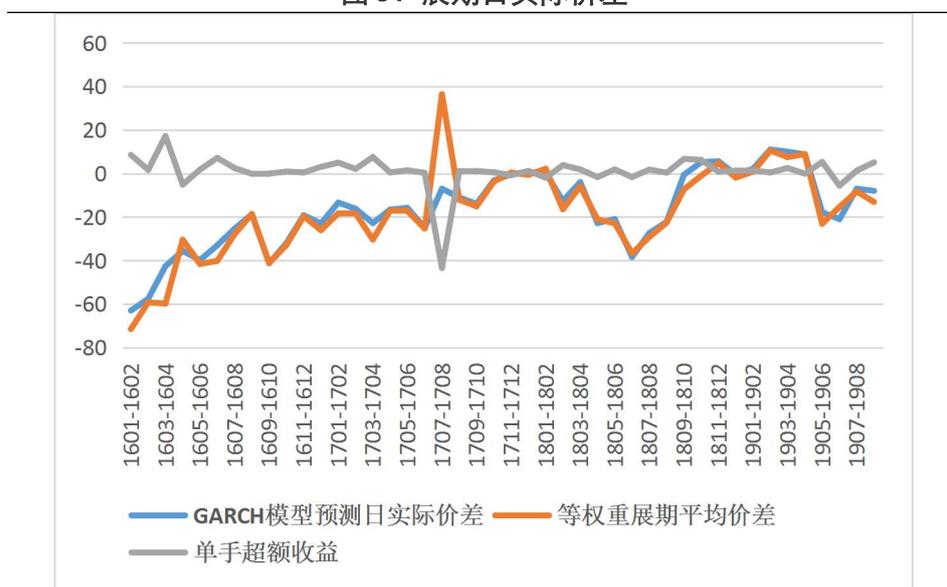
数据来源：Wind，兴证期货研发部

以 1907-1908 合约为例。由于在 7 月 15、16 日 GARCH 均预测价差会变小，没有进行展期。7 月 17 日预测价差将扩大，于是当天展期 1/2 的合约。18 日预测价差扩大，则展期剩余的合约。

### 1.3 回测表现

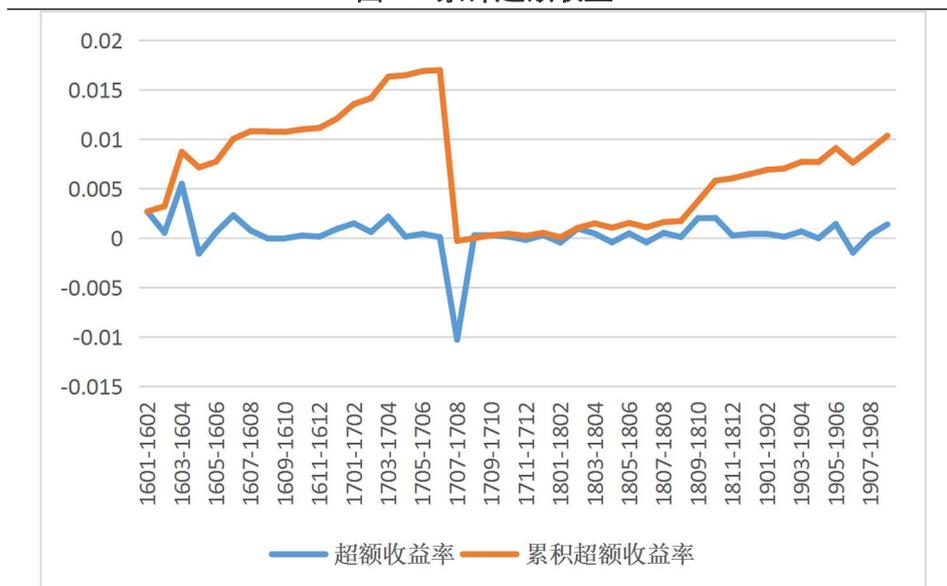
为了能独立客观地反映出模型本身所带来的收益变化，这里并未配合使用日内择时方法。回测中所用的开平仓价格来自交易日开盘后一小时内均值。同时我们也展示了基准策略（分四天等权重展期）在同一日内条件下的收益表现。

