



衍生品量化择时系列专题（六）：

基于繁微数据平台的豆粕基本面因子研究

报告日期：2022 年 2 月 23 日

王冬黎 高级分析师(金融工程)

从业资格号：F3032817

投资咨询号：Z0014348

Tel: 8621-63325888-3975

Email: dongli.wang@orientfutures.com

联系人：谢怡伦（分析师）

从业资格号：F03091687

Tel: 8621-63325888-1585

Email: yilun.xie@orientfutures.com

★基本面数据整理：

基于豆粕上下游产业链关系，梳理了 5 大类，共 144 个基本面/价量数据，包括期货市场，现货价格，进出口，产销库以及下游饲料养殖。对原始数据进行去极值，标准化，数据可得性调整，频率调整等处理，并新增各因子的月度环比，季度环比以及同比数据，作为模型的输入因子。

★降维与模型选择：

PCA 降维：对于数据进行主成分降维，以可解释性方差为依据截取 99% 的信息度，在进行滚动回归时实时对 144 个因子进行降维；

OLS 多元回归：利用普通最小二乘法进行多元回归，最优拟合曲线使各点到直线的距离的平方和（残差平方和 RSS）最小；

★模型结果：

综合模型回测结果达到 2019 年至今年年化收益 22.03%，年化波动 11.35%，夏普值 1.73，最大回撤-14.00%，胜率 0.60，盈亏比 0.94，收益风险比 1.57，sortino 比率 3.08，平均持仓时间 60 天。

★致谢：

感谢东方证券金融工程首席分析师朱剑涛老师的指导与帮助。

★风险提示：

市场风格的变换会造成特征有效性变化，导致模型效果下降。

相关报告

《衍生品量化择时系列专题（二）——螺纹钢指标筛选与大类因子研究》

《衍生品量化择时系列专题（五）——基于机器学习的螺纹钢价格周度预测》

目录

1、 基本面因子.....	5
1.1、 商品基本面量化原理.....	5
1.2、 繁微数据平台.....	5
1.3、 豆粕基本面因子选取.....	5
2、 降维及模型选择.....	10
3、 数据处理.....	11
4、 单因子预测能力分析.....	11
5、 OLS 多元回归.....	13
6、 总结及展望.....	15
7、 风险提示.....	15
8、 附录：部分因子回测结果展示.....	16

图表目录

图表 1: 饲料养殖产业链.....	6
图表 2: 豆粕价量因子 (期货市场)	6
图表 3: 豆粕价量因子 (现货价格)	7
图表 4: 豆粕基本面因子 (进出口)	8
图表 5: 豆粕基本面因子 (产销库)	8
图表 6: 豆粕基本面因子 (下游饲料养殖)	9
图表 7: 滚动回归.....	10
图表 8: 单因子回测结果展示 (预测周期 25, 滚动窗口 950)	12
图表 9: 单因子夏普值分布.....	12
图表 10: 多元回归回测分析.....	13
图表 11: 多元回归收益曲线图.....	14
图表 12: 因子相关性分布.....	15
图表 13: 期货结算价(1 月交割连续):豆粕单因子回测结果.....	16
图表 14: 注册仓单量:豆粕单因子回测结果.....	16
图表 15: CBOT 豆粕:非商业多头持仓量单因子回测结果.....	17
图表 16: CBOT 豆粕:非商业空头持仓量单因子回测结果.....	17
图表 17: CBOT 豆粕:非商业套期头寸单因子回测结果.....	18
图表 18: 豆粕: 卖持量: 前 5 名会员合计单因子回测结果.....	18
图表 19: 豆粕: 成交量: 前 5 名会员合计单因子回测结果.....	19
图表 20: 现货价:豆粕:哈尔滨单因子回测结果.....	19
图表 21: 现货价:豆粕:石家庄单因子回测结果.....	20
图表 22: 进口数量:豆粕:累计值单因子回测结果.....	20
图表 23: 进口数量:豆粕:累计同比单因子回测结果.....	21
图表 24: 进口数量:豆粕:进口地:中国台湾:累计值单因子回测结果.....	21
图表 25: 进口数量:豆粕:进口地:中国台湾:累计同比单因子回测结果.....	22
图表 26: 进口数量:豆粕:进口地:印度:累计同比单因子回测结果.....	22
图表 27: 进口到港数量:豆粕:华东区:当期值单因子回测结果.....	23
图表 28: 生猪存栏因子回测结果.....	23
图表 29: 生猪存栏变化率:环比增减单因子回测结果.....	24
图表 30: 生猪存栏变化率:同比增减单因子回测结果.....	24
图表 31: 生猪出栏:同比单因子回测结果.....	25
图表 32: 豆粕:库存:广西地区单因子回测结果.....	25
图表 33: 豆粕:库存:东北地区单因子回测结果.....	26

图表 34: 豆粕: 卖持量: 前 10 名会员合计(季度环比)单因子回测结果.....	26
图表 35: 油厂豆粕报价: 辽宁 (同比) 单因子回测结果.....	27
图表 36: 油厂豆粕报价: 广东 (季度环比) 单因子回测结果.....	27
图表 37: 国内 1135 家饲料产量: 总饲料 (月度环比) 单因子回测结果.....	28
图表 38: 国内 1135 家饲料产量: 总饲料 (季度环比) 单因子回测结果.....	28
图表 39: 国内 1135 家饲料产量: 禽料 (季度环比) 单因子回测结果.....	29
图表 40: 全国豆粕未执行合同 (同比) 单因子回测结果.....	29
图表 41: 豆粕: 库存: 广西地区 (同比) 单因子回测结果.....	30
图表 42: 全国豆粕周度产量 (同比) 单因子回测结果.....	30

1、基本面因子

1.1、商品基本面量化原理

基本面量化试图有效地结合“基本面研究”和“量化投资”两套独立而成熟的交易理论——以大量的基本面数据为基础对未来价格做出预测，并自动生成有效的交易信号，不仅遵从基本面研究的严谨逻辑，又兼具量化交易的高效分析。具体而言，在本研究报告中，以豆粕的基本面数据作为输入，通过模型对未来价格进行预测，接着根据预测价格生成相应的多空信号进行交易回测。

1.2、繁微数据平台

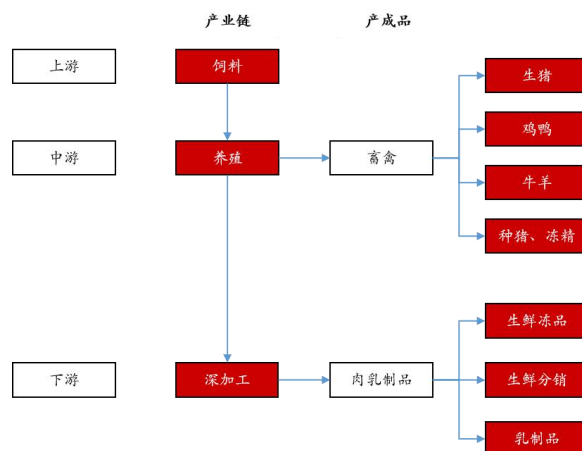
繁微数据平台（www.finoview.com.cn）是东证期货打造的智能投研数据一体化平台，提供商品策略、另类数据、研究报告等一系列卖方服务。繁微与大量第三方数据平台合作，提供从宏观到微观的多维度数据进行投研分析，此外，推出另类数据品牌，卫星数据、移动数据、航运数据，让投资者对大宗商品的研究更加多样化。繁微的背后是一支由30多位研究员组成的队伍，平台通过商品监控、实时策略、范式模板、新闻点评、每周观点、产业链图、研究报告、智能助手等功能，多角度、多方式地给投资者呈现最新的资讯和研究成果。

