

期货基本面量化系列专题（三）：

中州期货研究所 | 深度研究，投研一体

棉花择时策略探索之 XGBoost

概况：

棉花是重要的经济作物与天然纤维来源之一，主要应用于纺织等领域。棉花品种主要包括亚洲棉、草棉、陆地棉、海岛棉等，其中陆地棉是目前种植最为广泛的品种。近年来全球棉花产量保持相对高位，全球消费水平在 2019 年触底后回升明显。

本报告围绕 XGBoost 算法，以棉花的基差、进口利润、全球年产量预测值同比、全国年产量预测值同比、全国年消费量预测值同比、商业库存同比这六个因子为切入点，从单因子筛选、多因子组合两个方面介绍了 XGBoost 在棉花择时策略中的应用及其回测表现。从回测结果来看，由单因子信号组合的 XGBoost 模型虽然夏普比率较低，但在交易胜率、年化收益率、最大回撤等方面表现较为突出。

基于 XGBoost 模型，本报告初步构建了棉花择时框架，后续将逐步探索其他影响因子，并逐步拓展研究品种与研究方法。

风险提示：历史不会完全重演、市场不完全有效、政策等不可预见因素、模型失效等

目录

1. 棉花简介.....	3
1.1 历史背景.....	3
1.2 主要品种.....	3
1.3 产业链概述.....	4
1.4 全球供需概况.....	4
1.5 国内供需概况.....	5
2. 棉花单因子择时策略初探之 XGBoost.....	6
2.1 因子选取思路.....	6
2.2 XGBoost 择时之棉花基差.....	6
2.3 XGBoost 择时之棉花进口利润.....	8
2.4 XGBoost 择时之棉花全球年产量预测值同比.....	9
2.5 XGBoost 择时之棉花全国年产量预测值同比.....	11
2.6 XGBoost 择时之棉花全国年消费量预测值同比.....	12
2.7 XGBoost 择时之棉花商业库存同比.....	13
2.8 单因子总结.....	15
3. 棉花多因子组合策略初探之 XGBoost.....	15
3.1 单因子信号组合之加权组合.....	16
3.2 单因子信号组合之 XGBoost.....	16
3.3 多因子策略之 XGBoost.....	18
4. 小结.....	19

1. 棉花简介

1.1 历史背景

棉花是世界上广泛种植的经济作物之一，也是重要的天然纤维来源，其种植历史最早可追溯至公元前五千至四千年的印度河流域，并于 9 世纪前后传入欧洲。在我国，棉花种植历史可上溯至 2400 多年前的战国时代，其中亚洲棉、草棉分别有两千多年、一千五百多年的历史，陆地棉、海岛棉分别于 19 世纪 60 年代、20 世纪 20 年代引入我国，而新疆开始种植海岛棉则是在 20 世纪 50 年代。

1.2 主要品种

棉花，是锦葵科棉属植物的棉铃里保护种子的絮状纤维，这种纤维经过简单加工后即可用于纺纱织布，是重要的纺织原料之一。野生棉种类繁多，其中有四个品种被人类长期选择、驯化和栽培，分别是亚洲棉、草棉、陆地棉和海岛棉。

亚洲棉，因最早在亚洲种植而得名，又因纤维粗短而被称为**粗绒棉**，是我国利用较早的品种之一。亚洲棉一般为一年生草本，生长期短，色泽呆白，因其纤维粗、长度短、弹性好而被用作起绒纱用棉、医药用药棉、民用絮棉等。但因其产量、纤维长度、细度均不及陆地棉，且只适宜纺 28 号以上（21 英支以下）的中、粗号纱，亚洲棉现已逐渐被陆地棉所取代。

草棉，因原产于非洲而被称为非洲棉，在我国为一年生草本，生长期短，适于我国西北地区栽培，但种植面积不广。草棉株型矮小，棉铃偏小，其纤维短而细，可被用于纺纱、絮衣服被褥等。

陆地棉，原产于中美洲和加勒比海地区，又被称为高原棉、**细绒棉**。陆地棉一般为一年生草本，其植株健壮、生长期中、长、适应性广、结铃性强、铃大、衣分高、皮棉产量高、纤维品质好，是世界上栽培最广的棉种，可用于纺织 60-6 英支的细纱。在上世纪被引入我国后，由于其具有纤维长、韧度强、适合机器纺织等特点，陆地棉已逐步取代原本长期种植的亚洲棉与草棉，成为我国最主要的栽培品种。

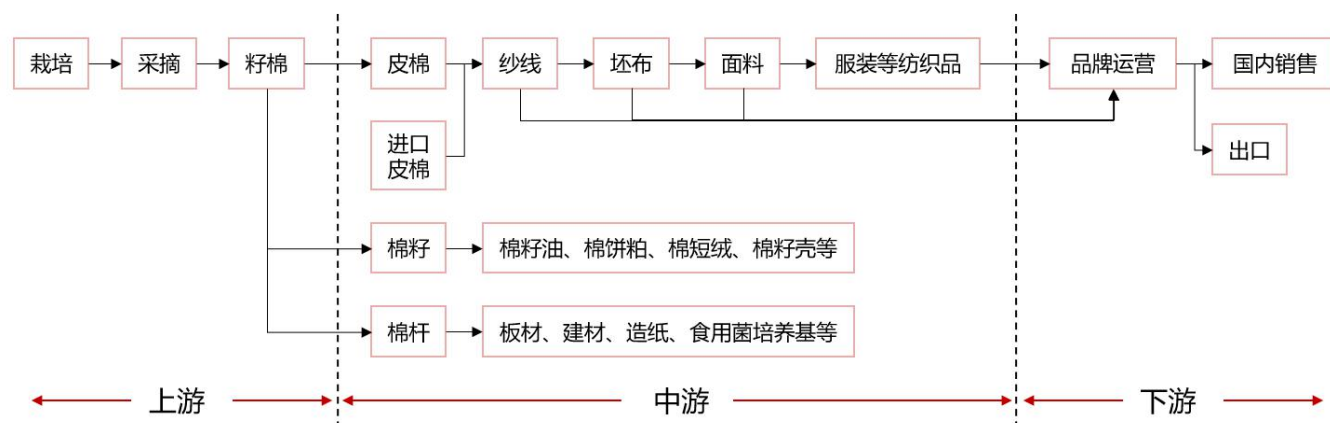
海岛棉，原产于南美洲、中美洲及加勒比地区，在我国又被称为**长绒棉**。海岛棉一般为多年生亚灌木或灌木，其纤维长且柔软、色乳白或淡黄、富有丝光、品质优良，可用于纺织 10 tex 以下的棉纱。根据纤维长度，海岛棉可进一步细分为中长绒棉与特长绒棉。我国新疆地区光照充足、温度较高、气候干燥，是适宜海岛棉种植的地区之一。

据中国农业科学院介绍，目前我国栽培的棉花 90% 以上为陆地棉，其次为海岛棉。郑商所一号棉花期货的基准交割品为符合 GB1103.1-2012《棉花 第 1 部分：锯齿加工细绒棉》规定的 3128B 级，且长度整齐度为 U3 档，断裂比强度为 S3 档，轧工质量为 P2 档的国产棉花。