图 3：展期日实际价差



数据来源：Wind，兴证期货研发部

图 4：累计超额收益



数据来源：Wind，兴证期货研发部

表 3：单次收益比较

单次最大正超额正收益	合约名称	单次最大负超额收益	合约名称	跑赢基准策略概率
0.65%	1603-1604	-1.01%	1707-1708	70.07%

数据来源：Wind，兴证期货研发部

从 2016 年 1 月至 2019 年 8 月共计 44 次展期中可以看出，基于 GARCH 模型的择时展期策略跑赢基准策略的概率达到了 70%以上。仅展期所带来的累计超额收益为 1.07%。值得注意的是在 1707-1708 合约期间有一次较大的回撤，回顾 2017 年 7 月 19 日的日线图，当天 IF1708 合约 3991 开 3720 收，IF1707 合约 3643 开 3729 收。由于 1708 合约高开盘后大幅回落的走势，使得短时间内远月大幅贴水于近月。基准策略在这一日获得了大额的展期收益，这一极值情况也反映了一定的市场操作风险。即便如此，在整个回测区间内 GARCH 模型的展期策略依然能够保持不低于基准策略的累计收益表现。

## 2. 基于随机森林 Bagging 算法的创业板指数择时对冲策略

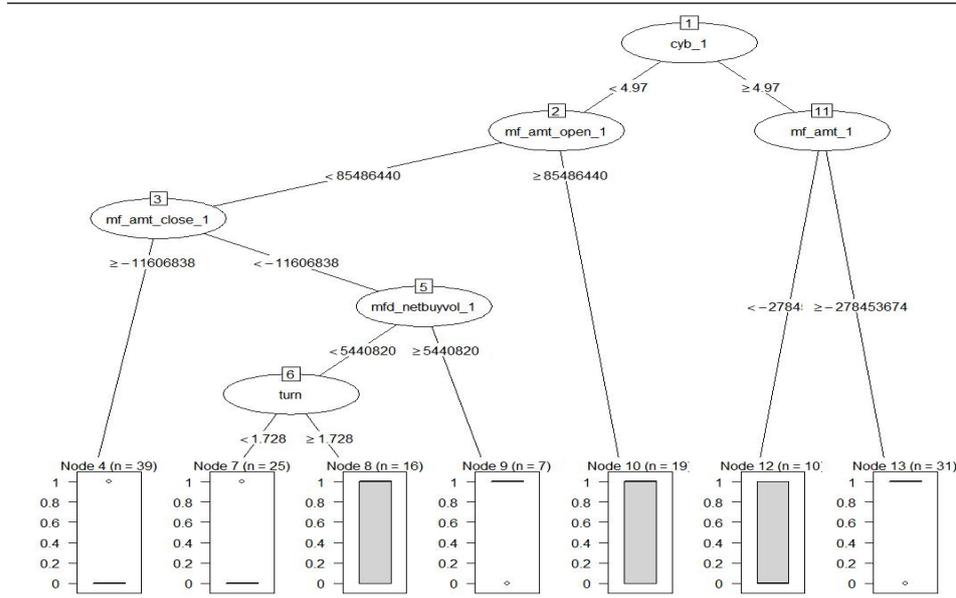
### 2.1 模型特点

随机森林(Random Forest)作为一种集成分类算法，最早由 Breiman 和 Culter (1995) 提出。近几年这种机器学习算法更是被广泛的运用到各个行业领域中。它的基本构建逻辑是运用 Bootstrap 方法有放回地抽取一部分样本构建 N 棵决策树，每一棵决策树会根据一定的判断标准从 M 个待定变量中选择最佳的分裂方式。

在这个课题中，由于存在非期货合约标的的情况，我们需要尽可能准确地判断现货指数的市场情绪以及行情走势。其中包含了大量的相关量价和宏观指标以及时间上的滞后变量，这就涉及到了高维数据所带来的多重共线性和过拟合等问题。随机森林的 Bagging 算法能比较好地处理这些情况：

1. 复杂数据处理能力。一方面，随机森林对多重共线性不敏感，只要树的数量足够多，随机森林能够处理极高维度的数据，不需要进行特征选择和降维的数据预处理。另一方面，随机森林体现出了能够处理数据集非线性和多输出的问题。
2. 预测能力稳定。由于在随机森林建立的过程中每一个决策树所用的数据仅采用总样本的一部分，最后通过大量样本投票输出预测结果（图 5）。因此不同于单一决策树需要剪枝，随机森林不容易产生过度拟合。此外，不同于 Boosting 算法会给予在训练过程中预测效果较好的弱分类器更大权重，Bagging 算法采用平均分配的机制，相对适用于分类不平衡的数据集。