本报告基于繁微数据平台，共选取5大类144个豆粕相关指标，对豆粕未来价格进行多周期的滚动预测，并根据相应的预测数据生成回测指标。

1.3、豆粕基本面因子选取

豆粕属于饲料养殖产业链中不可缺失的一环，该产业链可分为饲料企业、养殖企业和深加工企业，分别对应上游、中游、下游。从利润来源进行分析，上游企业主要使用豆粕及其他原料作为原材料进行加工，赚取加工费用；中游养殖企业购买饲料、仔猪等进行饲养，出售生猪等，赚取养殖利润；下游深加工企业以生猪肉作为原材料，进行深加工、分销等，赚取附加值。

图表 1：饲料养殖产业链



资料来源：东证衍生品研究院

基于繁微数据平台，整理豆粕相关因子共5大类144个。其中“期货市场”因子27个，涵盖了豆粕期货市场上的价量信息；“现货价格”数据37个，包含不同地区的豆粕现货价格；“进出口”数据17个，囊括了来自海外的进口价量以及国内各港口的到港数据；“产销库”18个，包括各地豆粕的产量、库存、销售数据；最后“下游饲料养殖”45个，覆盖了下端各环节的需求数据。具体指标见下列表格：

图表 2：豆粕价量因子（期货市场）

指标代码	分类	指标名称	指标来源	指标单位	指标频度
DZ02012852	期货市场	期货结算价(1月交割连续):豆粕	Wind	元/吨	日度
DZ02012853	期货市场	期货结算价(5月交割连续):豆粕	Wind	元/吨	日度
DZ02012854	期货市场	期货结算价(9月交割连续):豆粕	Wind	元/吨	日度
DZ02012855	期货市场	期货收盘价(1月交割连续):豆粕	Wind	元/吨	日度
DZ02012856	期货市场	期货收盘价(5月交割连续):豆粕	Wind	元/吨	日度
DZ02012857	期货市场	期货收盘价(9月交割连续):豆粕	Wind	元/吨	日度
DZ02012858	期货市场	期货结算价(活跃合约):豆粕	Wind	元/吨	日度
DZ02012859	期货市场	期货收盘价(活跃合约):豆粕	Wind	元/吨	日度
DZ02012861	期货市场	注册仓单量:豆粕	Wind	手	日度
DZ02012864	期货市场	CBOT 豆粕:非商业多头持仓量	Bloomberg	手	周度
DZ02012865	期货市场	CBOT 豆粕:非商业空头持仓量	Bloomberg	手	周度
DZ02012866	期货市场	CBOT 豆粕:非商业套期头寸	Bloomberg	手	周度
DZ02012867	期货市场	CBOT 豆粕:商业空头持仓量	Bloomberg	手	周度
DZ02012868	期货市场	CBOT 豆粕:商业多头持仓量	Bloomberg	手	周度
DZ02012869	期货市场	CBOT 豆粕:合计多头	Bloomberg	手	周度
DZ02012870	期货市场	CBOT 豆粕:合计空头	Bloomberg	手	周度
DZ02012871	期货市场	CBOT 豆粕:总持仓	Bloomberg	手	周度
DZ02012872	期货市场	CBOT 豆粕:非报告多头头寸	Bloomberg	手	周度
DZ02012873	期货市场	CBOT 豆粕:非报告空头头寸	Bloomberg	手	周度
DZ02012875	期货市场	豆粕: 买持仓: 前10名会员合计	DCE	手	日度
DZ02012876	期货市场	豆粕: 买持仓: 前20名会员合计	DCE	手	日度
DZ02012877	期货市场	豆粕: 卖持仓: 前5名会员合计	DCE	手	日度
DZ02012878	期货市场	豆粕: 卖持仓: 前10名会员合计	DCE	手	日度
DZ02012879	期货市场	豆粕: 卖持仓: 前20名会员合计	DCE	手	日度
DZ02012880	期货市场	豆粕: 成交量: 前5名会员合计	DCE	手	日度
DZ02012881	期货市场	豆粕: 成交量: 前10名会员合计	DCE	手	日度
DZ02012882	期货市场	豆粕: 成交量: 前20名会员合计	DCE	手	日度

资料来源：东证衍生品研究院

图表 3：豆粕价量因子（现货价格）

指标代码	分类	指标名称	指标来源	指标单位	指标频度
DZ02012884	现货价格	现货价:豆粕:哈尔滨	Wind	元/吨	日度
DZ02012885	现货价格	现货价:豆粕:四平	Wind	元/吨	日度
DZ02012886	现货价格	现货价:豆粕:沈阳	Wind	元/吨	日度
DZ02012887	现货价格	现货价:豆粕:大连	Wind	元/吨	日度
DZ02012888	现货价格	现货价:豆粕:北京	Wind	元/吨	日度
DZ02012889	现货价格	现货价:豆粕:天津	Wind	元/吨	日度
DZ02012890	现货价格	现货价:豆粕:石家庄	Wind	元/吨	日度
DZ02012891	现货价格	现货价:豆粕:秦皇岛	Wind	元/吨	日度
DZ02012892	现货价格	现货价:豆粕:太原	Wind	元/吨	日度
DZ02012893	现货价格	现货价:豆粕:兰州	Wind	元/吨	日度
DZ02012894	现货价格	现货价:豆粕:济宁	Wind	元/吨	日度
DZ02012895	现货价格	现货价:豆粕:日照	Wind	元/吨	日度
DZ02012896	现货价格	现货价:豆粕:烟台	Wind	元/吨	日度
DZ02012897	现货价格	现货价:豆粕:周口	Wind	元/吨	日度
DZ02012898	现货价格	现货价:豆粕:西安	Wind	元/吨	日度
DZ02012899	现货价格	现货价:豆粕:南京	Wind	元/吨	日度
DZ02012900	现货价格	现货价:豆粕:合肥	Wind	元/吨	日度
DZ02012901	现货价格	现货价:豆粕:成都	Wind	元/吨	日度
DZ02012902	现货价格	现货价:豆粕:连云港	Wind	元/吨	日度
DZ02012903	现货价格	现货价:豆粕:张家港	Wind	元/吨	日度
DZ02012904	现货价格	现货价:豆粕:上海	Wind	元/吨	日度
DZ02012905	现货价格	现货价:豆粕:宁波	Wind	元/吨	日度
DZ02012906	现货价格	现货价:豆粕:南昌	Wind	元/吨	日度
DZ02012907	现货价格	现货价:豆粕:长沙	Wind	元/吨	日度
DZ02012908	现货价格	现货价:豆粕:贵阳	Wind	元/吨	日度
DZ02012909	现货价格	现货价:豆粕:昆明	Wind	元/吨	日度
DZ02012910	现货价格	现货价:豆粕:南宁	Wind	元/吨	日度
DZ02012911	现货价格	现货价:豆粕:泉州	Wind	元/吨	日度
DZ02012912	现货价格	现货价:豆粕:湛江	Wind	元/吨	日度
DZ02012913	现货价格	现货价:豆粕:东莞	Wind	元/吨	日度
DZ02012914	现货价格	现货价:豆粕:防城	Wind	元/吨	日度
DZ02012915	现货价格	现货价:豆粕:武汉	Wind	元/吨	日度
DZ02012917	现货价格	油厂豆粕报价:辽宁	天下粮仓	元/吨	日度
DZ02012918	现货价格	油厂豆粕报价:天津	天下粮仓	元/吨	日度
DZ02012919	现货价格	油厂豆粕报价:江苏	天下粮仓	元/吨	日度
DZ02012920	现货价格	油厂豆粕报价:广东	天下粮仓	元/吨	日度
DZ02012921	现货价格	油厂豆粕报价:广西	天下粮仓	元/吨	日度