1.3 产业链概述

我国棉花产业链较为完整，上游环节包括棉花栽培、采摘、籽棉交售等，中游环节包括籽棉加工、流通、纺纱、织布、印染、制衣等，下游环节包括纺织品贸易商的运营、销售等。

图表 1：棉花上下游产业链



来源：《棉花产业链各环节相关性分析》，中州期货

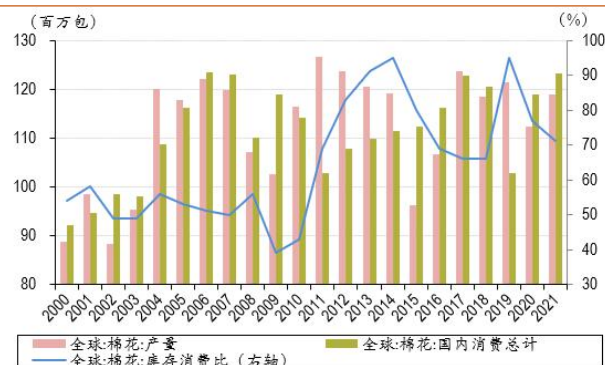
1.4 全球供需概况

近年来全球棉花产量保持相对高位，全球国内消费量在 2019 年触底后回升明显。具体来看，2020 年受疫情影响，全球棉花产量同比下滑 7.4%，但据美国农业部 (USDA) 预计，2021 年有望实现同比增长 5.7%。

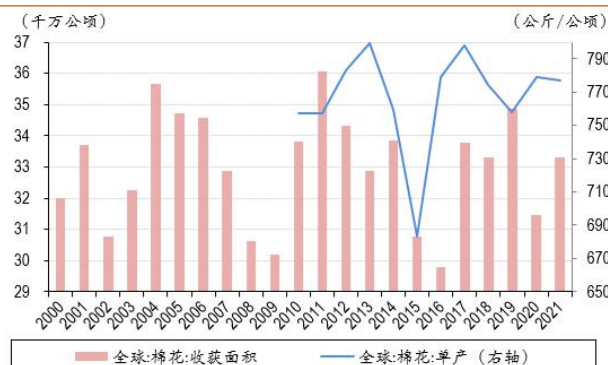
棉花的国内消费方面，2019 年同比下滑 14.6%至 2004 年以来的最低水平，随后 2020 年实现增长 15.7%，据 USDA 预计，2021 年有望进一步增长至 123.33 百万包，达到 2007 年以来的最高水平。值得注意的是，2020 与 2021 年全球棉花的产量均不及国内消费量，分别存在 6.59、4.49 百万包的缺口。全球棉花收获面积在 2020 年回落后有望于 2021 年迎来回升，而单产情况则在 2020 年回升后保持相对稳定。

图表 2：全球棉花产量与国内消费情况

图表 3：全球棉花收获面积与单产情况



来源：Wind，中州期货



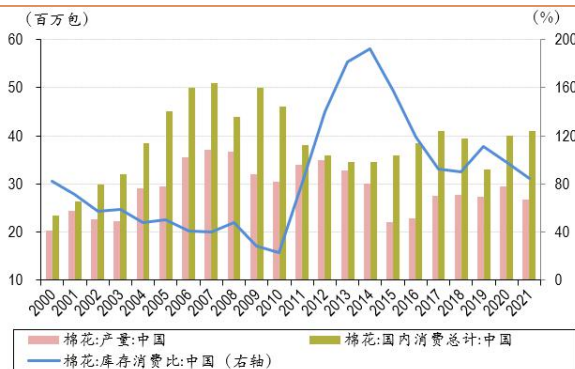
来源：Wind，中州期货

1.5 国内供需概况

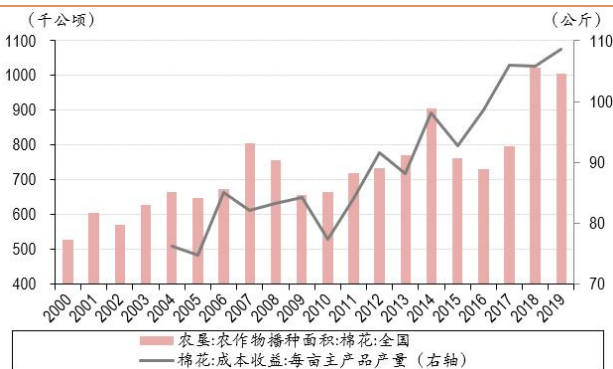
近两年我国棉花产量波动较大，国内消费量在 2019 年触底后回升。具体来看，2020 年虽面临疫情影响，我国棉花产量仍同比增长 8.3%，但 2021 年预计将同比下滑 9.3%。国内消费方面，2019 年同比下滑 16.5%至 2004 年以来的最低水平，随后 2020 年增长 21.2%，2021 年有望进一步增长至 41.00 百万包。2021 年我国棉花产量与国内消费缺口或将进一步加大，较 2020 年增加 4.25 百万包至 14.25 百万包。我国棉花播种面积在 2018 年增加较多，亩产情况则在 2010 年以来保持相对稳定的上升态势。

图表 4：我国棉花产量与国内消费情况

图表 5：我国棉花播种面积与亩产情况



来源：Wind，中州期货



来源：Wind，中州期货

2. 棉花单因子择时策略初探之 XGBoost

2.1 因子选取思路

从研究框架的角度出发，对品种的关注点大致可分为估值与驱动两个方面。其中，估值可进一步分为绝对价格与相对价格两方面，前者包括现货价格、期货价格等，而后者则包括下文将会分析的基差因子与进口利润因子。驱动可分为供应、库存、需求三方面，分别对应产量、库存、消费量等因子。

图表 6：因子选取思路



来源：中州期货

2.2 XGBoost 择时之棉花基差

此处所用的棉花基差计算方式为：

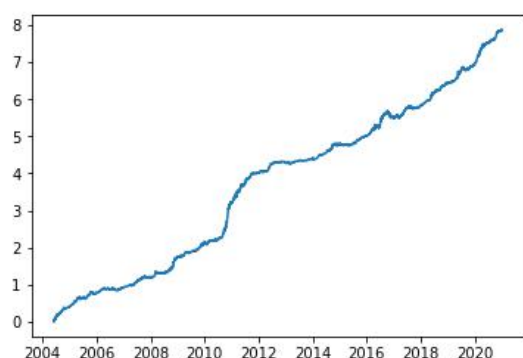
$$\text{棉花基差} = \text{中国棉花价格指数:328} - \text{期货收盘价(连续):棉花}$$

在进行数据处理后，考虑到现货价与收盘价获取的滞后性，这里将基差结果后移一日，即将基差应用于次日。故模型输入为日期、后移一日的基差，输出为 n 个交易日后 Wind 郑棉指数 (CFFI.WI) 较该日的涨跌方向，若上涨或持平则记为 +1，若下跌则记为 0。这里将 2004 年 6 月 2 日至 2020 年 12 月 31 日 3932 个样本作为训练集，将 2021 年 1 月 4 日至 9 月 3 日的 165 个样本作为测试集，并以 Wind 郑棉指数作为收益