图 5：决策树举例



数据来源：Wind，兴证期货研发部

## 2.2 策略构建

我们选择 IC 合约上市以来的 2015 年 6 月至 2018 年 12 月作为训练集。其中因变量为创业板指数每周四的开盘价。若一周跌幅超过  $d\%$ ，记为负号；反之则记为正号。跌幅  $d$  反映了策略实现的效果，如果跌幅设置过小，则难以与区间震荡行情区分，容易造成模型解释及预测能力较差；如果跌幅设置过大，由于数据在时间轴上的数量有限，跌幅超过  $d\%$  的次数太少，很难检验模型是否显著有效。综合样本内的频数以及对应模型的 AIC 值，选择 2% 的跌幅作为建立因变量的阈值。

表 4：阈值选择

跌幅 $d\%$	样本内频数	AIC
1%	53 次	-807
2%	39 次	-1256
3%	27 次	-1398
4%	21 次	-965
5%	18 次	-1068

数据来源：Wind，兴证期货研发部

待定变量方面，我们选取了创业板 100 指数中的 59 个因子以及对应一阶和二阶分值，共计 177 个。通过 Bootstrap 方法构建子样本并观察每一棵决策树的分支节点所用因子，剔除了那些在 600 棵决策树中所用次数在 10

次以下的因子，并保留了剩余的 92 个。这其中包含了一些可以直接从数据来源获得的变量包括市盈率、换手率、净主动买入量、上涨/下跌的成分股占总数的比例等量价指标。同时，我们也通过算法构造了一些新的变量包括机构重仓股持仓变化、指数成分股收益率与波动率的秩相关系数等。

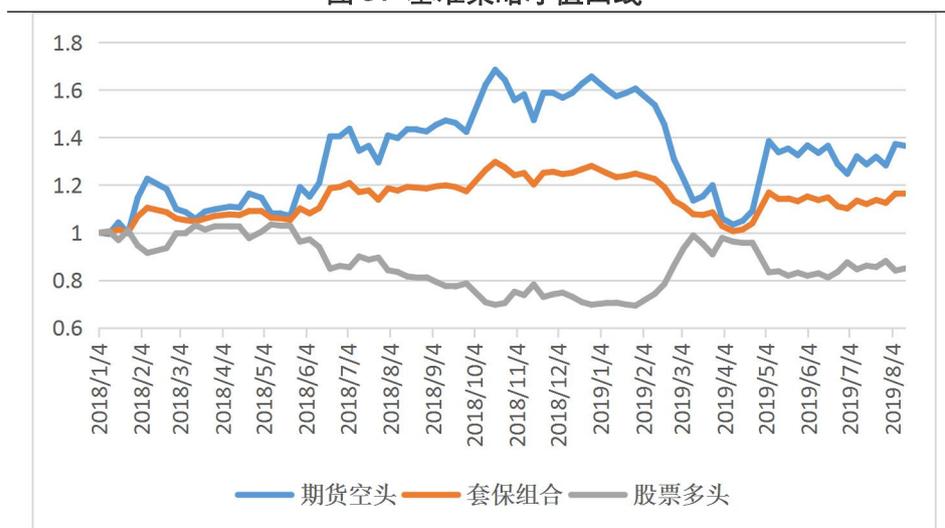
随机森林中包含了一些构建的参数如决策树数量、每个分支节点待选因子数量、因子选择标准（Gini、AIC/BIC、信息熵）、终端节点包含的最小样本数量等。这里采用 10 折的交叉验证（Cross Validation）在训练集中构建子样本，检验不同参数组合下模型的袋外误差率（out-of-bag error rate）并获得最小的一个参数组合。这从一定程度上保证了模型具有一定的拟合程度和泛化能力。