资料来源：东证衍生品研究院

图表 4：豆粕基本面因子（进出口）

指标代码	分类	指标名称	指标来源	指标单位	指标频度
DZ02012924	进出口	进口平均单价:豆粕:当月值	Wind	美元/吨	月度
DZ02012927	进出口	进口数量:豆粕:当月值	Wind	吨	月度
DZ02012928	进出口	进口数量:豆粕:累计值	Wind	吨	月度
DZ02012929	进出口	进口数量:豆粕:累计同比	Wind	%	月度
DZ02012930	进出口	进口数量:豆粕:进口地:丹麦:累计值	Wind	吨	月度
DZ02012931	进出口	进口数量:豆粕:进口地:中国台湾:累计值	Wind	吨	月度
DZ02012932	进出口	进口数量:豆粕:进口地:印度:累计值	Wind	吨	月度
DZ02012933	进出口	进口数量:豆粕:进口地:丹麦:累计同比	Wind	%	月度
DZ02012934	进出口	进口数量:豆粕:进口地:中国台湾:累计同比	Wind	%	月度
DZ02012935	进出口	进口数量:豆粕:进口地:印度:累计同比	Wind	%	月度
DZ02012937	进出口	进口到港数量:豆粕:当期值	Wind	吨	周度
DZ02012938	进出口	进口到港数量:豆粕:华北区:当期值	Wind	吨	周度
DZ02012939	进出口	进口到港数量:豆粕:东北区:当期值	Wind	吨	周度
DZ02012940	进出口	进口到港数量:豆粕:华东区:当期值	Wind	吨	周度
DZ02012941	进出口	进口到港数量:豆粕:中南区:当期值	Wind	吨	周度
DZ02012942	进出口	进口到港数量:豆粕:西南区:当期值	Wind	吨	周度
DZ02012943	进出口	进口到港数量:豆粕:西北区:当期值	Wind	吨	周度

资料来源：东证衍生品研究院

图表 5：豆粕基本面因子（产销库）

指标代码	分类	指标名称	指标来源	指标单位	指标频度
DZ02013026	产销存	全国豆粕月度产量	天下粮仓	吨	周度
DZ02013027	产销存	全国豆粕周度产量	天下粮仓	吨	周度
DZ02013006	产销存	全国豆粕日度成交量	天下粮仓	吨	日度
DZ02013007	产销存	豆粕提货量（38家）	天下粮仓	吨	日度
DZ02013009	产销存	全国豆粕未执行合同	天下粮仓	万吨	周度
DZ02013011	产销存	豆粕:库存:华北地区	天下粮仓	万吨	周度
DZ02013012	产销存	豆粕:库存:广东地区	天下粮仓	万吨	周度
DZ02013013	产销存	豆粕:库存:华东地区	天下粮仓	万吨	周度
DZ02013014	产销存	豆粕:库存:广西地区	天下粮仓	万吨	周度
DZ02013015	产销存	豆粕:库存:山东地区	天下粮仓	万吨	周度
DZ02013016	产销存	豆粕:库存:东北地区	天下粮仓	万吨	周度
DZ02013017	产销存	豆粕:库存:福建地区	天下粮仓	万吨	周度
DZ02013018	产销存	豆粕:库存:沿海总计	天下粮仓	万吨	周度
DZ02013019	产销存	豆粕:库存:全国	天下粮仓	万吨	周度
DZ02013021	产销存	豆粕:库存:山东饲料企业库存平均天数	天下粮仓	天	周度
DZ02013022	产销存	豆粕:库存:大型饲料企业1库存量	天下粮仓	万吨	周度
DZ02013023	产销存	豆粕:库存:大型饲料企业2库存量	天下粮仓	万吨	周度
DZ02013024	产销存	豆粕:库存:中小饲料企业平均库存量	天下粮仓	万吨	周度

资料来源：东证衍生品研究院

图表 6：豆粕基本面因子（下游饲料养殖）

指标代码	分类	指标名称	指标来源	指标单位	指标频度
DZ02012961	下游饲料养殖	产量:饲料:当月值	Wind	吨	月度
DZ02012962	下游饲料养殖	产量:饲料:当月同比	Wind	%	月度
DZ02012996	下游饲料养殖	饲料产量分省份:全国	天下粮仓	万吨	年度
DZ02012997	下游饲料养殖	饲料产量分省份:山东	天下粮仓	万吨	年度
DZ02012998	下游饲料养殖	饲料产量分省份:广东	天下粮仓	万吨	年度
DZ02013000	下游饲料养殖	国内1135家饲料产量:总饲料	天下粮仓	吨	月度
DZ02013001	下游饲料养殖	国内1135家饲料产量:猪料	天下粮仓	吨	月度
DZ02013002	下游饲料养殖	国内1135家饲料产量:禽料	天下粮仓	吨	月度
DZ02013003	下游饲料养殖	国内1135家饲料产量:水产料	天下粮仓	吨	月度
DZ02012979	下游饲料养殖	天邦股份:销量:饲料:合计	Wind	吨	半年度
DZ02012980	下游饲料养殖	天邦股份:销量:饲料:油脂	Wind	吨	半年度
DZ02012981	下游饲料养殖	天邦股份:销量:水产品	Wind	吨	半年度
DZ02012982	下游饲料养殖	海大集团:销量:水产配合饲料	Wind	万吨	半年度
DZ02012983	下游饲料养殖	海大集团:销量:饲料	Wind	万吨	半年度
DZ02012984	下游饲料养殖	海大集团:销量:畜禽配合饲料	Wind	万吨	半年度
DZ02012985	下游饲料养殖	大北农:销量:猪饲料	Wind	万吨	半年度
DZ02012986	下游饲料养殖	大北农:销量:禽用饲料	Wind	万吨	半年度
DZ02012987	下游饲料养殖	大北农:销量:饲料	Wind	万吨	半年度
DZ02012988	下游饲料养殖	大北农:销量:水产饲料	Wind	万吨	半年度
DZ02012989	下游饲料养殖	大北农:销量:反刍饲料	Wind	万吨	半年度
DZ02012990	下游饲料养殖	大北农:销量:其他饲料	Wind	万吨	半年度
DZ02012991	下游饲料养殖	大北农:销量:种子	Wind	万公斤	半年度
DZ02012992	下游饲料养殖	金新农:销量:饲料	Wind	万吨	半年度
DZ02012993	下游饲料养殖	正邦科技:销量:饲料	Wind	万吨	半年度
DZ02012994	下游饲料养殖	正邦科技:出栏量:生猪	Wind	万头	年度
DZ02012957	下游饲料养殖	全国大中城市:出场价格:生猪	Wind	元/公斤	周度
DZ02012958	下游饲料养殖	全国大中城市:批发价:玉米	Wind	元/公斤	周度
DZ02012959	下游饲料养殖	全国大中城市:猪粮比价	Wind		周度
DZ02012946	下游饲料养殖	生猪存栏	Wind	万头	月度
DZ02012947	下游饲料养殖	生猪存栏:能繁母猪	Wind	万头	月度
DZ02012948	下游饲料养殖	生猪存栏变化率:环比增减	Wind	%	月度
DZ02012949	下游饲料养殖	生猪存栏变化率:同比增减	Wind	%	月度
DZ02012950	下游饲料养殖	能繁母猪存栏变化率:环比增减	Wind	%	月度
DZ02012951	下游饲料养殖	能繁母猪存栏变化率:同比增减	Wind	%	月度
DZ02012952	下游饲料养殖	生猪存栏:同比	Wind	%	季度
DZ02012953	下游饲料养殖	生猪出栏:同比	Wind	%	季度
DZ02012955	下游饲料养殖	生猪定点屠宰企业屠宰量	Wind	万头	月度
DZ02012972	下游饲料养殖	产量:肉类	Wind	万吨	年度
DZ02012973	下游饲料养殖	产量:水产品	Wind	万吨	年度
DZ02012974	下游饲料养殖	产量:鸡蛋:中国	Wind	吨	年度
DZ02012975	下游饲料养殖	产量:鲜牛奶:中国	Wind	吨	年度
DZ02012968	下游饲料养殖	产量:水产品	Wind	万吨	年度
DZ02012969	下游饲料养殖	产量:海水产品	Wind	万吨	年度
DZ02012970	下游饲料养殖	产量:淡水产品	Wind	万吨	年度
DZ02012977	下游饲料养殖	渔业总产值:全国:累计值	Wind	亿元	季度