率计算因子。

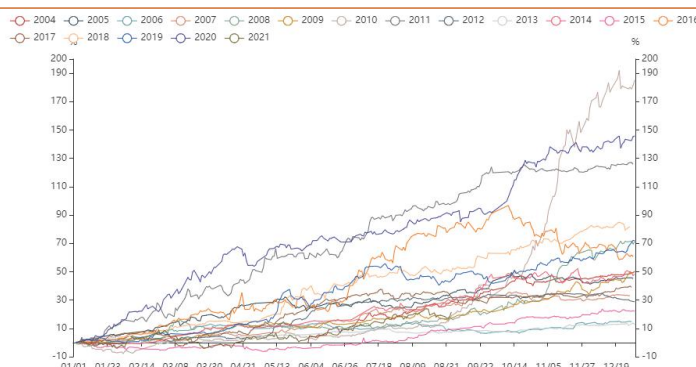
经过比较，这里将以间隔 2 个交易日，即 $n=2$ 的 Wind 郑棉指数涨跌方向为预测因子来进行择时策略检验。在模型训练中，为进一步降低过度拟合的可能性，设置参数 `'max_depth': 5`、`'objective': 'multi: softmax'`、`'num_class': 2`、`'subsample': 0.7`、`num_rounds = 20`。仓位变动标准为，若信号等于 1 则多头开仓或持有，若等于 0 则空头开仓或持有。从回测表现来看，训练集内，净值从 1 逐步增长至 2620.501，交易胜率为 64%，复合年增长率为 60.7%，夏普比率为 5.381，最大回撤为 20%，年均交易 43 次。测试集检验方面，净值从 1 逐步增长至 1.214，交易胜率为 75%，复合年增长率为 33.5%，夏普比率为 2.132，最大回撤为 12%，年均交易 12 次。分年度来看，2010 年该策略收益率表现最为突出，为 185%，其次为 2020 年（146%）与 2011 年（126%），余下各年收益率均位于 13%至 82%范围内。

图表 7：棉花基差策略训练集净值的自然对数走势



来源：Python，中州期货

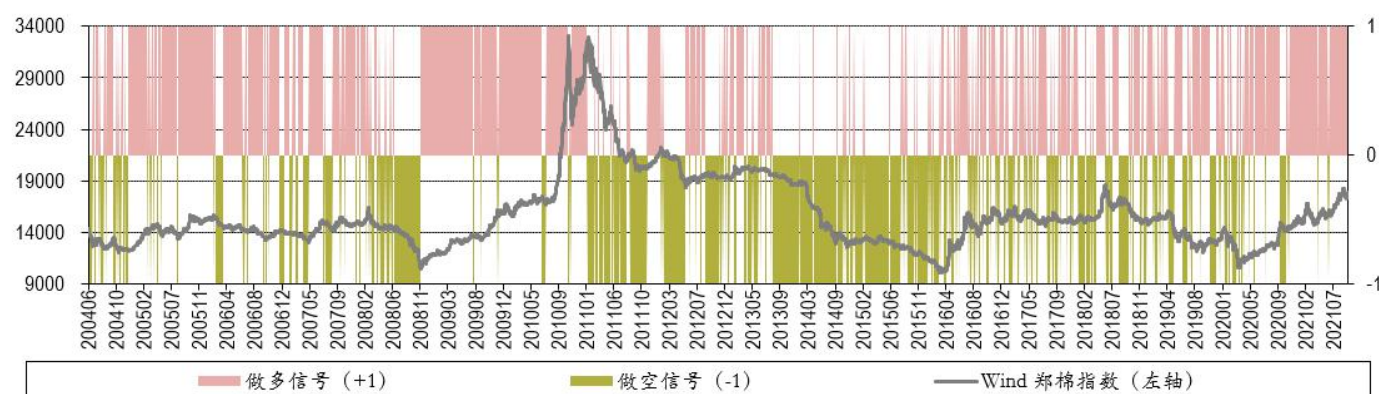
图表 8：棉花基差策略各年收益表现



来源：Python，中州期货

将策略信号与 Wind 郑棉指数走势相比较可以看出，该棉花基差策略在捕捉指数的短期趋势方面展现出了一定的效果。

图表 9: Wind 郑棉指数走势与棉花基差的 XGBoost 策略信号比较



来源：中州期货

2.3 XGBoost 择时之棉花进口利润

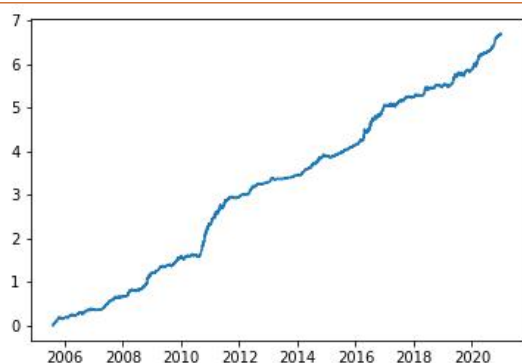
此处所用的棉花进口利润计算方式为：

$$\text{棉花进口利润} = \text{中国棉花价格指数:328} - \text{进口棉价格指数(FC Index):M:滑准税港口提货价}$$

对数据进行处理后，同样再将进口利润结果后移一日，模型输入为**日期、后移一日的进口利润**，输出为 **n 个交易日后 Wind 郑棉指数较该日的涨跌方向**，同样记为+1 或 0。这里将 2005 年 8 月 2 日至 2020 年 12 月 31 日 3751 个样本作为训练集，将 2021 年 1 月 4 日至 9 月 3 日的 165 个样本作为测试集，并以 **Wind 郑棉指数作为收益率计算因子**。

经过比较，这里将以间隔 7 个交易日，即 $n=7$ 的 Wind 郑棉指数涨跌方向为预测因子来进行择时策略检验。模型参数设置、仓位变动标准等均同上。从回测表现来看，训练集内，净值从 1 逐步增长至 804.198，交易胜率为 63%，复合年增长率为 54.3%，夏普比率为 5.233，最大回撤为 9%，年均交易 23 次。测试集检验方面，净值从 1 逐步增长至 1.213，交易胜率为 75%，复合年增长率为 33.8%，夏普比率为 2.131，最大回撤为 12%，年均交易 6 次。分年度来看，2016 年该策略收益率表现最为突出，为 144%，其次为 2020 年（120%）与 2010 年（107%），余下各年收益率均位于 18%至 88%之间。