具体交易策略为市值对冲，周频调仓。每周四用模型预测下一周的现货走势。如果已经空开，预测为正号，则平仓；预测为负号，则继续持有（如果一周内到期则先展期到次月）。

## 2.3 回测表现

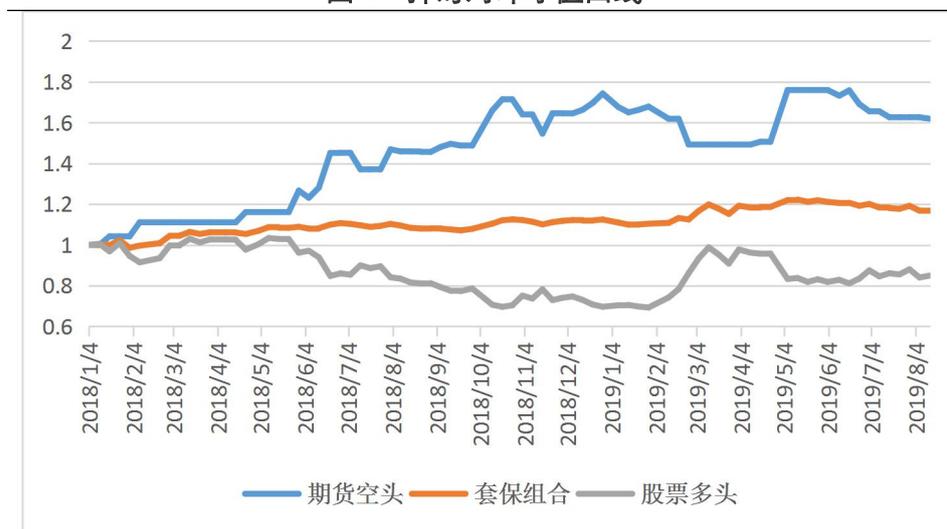
为了尽可能模拟实盘交易行为，回测中我们选取周四 9:30 至 11:30 间每 15 分钟的均价开平仓或展期。其中每手冲击成本 0.03%，手续费 0.002%。初始资金 5000 万，期货账户 1500 万，股票账户 3500 万，20%保证金。也就是在起始日股票账户 3500 万满仓持有现货，期货账户 700 万做保证金，其余为备用金。同时也展示了相同条件下的基准策略，即回测初始时间开始做空期货且之后只进行换月展期。测试集时间 2018 年 1 月至 2019 年 8 月。

图 5：基准策略净值曲线



数据来源：Wind，兴证期货研发部

图 6：择时对冲净值曲线



数据来源：Wind，兴证期货研发部

表 5：基准策略

	累计收益率	年化收益率	最大回撤	夏普比率
期货空头	35.90%	20.32%	39.23%	0.52
套保组合	14.84%	6.18%	21.39%	0.50

数据来源：Wind，兴证期货研发部

表 6：择时对冲策略

	累计收益率	年化收益率	最大回撤	夏普比率	择时胜率	开仓次数
期货空头	61.10%	35.17%	19.01%	1.20	68.45%	19
套保组合	18.92%	9.14%	6.45%	1.14		

数据来源：Wind，兴证期货研发部

可以看出基准策略在没有择时的条件下，面临将近最大 40% 回撤，即便在叠加现货端之后整个仓位的回撤也达到了 21%。相较而言，择时对冲策略期货端最大回撤 19%，整个套保组合的回撤为 6.45%。通过模型预测行情跌幅超过 2% 的准确率有 68%，而且在抵消创业板现货亏损的同时，能保持一定的盈利。

在未来的拓展与应用中，如果择时对冲策略的调仓频率低于周频，那么便可以与上文呈现的 GARCH 模型展期策略结合使用。举例来说，如果预测下月走势有下行风险，可以在信号产生后的 5 天，预测跨月价差并选取价差较优日建仓。

## 分析师承诺

本人以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。报告所采用的数据均来自公开资料，分析逻辑基于本人的职业理解，通过合理判断的得出结论，力求客观、公正，结论，不受任何第三方的授意影响。本人不曾因也将不会因本报告中的具体推荐意见或观点而直接或间接接收到任何形式的报酬。

## 免责声明

本报告的信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。文中的观点、结论和建议仅供参考。兴证期货可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告反映编写分析员的不同设想、见解及分析方法。报告所载资料、意见及推测仅反映分析员于发出此报告日期当日的独立判断。

客户不应视本报告为作出投资决策的惟一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的损失负任何责任。

本报告的观点可能与资管团队的观点不同或对立，对于基于本报告全面或部分做出的交易、结果，不论盈利或亏损，兴证期货研究发展部不承担责任。

本报告版权仅为兴证期货有限公司所有，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，需注明出处兴证期货研究发展部，且不得对本报告进行有悖原意的引用、删节和修改。