资料来源：东证衍生品研究院

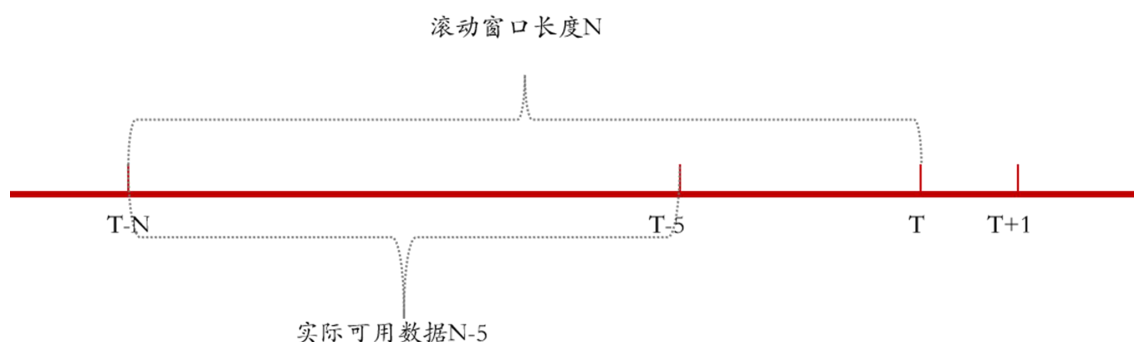
2、降维及模型选择

本报告依据 OLS 进行滚动回归预测。OLS（普通最小二乘法）多元回归的原理为，最优拟合曲线使得各点到直线的距离的平方和（残差平方和 RSS）最小：

$$RSS = \sum_{t=1}^T (y_t - \hat{y}_t)^2 = \sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta}x_t)^2$$

本报告采用滚动回归的方式进行回测，以避免使用未来数据。以周度预测为例，首先设置相应的滚动回归窗口长度 N，对每一天 T 都截取 T-N 到 T 时间段的基本面数据，由于为周度预测，需要对基本面数据进行 5 天的移动处理，所以实际可用数据点为 N-5 天内的数据，随后根据训练模型得到一系列预测值，再根据预测值与真实值的比较去构造回测模型。

图表 7：滚动回归



资料来源：东证衍生品研究院

在进行多元回归之前，报告采用主成分分析法（PCA）进行降维处理，PCA 通过线性变换将原始数据变换为一组各维度线性无关的表示，可用于提取数据的主要特征分量，常用于高维数据的降维。其算法步骤如下：

设有 m 条 n 维数据：

- 将原始数据按列组成 n 行 m 列矩阵 X；
- 将 X 的每一行进行零均值化，即减去这一行的均值；
- 求出协方差矩阵 $C = \frac{1}{m}XX^T$ ；
- 求出协方差矩阵的特征值及对应的特征向量；
- 将特征向量按对应特征值大小从上到下按行排列成矩阵，取前 k 行组成矩阵 P；

- $Y = PX$ 即为降维到 k 维后的数据。

在进行 PCA 降维之后，可以缓解维度灾难，并对数据进行降噪，同时将数据压缩到低维之后，使得降维之后的数据各特征相互独立。但是在另一方面，由于 PCA 保留了主要信息，舍弃了一些看似无用的信息，但这些“无用信息”只是在训练集上没有有效表现，因此产生了过拟合的可能性，这一问题在模型训练时需要注意。

3、数据处理

频率调整：基本面的原始数据多为低频数据（月频或周频），为便于处理，将所有数据前值填充为日频数据；

标准化：对所有填充后的数据进行 z-score 标准化处理，提高数据之间的可比性；

异常值处理：对所有偏离均值 3 个标准差的数据进行处理；

移仓换月处理：为避免期货展期导致的价格影响，本报告以豆粕复权价格进行回测；

周期性影响：为剔除数据的周期性影响，对数据分别进行环比，同比等处理；

可得性处理：按指标具体可得性进行相应滞后处理；

回测参数：本报告基于日频数据生成周频级别的多空信号，每周更新仓位，手续费设置为双边万三。

4、单因子预测能力分析

首先分别对 144 个因子进行单独回归考察其各自的预测能力，在原有 144 个因子的基础上，添加各个因子月度环比，季度环比，以及同比值，构造新的衍生因子。经过扩充后的因子数量达到 576 (144×4) 个。下表分别列出了收益前十和后十名的因子回测数据。收益较好的因子主要是 DZ02013000, DZ02013001, DZ02013002, DZ02012986, DZ02012861 以及其衍生因子，对照上文因子表格，这些因子分别对应为国内 1135 家饲料产量:总饲料 (DZ02013000)，国内 1135 家饲料产量:猪料 (DZ02013001)，国内 1135 家饲料产量:禽料 (DZ02013002)，大北农:销量:禽用饲料 (DZ02012986)，注册仓单量:豆粕 (DZ02012861)，其中除了注册仓单外，其余均为“下游饲料养殖”类因子。说明在豆粕的上下游产业链中，下游段的饲料需求对豆粕价格的涨跌具有较强的预测能力。通过对五大类因子的平均表现分析，也同样可以发现，“下游饲料养殖”类因子的平均表现明显好于其他四大类因子。对应表现较好的因子，其胜率可以达到 0.59，年化波动相对而言也更低，但是另一方面，平均持仓时间相对较短，也就意味这换手率提高，拉升了整体的交易成本，但是考虑到其平均持仓时间均在一个月以上，换手成本仍可维持在较低程度。