图表 10: 棉花进口利润策略训练集净值自然对数走势



来源: Python, 中州期货

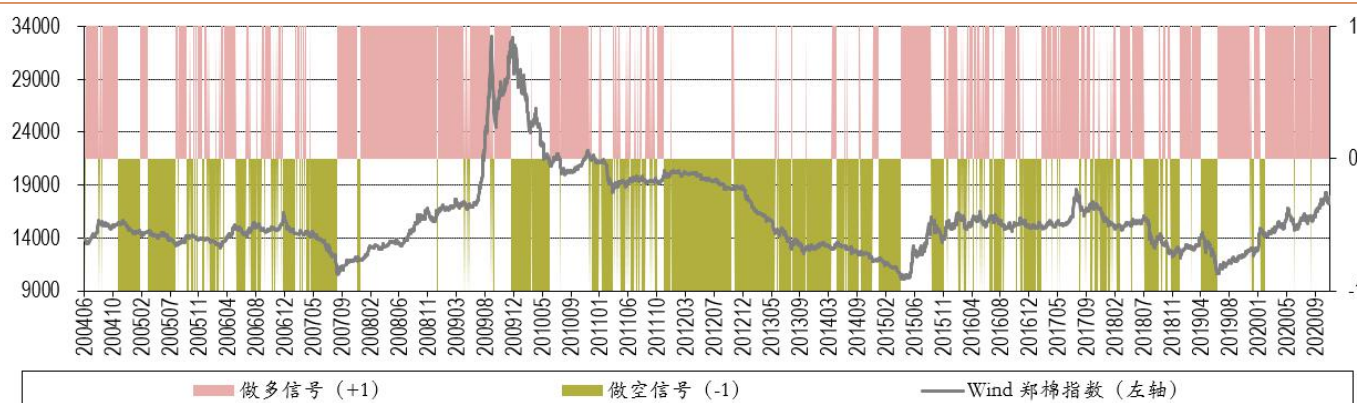
图表 11: 棉花进口利润策略各年收益表现



来源: Python, 中州期货

将策略信号与 Wind 郑棉指数走势相比较可以看出, 该进口利润策略在捕捉指数的中短期趋势性波动方面展现出了一定的效果。

图表 12: Wind 郑棉指数走势与棉花进口利润的 XGBoost 策略信号比较



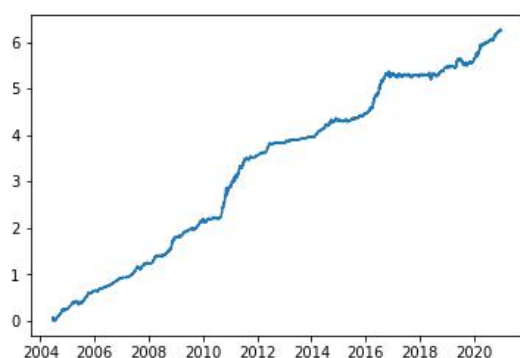
来源: 中州期货

2.4 XGBoost 择时之棉花全球年产量预测值同比

此处所用的棉花全球年产量预测值 (全球:产量:市场年度:棉花) 为美国农业部 (USDA) 月度更新的预测数据。假设当月的年产量预测数据于该月 14 日更新, 故将该月预测数据的同比应用于当月 14 日至下月 13 日。对数据进行处理后, 模型输入为**日期**、**后移的棉花全球年产量预测值同比**, 输出为**n 日后 Wind 郑棉指数较该日的涨跌方向**, 同样记为+1 或 0。这里将 2004 年 6 月 14 日至 2020 年 12 月 31 日 4029 个样本作为训练集, 将 2021 年 1 月 4 日至 9 月 3 日的 165 个样本作为测试集。

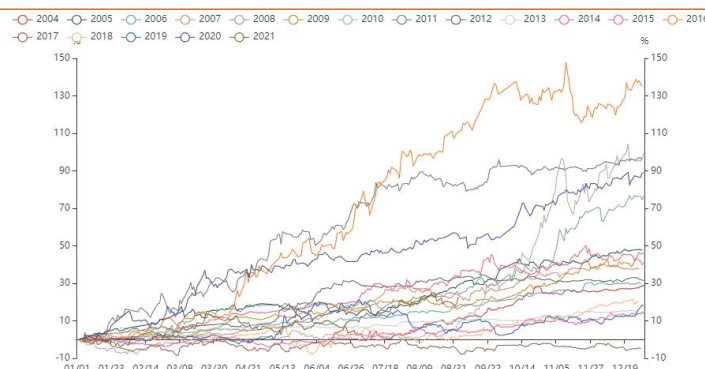
经过比较，这里将以间隔 6 个交易日，即 $n=6$ 的 Wind 郑棉指数涨跌方向为预测因子来进行择时策略检验。模型参数设置、仓位变动标准、收益率指标等均同上。从回测表现来看，训练集内，净值从 1 逐步增长至 524.566，交易胜率为 89%，复合年增长率为 45.97%，夏普比率为 4.574，最大回撤为 17%，年均交易 4 次。测试集检验方面，净值从 1 逐步增长至 1.226，交易胜率为 50%，复合年增长率为 35.98%，夏普比率为 2.431，最大回撤为 10%，年均交易 3 次。分年度来看，2016 年该策略收益率表现最为突出，为 135%，其次为 2010 年（99%）与 2011 年（96%）。值得注意的是，在训练集回测中，2017 年该策略收益为 -5%，表现不佳。

图表 13：棉花全球年产量预测值同比策略训练集净值的自然对数走势



来源：Python，中州期货

图表 14：棉花全球年产量预测值同比策略各年收益表现



来源：Python，中州期货

将策略信号与 Wind 郑棉指数走势相比较可以看出，该全球年产量预测值同比策略在捕捉指数的中长期走势上展现出了一定的效果。

图表 15：Wind 郑棉指数走势与棉花全球年产量预测值同比的 XGBoost 策略信号比较



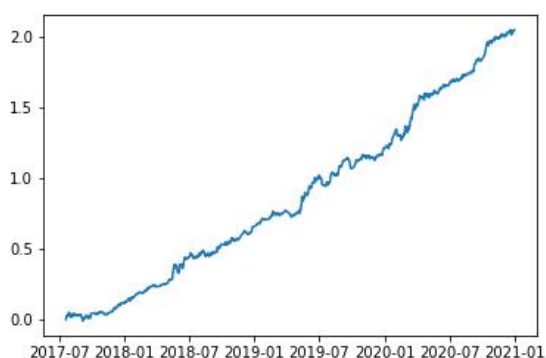
来源：中州期货

2.5 XGBoost 择时之棉花全国年产量预测值同比

此处所用的棉花全国年产量预测值（产量:棉花:年度预测值）为中国农业信息网月度更新的预测数据。假设当月的年产量预测数据于该月 20 日更新，故将该月预测数据的同比应用于当月 20 日至下月 19 日。对数据进行处理后，模型输入为**日期、后移的棉花全国年产量预测值同比**，输出为**n 日后 Wind 郑棉指数较该日的涨跌方向**，同样记为+1 或 0。这里将 2017 年 7 月 20 日至 2020 年 12 月 31 日 842 个样本作为训练集，将 2021 年 1 月 4 日至 9 月 3 日的 165 个样本作为测试集。

经过比较，这里将以间隔 4 个交易日，即 $n=4$ 的 Wind 郑棉指数涨跌方向为预测因子来进行择时策略检验。模型参数设置、仓位变动标准、收益率指标等均同上。从回测表现来看，训练集内，净值从 1 逐步增长至 7.739，交易胜率为 97%，复合年增长率为 80.9%，夏普比率为 3.567，最大回撤为 8%，年均交易 9 次。测试集内，仓位一直为多头且未发生变动，净值从 1 逐步增长至 1.165，复合年增长率为 25.9%，夏普比率为 1.956，最大回撤为 12%。分年度来看，2020 年该策略收益率表现最为突出，为 131%，其次为 2019 年（74%）与 2018 年（71%）。