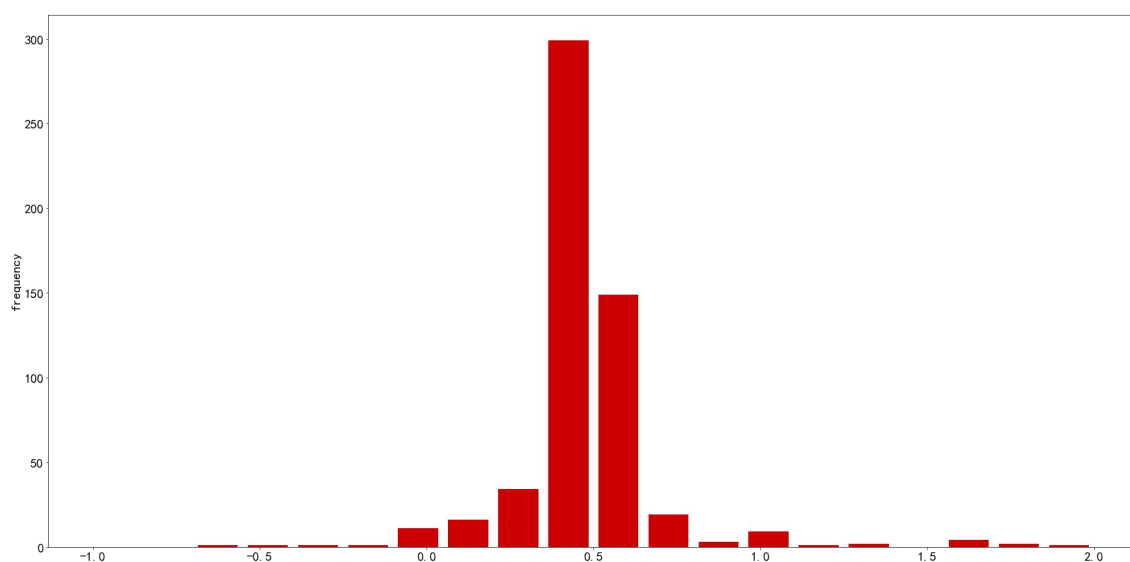
图表 8：单因子回测结果展示（预测周期 25，滚动窗口 950）

Code	总收益	年化收益	年化波动	夏普值	最大回撤	收益风险比	胜率	sortino比率	平均持仓时间
DZ02013000_ratio_20	102.77%	25.66%	11.15%	2.08	-9.02%	2.85	0.59	4.02	48.75
DZ02013001_ratio_20	102.77%	25.66%	11.15%	2.08	-9.02%	2.85	0.59	4.02	48.75
DZ02013002_ratio_60	93.05%	23.68%	11.45%	1.86	-9.53%	2.48	0.59	3.38	55.71
DZ02013002_ratio_20	90.53%	23.15%	11.31%	1.84	-9.02%	2.57	0.58	3.48	55.71
DZ02012986_ratio_250	83.26%	21.62%	11.07%	1.74	-11.08%	1.95	0.58	3.17	60
DZ02013002	80.80%	21.09%	11.62%	1.61	-9.53%	2.21	0.58	2.88	65
DZ02013001	79.65%	20.84%	11.57%	1.59	-9.53%	2.19	0.58	2.87	55.71
DZ02013000_ratio_60	79.65%	20.84%	11.57%	1.59	-9.53%	2.19	0.58	2.87	55.71
DZ02013001_ratio_60	79.65%	20.84%	11.57%	1.59	-9.53%	2.19	0.58	2.87	55.71
DZ02012861	68.76%	18.42%	12.04%	1.33	-14.69%	1.25	0.58	2.33	65
.....								
DZ02012973_ratio_250	9.52%	2.98%	13.90%	0.04	-24.12%	0.12	0.56	0.06	156
DZ02012981_ratio_20	9.52%	2.98%	13.90%	0.04	-24.12%	0.12	0.56	0.06	156
DZ02013019	8.08%	2.54%	14.07%	0.01	-24.37%	0.1	0.56	0.02	130
DZ02012949	5.96%	1.89%	18.34%	-0.03	-28.52%	0.07	0.55	-0.05	97.5
DZ02012957_ratio_250	5.01%	1.59%	14.46%	-0.06	-25.26%	0.06	0.55	-0.08	130
DZ02012994_ratio_250	3.15%	1.01%	14.47%	-0.1	-27.02%	0.04	0.55	-0.15	195
DZ02012981_ratio_250	0.24%	0.08%	14.77%	-0.16	-25.80%	0	0.54	-0.24	97.5
DZ02012951	-10.58%	-3.55%	17.91%	-0.33	-28.52%	-0.12	0.55	-0.5	78
DZ02012865	-14.03%	-4.77%	17.43%	-0.41	-27.61%	-0.17	0.55	-0.62	130
DZ02012981_ratio_60	-27.15%	-9.73%	19.20%	-0.63	-34.52%	-0.28	0.53	-0.94	195

资料来源：东证衍生品研究院

下图用直方图展示了全部 144 个因子的夏普值分布，其均值为 0.48。大部分因子对豆粕期货价格数据具有正向的预测能力，然而绝大部分因子的夏普值在 0.6 以下，只有少数因子夏普值超过 1。

图表 9：单因子夏普值分布



资料来源：东证衍生品研究院

5、OLS 多元回归

在该部分，针对全部因子进行多元回归，观测其回测结果。

在进行多元回归之前，需要对数据进行降维，通过计算因子的可解释性变量，截取99%的信息度。在信号处理过程中，有效信号具有较大方差，而噪声具有较小方差，信噪比就是信号和噪声的方差比，该值越大越好。为了确定PCA降维的维度数量，需要计算可解释性方差，即将各个特征值大小除以求和的总特征值得到各个特征重要性占比。通过累计可解释性方差，截取99%的信息度。

针对不同的滚动窗口和预测周期，对豆粕进行多周期的滚动回测，结果如下：

图表 10：多元回归回测分析

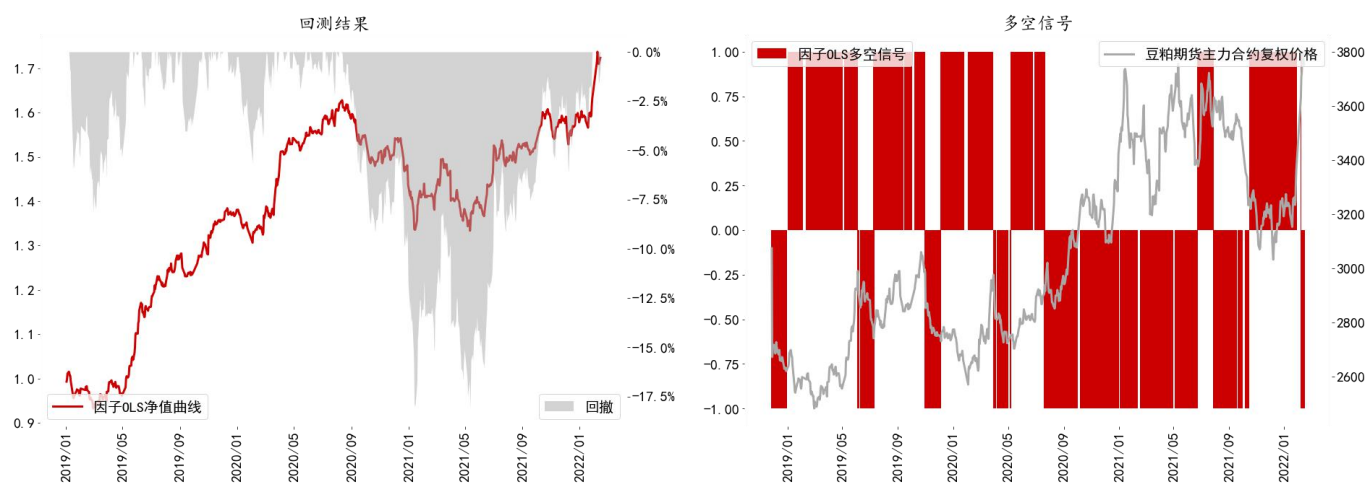
预测周期(天)	滚动周期(天)																
	200	250	300	350	400	450	500	550	600	650	700	750	800	850	900	950	1000
1	-0.42	-0.27	-0.43	0.13	0.34	0.5	0.56	0.21	-0.04	0.04	0.07	0.15	0.31	0.3	0.18	-0.04	-0.37
5	-0.13	0.29	-0.13	0.11	0.85	0.43	0.67	0.49	0.32	0.02	0.07	0.02	0.56	0.25	0.32	0.14	0.11
10	-0.28	0.35	0.07	0.6	0.41	-0.11	0.21	0.25	0.21	0.16	0.2	0.01	0.47	0.14	0.14	-0.22	-0.08
15	0.62	0.4	0.08	1.03	0.03	0.37	0.39	-0.49	0.27	0.08	-0.55	1.09	0.61	-0.44	1.17	0.59	0.04
20	0.15	0.14	0.39	0.65	0.15	0.84	-0.21	0.13	-0.64	0.16	-0.68	-0.04	-0.48	0.19	0.44	0.55	0.15
25	0.32	0.22	0.11	0.25	0.49	0.85	0.74	0.52	0.92	0.73	0.7	0.95	0.83	0.91	1.34	1.73	1.18
30	0.39	0.15	0.51	0.99	0.29	0.91	1.05	0.26	0.34	0.56	0.28	0.71	0.33	0.66	0.8	-0.15	0.18