图表 16：棉花全国年产量预测值同比策略训练集净值的自然对数走势



来源：Python，中州期货

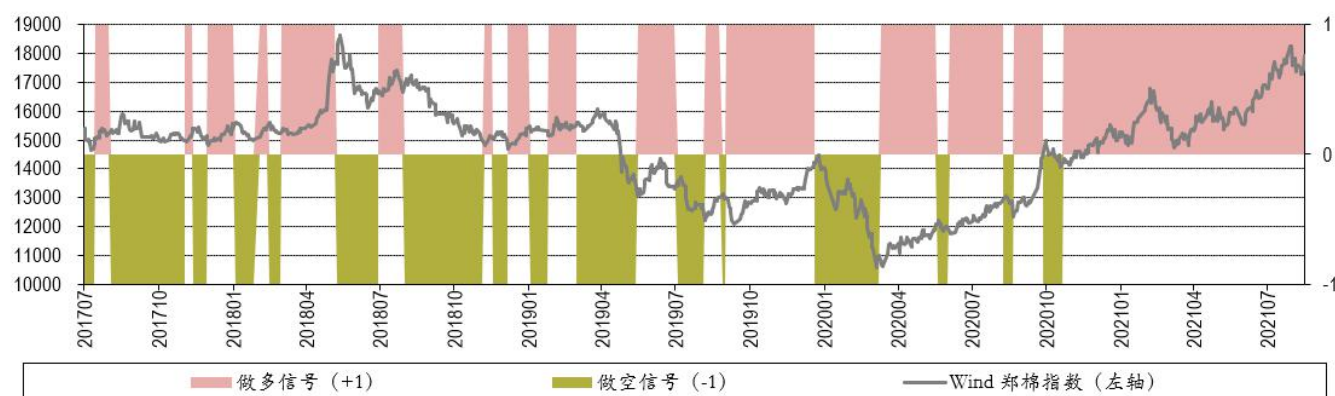
图表 17：棉花全国年产量预测值同比策略各年收益表现



来源：Python，中州期货

将策略信号与 Wind 郑棉指数走势相比较可以看出，该全国年产量预测值同比策略在捕捉指数的长期走势上存在一定的效果。

图表 18: Wind 郑棉指数走势与棉花全国年产量预测值同比的 XGBoost 策略信号比较



来源：中州期货

2.6 XGBoost 择时之棉花全国年消费量预测值同比

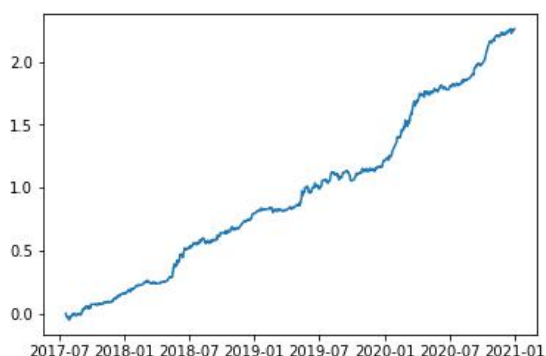
此处所用的棉花全国年消费量预测值 (消费量:棉花:年度预测值) 为中国农业信息网月度更新的预测数据。

假设当月的年消费量预测数据于该月 20 日更新, 故将该月预测数据的同比应用于当月 20 日至下月 19 日。

对数据进行处理后, 模型输入为**日期**、**后移的棉花全国年消费量预测值同比**, 输出为 **n 日后 Wind 郑棉指数较该日的涨跌方向**, 同样记为+1 或 0。这里将 2017 年 7 月 20 日至 2020 年 12 月 31 日 842 个样本作为训练集, 将 2021 年 1 月 4 日至 9 月 3 日的 165 个样本作为测试集。

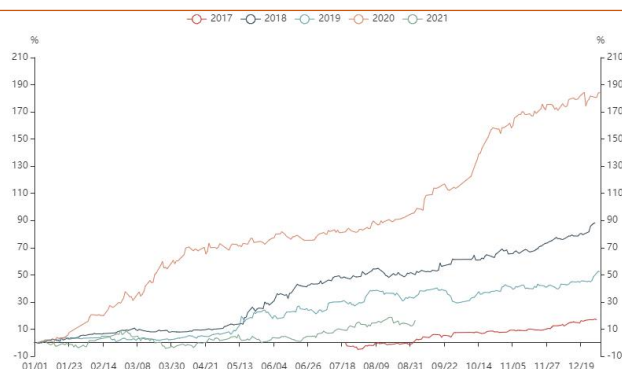
经过比较, 这里将以间隔 5 个交易日, 即 $n=5$ 的 Wind 郑棉指数涨跌方向为预测因子来进行择时策略检验。模型参数设置、仓位变动标准、收益率指标等均同上。从回测表现来看, 训练集内, 净值从 1 逐步增长至 9.594, 交易胜率为 93%, 复合年增长率为 92.52%, 夏普比率为 3.701, 最大回撤为 8%, 年均交易 9 次。测试集内, 仓位一直为多头且未发生变动, 净值从 1 逐步增长至 1.165, 复合年增长率为 25.9%, 夏普比率为 1.956, 最大回撤为 12%。分年度来看, 2020 年该策略收益率表现最为突出, 为 184%, 其次为 2018 年 (88%) 与 2019 年 (52%)。

图表 19：棉花全国年消费量预测值同比策略训练集净值的自然对数走势



来源：Python，中州期货

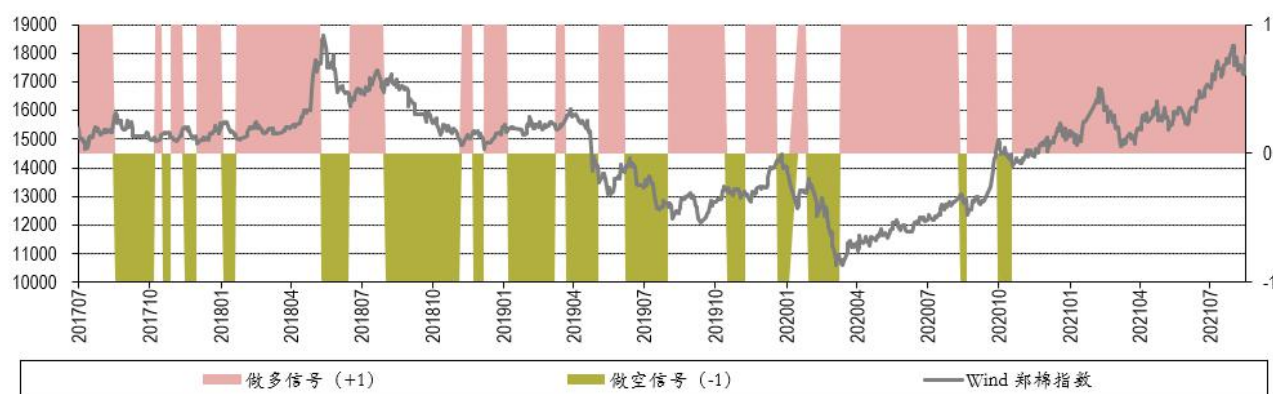
图表 20：棉花全国年消费量预测值同比策略各年收益表现



来源：Python，中州期货

将策略信号与 Wind 郑棉指数走势相比较可以看出，该全国年消费量预测值同比策略同样在捕捉指数的长期走势上存在一定的效果。

图表 21：Wind 郑棉指数走势与棉花全国年消费量预测值同比的 XGBoost 策略信号比较



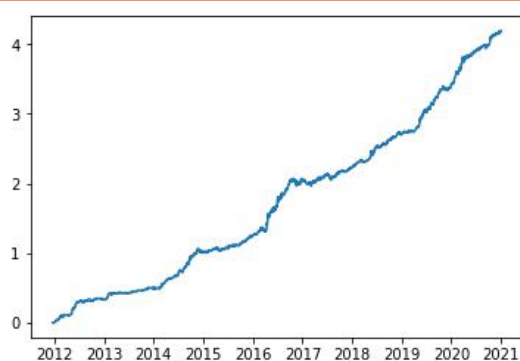
来源：中州期货

2.7 XGBoost 择时之棉花商业库存同比

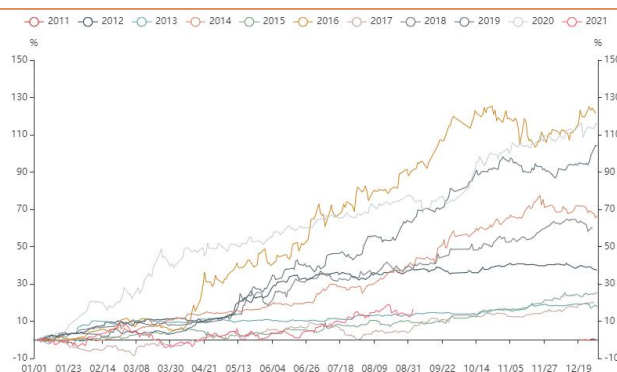
此处所用的棉花商业库存（商业库存:棉花）为中国棉花信息网月度公布的数据。假设当月的库存数据于次月 20 日更新，故将该月的库存同比应用于数据公布日（次月 20 日）至次月数据的公布日（往后第二个月 20 日）之间。对数据进行处理后，模型输入为**日期、后移的棉花商业库存同比**，输出为**n 日后 Wind 郑棉指数较该日的涨跌方向**，同样记为+1 或 0。这里将 2011 年 12 月 20 日至 2020 年 12 月 31 日 2197 个样本作为训练集，将 2021 年 1 月 4 日至 9 月 3 日的 165 个样本作为测试集。

经过比较，这里将以间隔 6 个交易日，即 $n=6$ 的 Wind 郑棉指数涨跌方向为预测因子来进行择时策略检验。模型参数设置、仓位变动标准、收益率指标等均同上。从回测表现来看，训练集内，净值从 1 逐步增长至 66.145，交易胜率为 96%，复合年增长率为 59.01%，夏普比率为 4.264，最大回撤为 11%，年均交易 6 次。测试集内，仓位一直为多头且未发生变动，净值从 1 逐步增长至 1.165，复合年增长率为 25.9%，夏普比率为 1.956，最大回撤为 12%。分年度来看，2016 年该策略收益率表现最为突出，为 122%，其次为 2020 年 (116%) 与 2019 年 (104%)。

图表 22：棉花商业库存同比策略训练集净值的自然对数走势 图表 23：棉花商业库存同比策略各年收益表现



来源：Python，中州期货



来源：Python，中州期货

将策略信号与 Wind 郑棉指数走势相比较可以看出，该商业库存同比策略同样在捕捉指数的中长期走势上存在一定的效果。

图表 24：Wind 郑棉指数走势与棉花商业库存同比的 XGBoost 策略信号比较



来源：中州期货

2.8 单因子总结

回顾这六个因子各自的择时策略表现，可以看出 XGBoost 模型对棉花单因子择时策略的构建存在一定的参考价值，后文将继续探索 XGBoost 对多因子组合择时策略的构建。

图表 25：前文 XGBoost 各单因子择时策略训练集回测表现

因子	预测间隔天数	回测表现							
		回测时间范围	回测天数	最终净值	交易胜率	年化收益率	夏普比率	最大回撤	年均交易频率
棉花基差	2	2004/06/02-2020/12/31	3932	2621	64%	61%	5.381	20%	43
棉花进口利润	7	2005/08/02-2020/12/31	3751	804	63%	54%	5.233	9%	23
棉花全球年产量预测值同比	6	2004/06/14-2020/12/31	4029	525	89%	46%	4.574	17%	4
棉花全国年产量预测值同比	4	2017/07/20-2020/12/31	842	7.739	97%	81%	3.567	8%	9
棉花全国年消费量预测值同比	5	2017/07/20-2020/12/31	842	9.594	93%	93%	3.701	8%	9
棉花商业库存同比	6	2011/12/20-2020/12/31	2197	66	96%	59%	4.264	11%	6

来源：中州期货

图表 26：前文 XGBoost 各单因子择时策略测试集检验表现

因子	预测间隔天数	测试集检验表现							
		回测时间范围	回测天数	最终净值	交易胜率	年化收益率	夏普比率	最大回撤	年均交易频率
棉花基差	2	2021/01/04-2021/09/03	165	1.214	75%	34%	2.132	12%	12
棉花进口利润	7	2021/01/04-2021/09/03	165	1.213	75%	34%	2.131	12%	6
棉花全球年产量预测值同比	6	2021/01/04-2021/09/03	165	1.226	50%	36%	2.431	10%	3
棉花全国年产量预测值同比	4	2021/01/04-2021/09/03	165	1.165	-	26%	1.956	12%	-
棉花全国年消费量预测值同比	5	2021/01/04-2021/09/03	165	1.165	-	26%	1.956	12%	-
棉花商业库存同比	6	2021/01/04-2021/09/03	165	1.165	-	26%	1.956	12%	-

注：后三个因子在测试集中一直持有多头仓位

来源：中州期货

3. 棉花多因子组合策略初探之 XGBoost

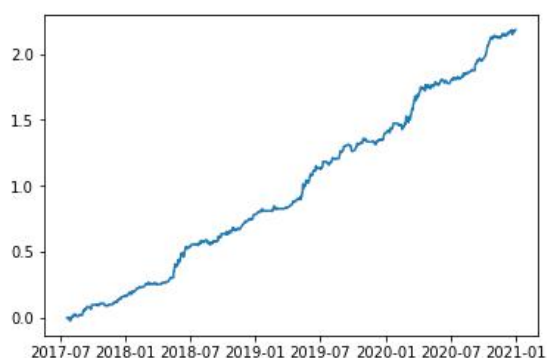
这里主要尝试三种方式来探索 XGBoost 模型对棉花多因子组合择时策略的构建，一是将上文六个单因子策略的择时信号进行**加权组合**进而得到新的信号，二是将**六个因子的信号作为输入**来构建新的 XGBoost 模型，三是将**六个因子作为输入**来构建新的 XGBoost 模型。

3.1 单因子信号组合之加权组合

此部分的单因子策略均与上文一致，这里将其产生的六个择时信号进行加权组合，即对基差信号、进口利润信号、全球年产量预测值同比信号、全国年产量预测值同比信号、全国年消费量预测值同比与商业库存同比信号分别赋予 $1/6$ 的权重来进行组合信号的生成，在此基础上以组合信号大于 0 为多头开仓与持仓标准、小于 0 为空头开仓与持仓标准。

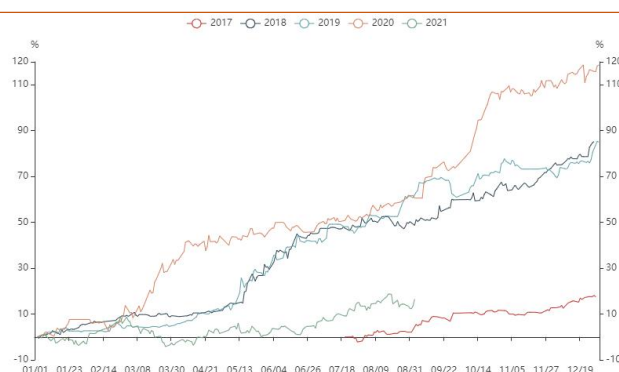
从回测表现来看，训练集（2017 年 7 月 20 日至 2020 年 12 月 31 日）内，净值从 1 逐步增长至 8.862，交易胜率为 42%，复合年增长率为 88.14%，夏普比率为 3.537，最大回撤为 5%，年均交易 29 次。测试集（2021 年 1 月 4 日至 2020 年 9 月 3 日）内，仓位一直为多头且未发生变动，净值从 1 逐步增长至 1.165，复合年增长率为 25.9%，夏普比率为 1.956，最大回撤为 12%。分年度来看，2020 年该策略收益率表现最为突出，为 119%，其次为 2019 年（85%）与 2018 年（85%）。与单因子策略相比，该信号组合方法的交易胜率、夏普比率最低，但其年化收益率、最大回撤表现较为突出，赋予各因子的权重以及新信号的判断标准或还有进一步尝试的空间。