资料来源：东证衍生品研究院

观察不同预测周期及滚动周期的夏普值可以发现，回测效果随参数变化的敏感性较强，且并未出现明显的规律性。从预测周期长度而言，当预测周期为25时，其总体预测能力更强，显示对于豆粕基本面因子，长期的预测能力好于短期预测能力。表现最好的组合为预测周期25天，滚动周期950天，具体表现为2019年至今年年化收益22.03%，年化波动11.35%，夏普值1.73，最大回撤-14.00%，胜率0.60，盈亏比0.94，收益风险比1.57，sortino比率3.08，平均持仓时间60天。

下图展示了具体的收益曲线图，可以发现三次较大幅度的回撤发生在2020年1月，2021年1月以及2021年5月，参考豆粕期货价格数据，在这三段时期内，豆粕期货价格均发生了较大幅度的波动。

图表 11: 多元回归收益曲线图

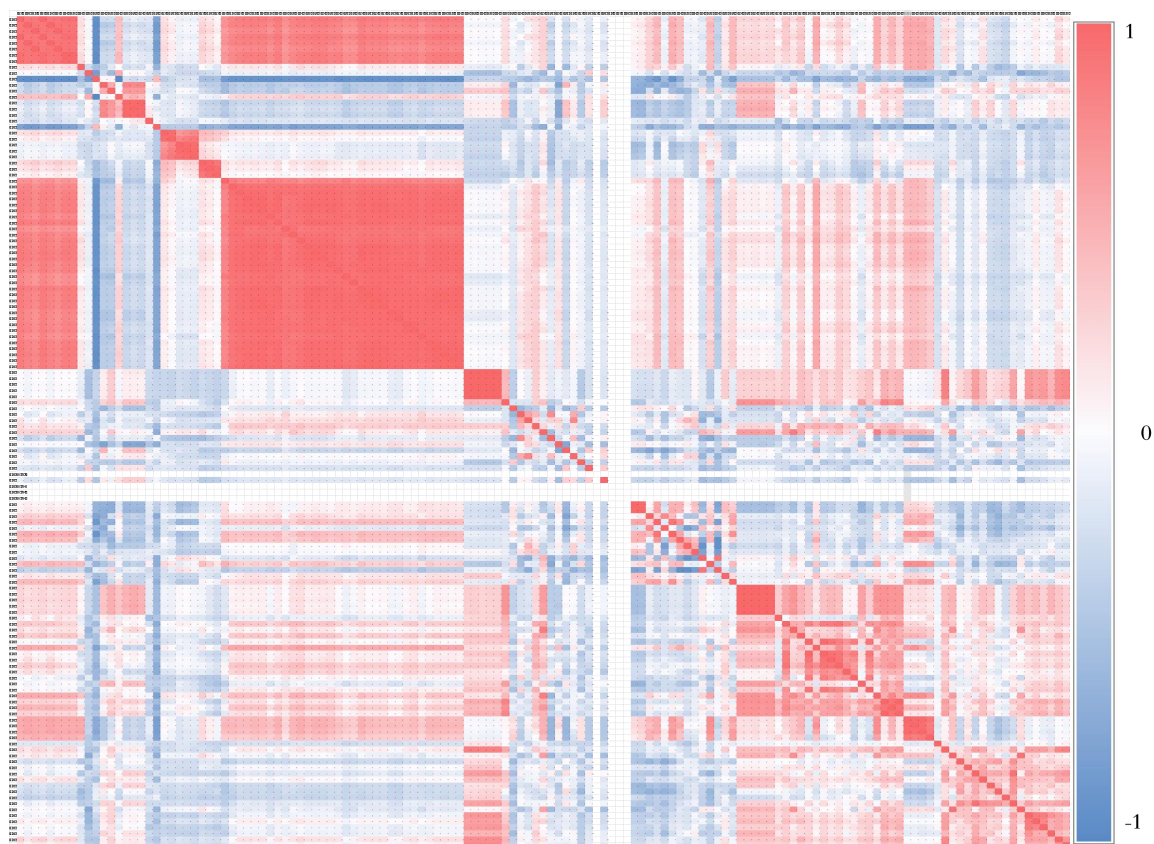


资料来源: 东证衍生品研究院

在对豆粕进行多因子回归后, 出现了一个异常现象, 即多因子回归后的结果并不好于预测能力较强的个别单因子, 我们通过考察各因子的相关性可以发现其潜在原因。由下表各因子相关性色阶图可以发现, 在单因子内部存在大量相关性极强的因子。在去除相关性大于 0.7 的强相关性因子后, 可用因子只剩下 37 个, 接下来, 我们进一步在 37 个因子的基础上筛选出年化收益高于 10%, 夏普值大于 0.6 的具有较强预测能力的因子, 经过第二轮筛选, 最后剩下的因子只有三个, 分别是: **DZ02012861** (注册仓单量: 豆粕), **DZ02012946** (生猪存栏), **DZ02012962** (产量: 饲料: 当月同比), 即便如此, 这三者之间平均相关系数也达到了 0.45。显然这三个因子对于因子组合的贡献是有限的。

至于如何提升因子组合的预测能力, 一个可行的方法就是先构造足够多的有效单因子, 可以通过对现有基本面因子进行运算来构造, 比如计算两个因子的比值或者相关性, 如通过计算库存与销量的比值得到期货品种的库销比。这需要对相关期货品种有一定的认知度。在后续的研究中, 我们会与相关商品分析师进行合作去构造各品种的因子库, 以丰富投资者的策略选择。

图表 12: 因子相关性分布



资料来源：东证衍生品研究院

6、总结及展望

相较于本系列之前的螺纹钢基本面量化，豆粕作为农产品中上下游产业链相对复杂的期货品种，其价格的精确预测更为复杂，预测的准确度也更低。研究过程中发现了一个有趣的现象：对于月度级别的豆粕价格预测，个别单因子的预测能力甚至高于组合因子。基于上文分析，我们发现在删除强相关性的因子之后，真正可用的有效因子数量并不多。在后续的研究中，我们立志于建立起丰富的商品期货因子库，在此基础上进一步深化期货各品种的基本面量化研究。