图表 27：六因子信号线性组合后的训练集净值自然对数走势



来源：Python，中州期货

图表 28：六因子信号线性组合后新信号的各年收益表现



来源：Python，中州期货

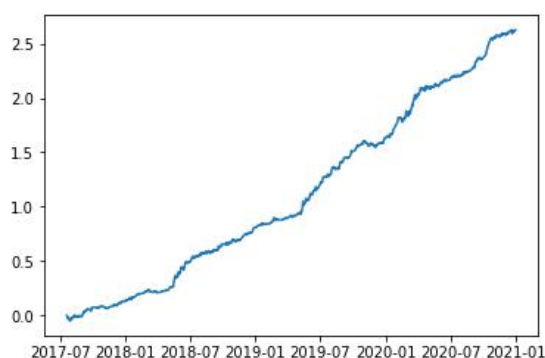
3.2 单因子信号组合之 XGBoost

此部分的单因子策略均与上文一致，这里将其产生的六个信号输入新的 XGBoost 模型进行训练并生成新的择时信号。经过比较，这里将以间隔 6 个交易日，即 $n=6$ 的 Wind 郑棉指数涨跌方向为预测因子来进行择

时策略检验。模型参数设置、仓位变动标准、收益率指标等均同上。

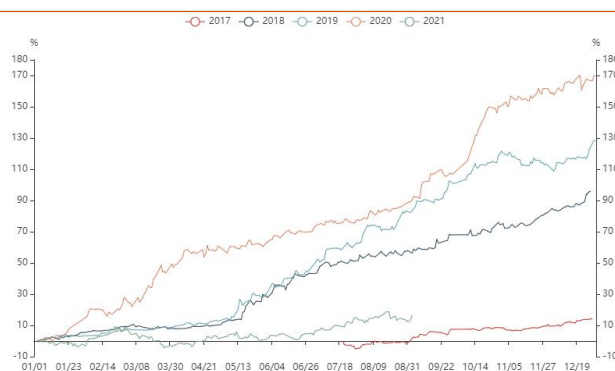
从回测表现来看, 训练集 (2017 年 7 月 20 日至 2020 年 12 月 31 日) 内, 净值从 1 逐步增长至 13.835, 交易胜率为 91%, 复合年增长率为 114.05%, 夏普比率为 3.519, 最大回撤为 6%, 年均交易 13 次。测试集 (2021 年 1 月 4 日至 2020 年 9 月 3 日) 内, 仓位一直为多头且未发生变动, 净值从 1 逐步增长至 1.165, 复合年增长率为 25.9%, 夏普比率为 1.956, 最大回撤为 12%。分年度来看, 2020 年该策略收益率表现最为突出, 为 170%, 其次为 2019 年 (128%) 与 2018 年 (96%)。与单因子策略相比, 该信号组合方法的夏普比率最低, 但其交易胜率、年化收益率、最大回撤表现较为突出, 其表现比 3.1 中的信号加权组合亦有所提升。

图表 29: 六因子信号线性组合后的训练集净值自然对数走势



来源: Python, 中州期货

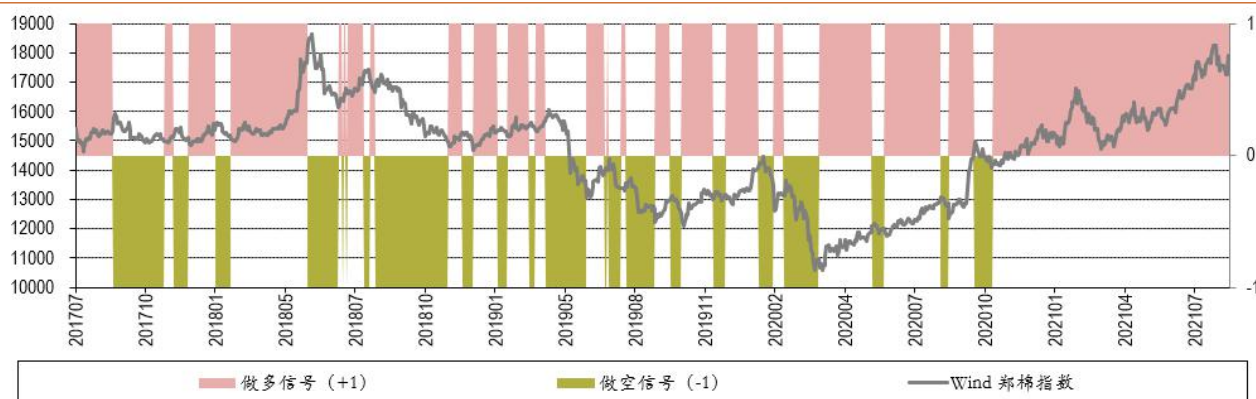
图表 30: 六因子信号线性组合后新信号的各年收益表现



来源: Python, 中州期货

将策略信号与 Wind 郑棉指数走势相比较可以看出, 该组合策略同样在捕捉指数的中长期走势上存在一定的效果。

图表 31: Wind 郑棉指数走势与该 XGBoost 策略信号比较



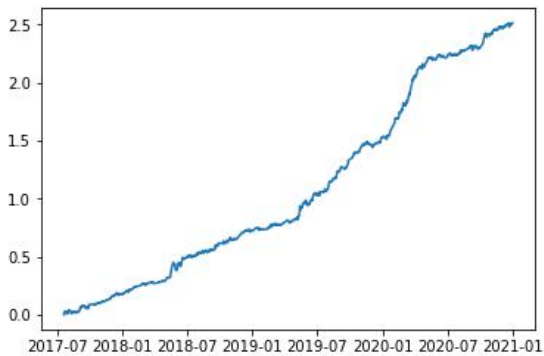
来源：中州期货

3.3 多因子策略之 XGBoost

这里将处理并后移的棉花基差、进口利润、全球年产量预测值同比、全国年产量预测值同比、全国年消费量预测值同比、商业库存同比及日期作为输入， n 日后 Wind 郑棉指数较该日的涨跌方向（上涨或持平为 1，下跌为 0）作为输出，来构建新的 XGBoost 模型。经过循环遍历与比较，这里将以间隔 5 个交易日，即 $n=5$ 的 Wind 郑棉指数涨跌方向为预测因子来进行择时策略检验，模型参数不变，仓位变动标准与收益率计算因子同上。

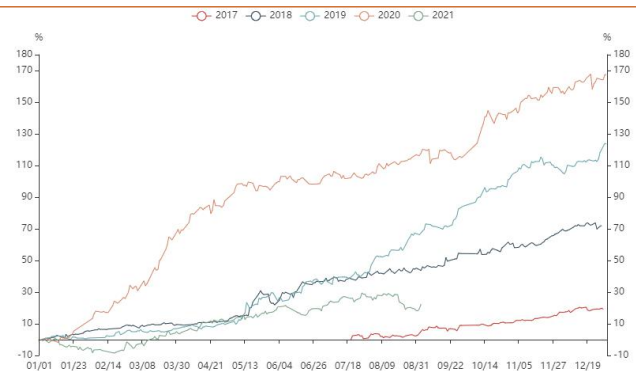
从回测表现来看，训练集（2017 年 7 月 20 日至 2020 年 12 月 31 日）内，净值从 1 逐步增长至 12.360，交易胜率为 76%，复合年增长率为 107.18%，夏普比率为 3.265，最大回撤为 7%，年均交易 35 次。测试集（2021 年 1 月 4 日至 2020 年 9 月 3 日）内，净值从 1 逐步增长至 1.225，交易胜率为 57%，复合年增长率为 35.81%，夏普比率为 1.485，最大回撤为 9%，年均交易 91 次。分年度来看，2020 年该策略收益率表现最为突出，为 168%，其次为 2019 年（124%）与 2018 年（72%）。与单因子策略相比，该信号组合方法的夏普比率、交易胜率最低，但其年化收益率、最大回撤表现较为突出。

图表 32：多因子 XGBoost 策略净值的自然对数走势



来源：Python，中州期货

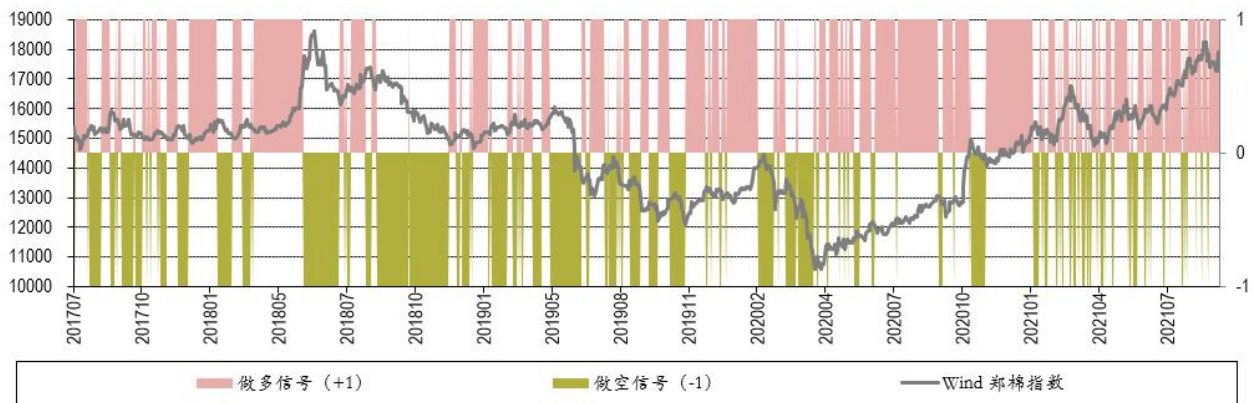
图表 33：棉花多因子 XGBoost 模型的各年收益表现



来源：Python，中州期货

将策略信号与 Wind 郑棉指数走势相比较可以看出，该多因子策略主要捕捉指数的中短期波动并取得一定效果。

图表 34：Wind 郑棉指数走势与该多因子 XGBoost 策略信号比较



来源：中州期货

4. 小结

这里对上述九个择时信号生成方法进行整体回顾。就回测表现来看，虽然单因子信号组合的 XGBoost 模型（下表⑧）的夏普比例较低，但其在交易胜率、年化收益率、最大回撤上具有明显优势。

图表 35：六个择时策略的回测表现比较

指标	预测间隔天数	回测表现							
		回测时间范围	回测天数	最终净值	交易胜率	年化收益率	夏普比率	最大回撤	年均交易频率
① 棉花基差	2	2004/06/02-2020/12/31	3932	2621	64%	61%	5.381	20%	43
② 棉花进口利润	7	2005/08/02-2020/12/31	3751	804	63%	54%	5.233	9%	23
③ 棉花全球年产量预测值同比	6	2004/06/14-2020/12/31	4029	525	89%	46%	4.574	17%	4
④ 棉花全国年产量预测值同比	4	2017/07/20-2020/12/31	842	7.739	97%	81%	3.567	8%	9
⑤ 棉花全国年消费量预测值同比	5	2017/07/20-2020/12/31	842	9.594	93%	93%	3.701	8%	9
⑥ 棉花商业库存同比	6	2011/12/20-2020/12/31	2197	66	96%	59%	4.264	11%	6
⑦ 单因子信号组合之加权组合	-	2017/07/20-2020/12/31	842	8.862	42%	88%	3.537	5%	29
⑧ 单因子信号组合之XGBoost	6	2017/07/20-2020/12/31	842	13.835	91%	114%	3.519	6%	13
⑨ 多因子策略之XGBoost	5	2017/07/20-2020/12/31	842	12.360	76%	107%	3.265	7%	35

来源：中州期货

本报告围绕 XGBoost 算法，以棉花的基差、进口利润、全球年产量预测值同比、全国年产量预测值同比、全国年消费量预测值同比、商业库存同比这六个因子为切入点，从单因子筛选、多因子组合两个方面介绍了 XGBoost 在棉花择时策略中的应用及其回测表现。基于 XGBoost 模型，本报告初步构建了棉花择时框架，后续将逐步探索其他影响因子，并逐步拓展研究品种与研究方法。

■ 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做任何保证，也不保证所包含的信息和建议不会发生任何变更。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，报告中的信息或意见并不构成所述证券或期货的买卖出价或征价，据此作出的任何投资决策与本公司和作者无关。

本报告版权仅为我公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制发布。如引用、刊发，须注明出处为中州期货公司，且不得对本报告进行有悖原意的引用、删节和修改。

■ 联系我们

管理总部 200120

地址：上海市浦东新区民生路 1299 号丁香国际商业中心西楼 1502 室

Tel: 021-68395023

Email: zzzjb@zzfco.net

烟台营业部 264000

地址：烟台市芝罘区西关南街 2 号万达金融中心 20 层

Tel: 0535-6695733 传真: 0535-6692349

Email: zzyt@zzfco.net

青岛营业部 266071

地址：青岛市南京路 9 号联合大厦 19 层

Tel: 0532-85753103 传真: 0532-85753173

Email: zzqd@zzfco.net

临沂营业部 276004

地址：临沂市兰山区沂蒙路 454 号华泰(嘉锐)大厦 801

Tel: 0539-8053280 传真: 0539-8053280

Email: zzly@zzfco.net

大连营业部 116023

地址：大连市沙河口区会展路 129 号大连国际金融中心 A 座 2208 室

Tel: 0411-84806638 传真: 0411-84806338

济南营业部 250000

地址：济南市市中区阳光新路 67 号济南欧亚大观商都 A1-1105

Tel: 0531-83165837 传真: 0531-83165836

Email: zzjn@zzfco.net

广西分公司 116023

地址：广西南宁五象新区五象大道 403 号富雅国际金融中心 6512-6513 室

Tel: 0771-5505082 Email: zzgx@zzfco.net

烟台总部 264010

地址：山东省烟台市芝罘区万达金融中心 A 座 20 楼

Tel: 0535-6692646

Email: zzqh@zzfco.net

上海分公司 200120

上海市浦东新区民生路 1299 号丁香国际商业中心西楼 1503 室

Tel: 021-68580502 Email: zzsh@zzfco.net

沈阳营业部 110004

地址：沈阳市和平区三好街 100-4 号 702-704 室

Tel: 024-31258299 传真: 024-31258292

Email: zzsy@zzfco.net

烟台分公司 265701

地址：烟台市开发区长江路天马中心 1 号楼 23 层 2301 号房

Tel: 0535-2169566 传真: 0535-2169566

Email: zzlk@zzfco.net

江苏分公司 210019

地址：南京市建邺区兴隆大街 188-1 号

Tel: 025-66687725

Email: zzjsu@zzfco.net

华中分公司 430000

地址：武汉市江汉区中央商务区泛海国际 SOHO 城第 3、4、6 幢 4 号楼 9 层 7 号

Tel: 027-83555585

Email: zzhz@zzfco.net

宜昌分公司 443000

地址：宜昌市伍家岗区外滩商业中心楼 000301-01

Tel: 0717-6229585

Email: zzyc@zzfco.net