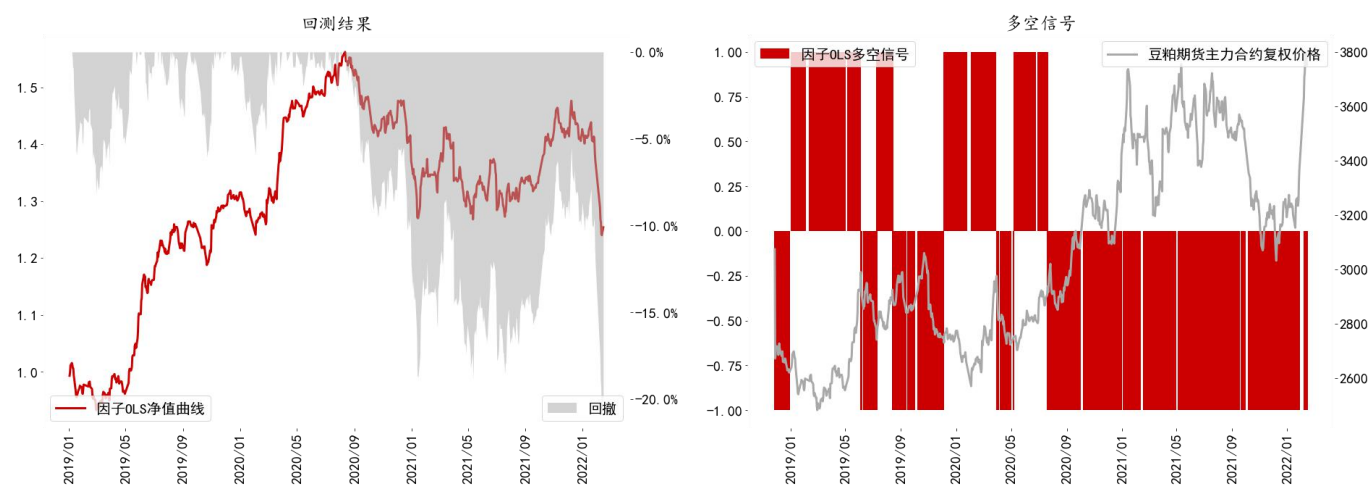
7、风险提示

市场风格的变换会造成特征有效性变化，导致模型效果下降。

8、附录：部分因子回测结果展示

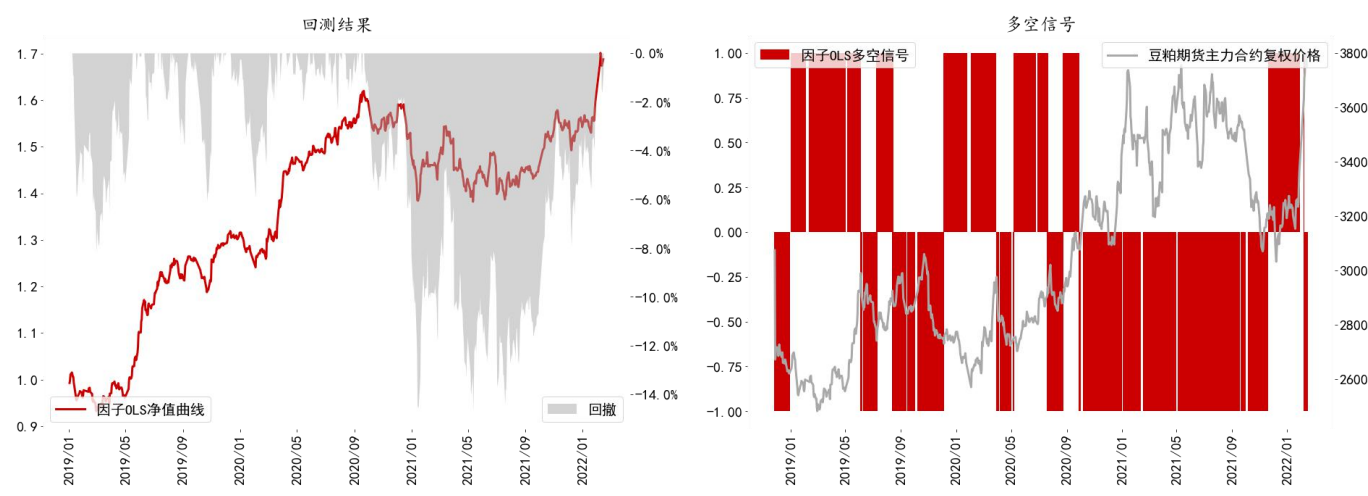
在附录中给出部分单因子的回测结果展示。

图表 13: 期货结算价(1月交割连续):豆粕单因子回测结果



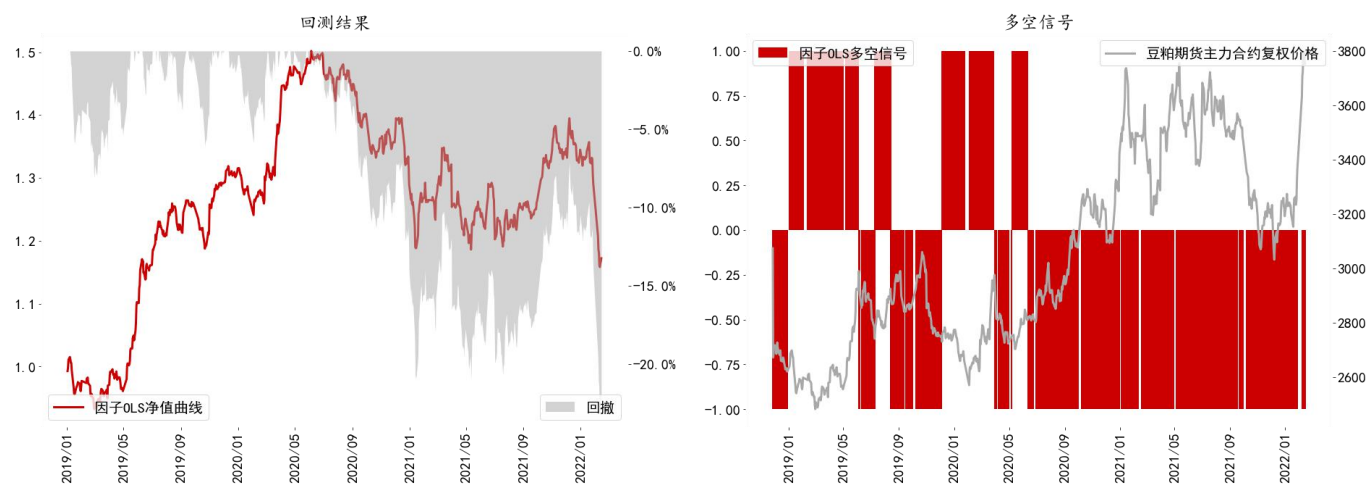
资料来源：东证衍生品研究院

图表 14: 注册仓单量:豆粕单因子回测结果



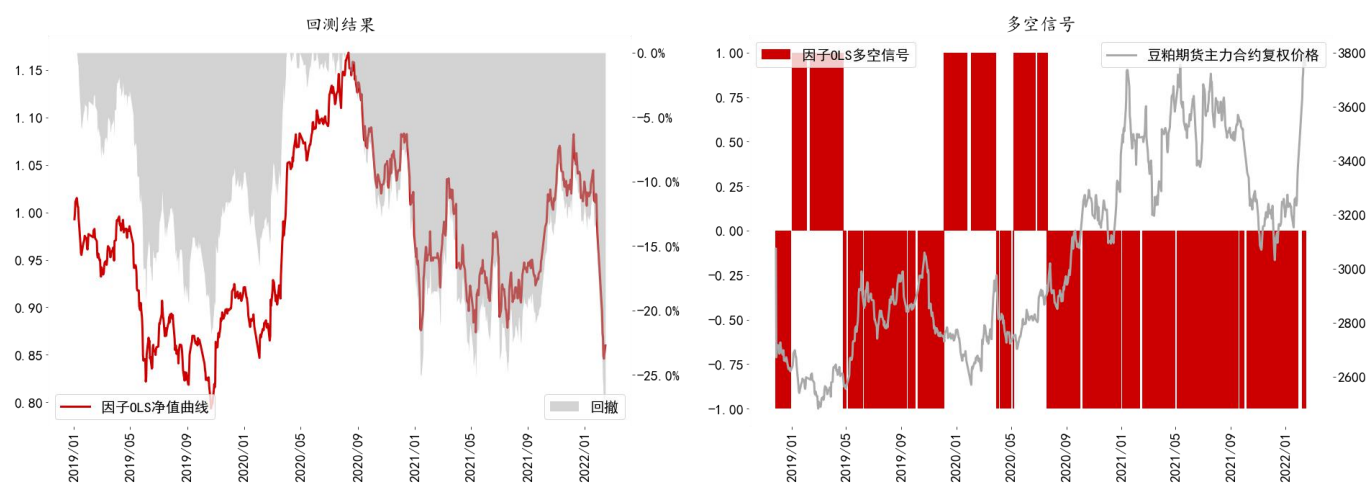
资料来源：东证衍生品研究院

图表 15: CBOT 豆粕:非商业多头持仓量单因子回测结果



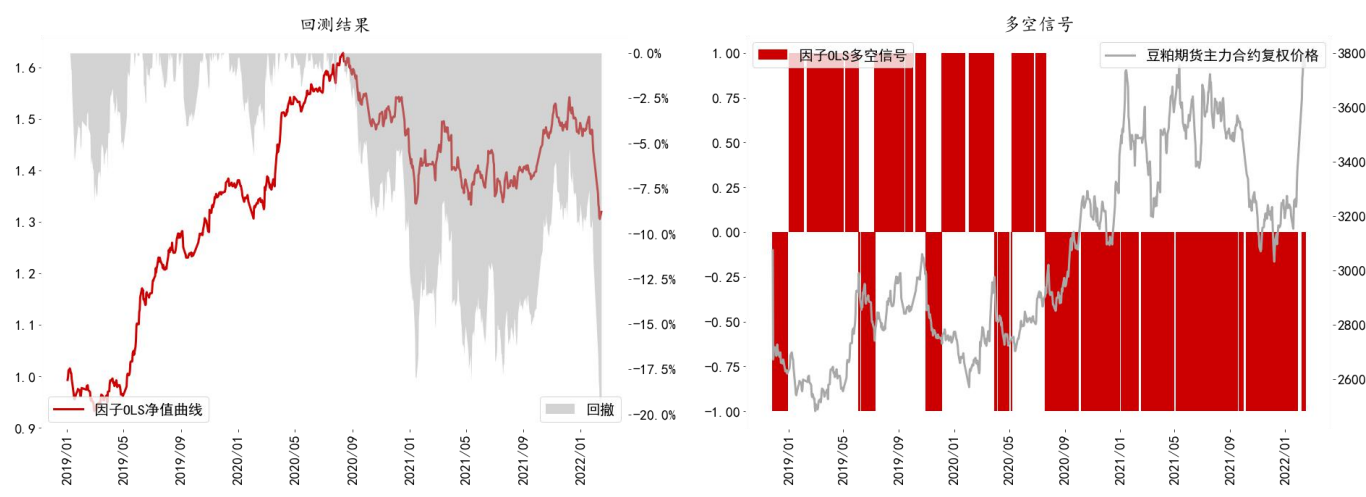
资料来源: 东证衍生品研究院

图表 16: CBOT 豆粕:非商业空头持仓量单因子回测结果



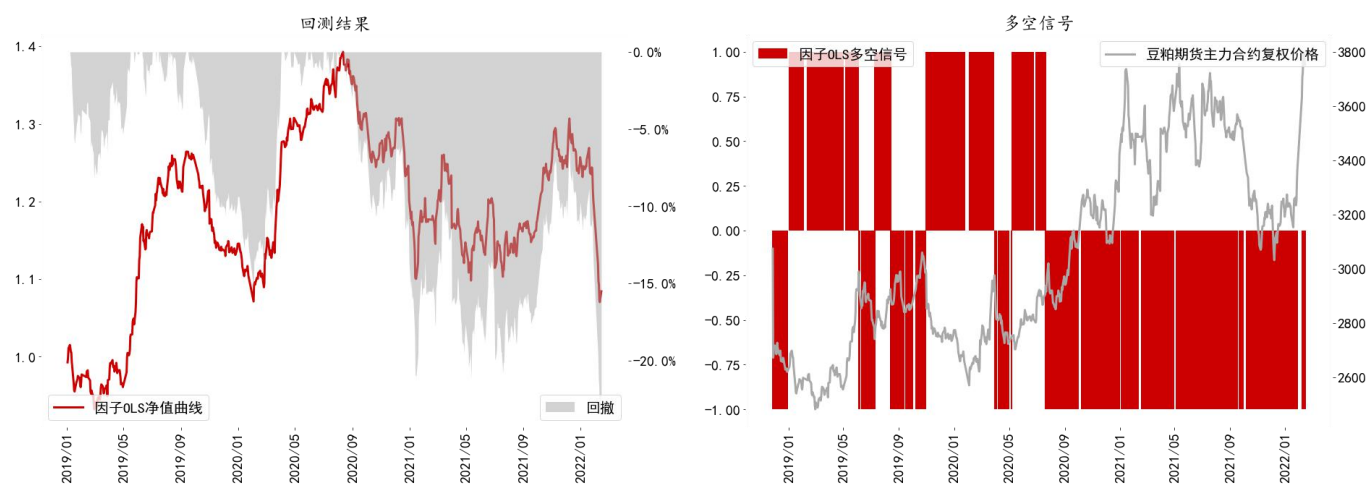
资料来源: 东证衍生品研究院

图表 17：CBOT 豆粕:非商业套期头寸单因子回测结果



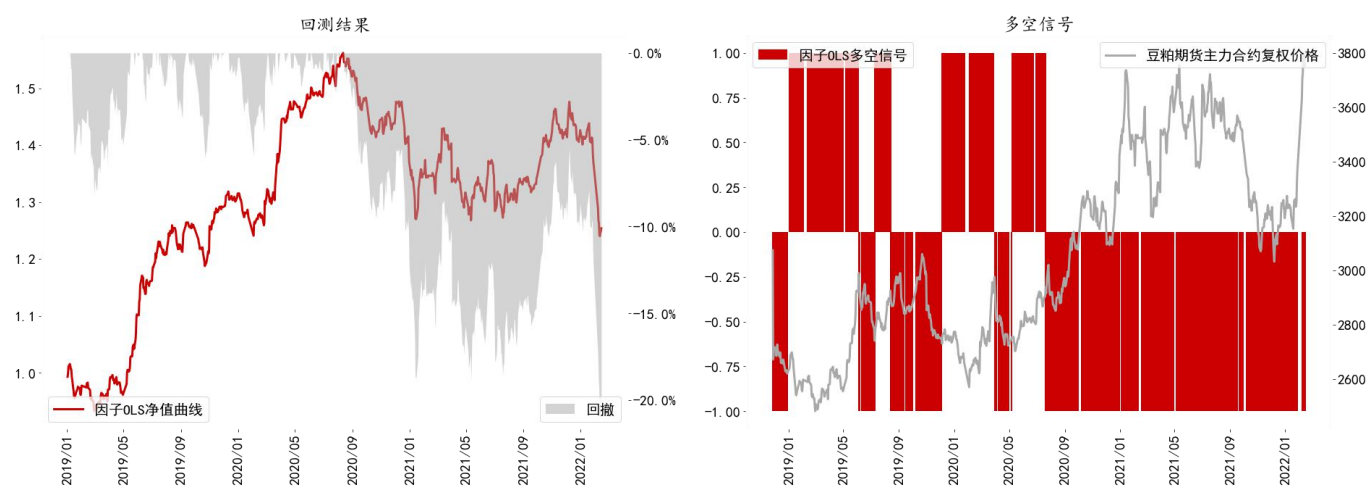
资料来源：东证衍生品研究院

图表 18：豆粕：卖持仓：前5名会员合计单因子回测结果



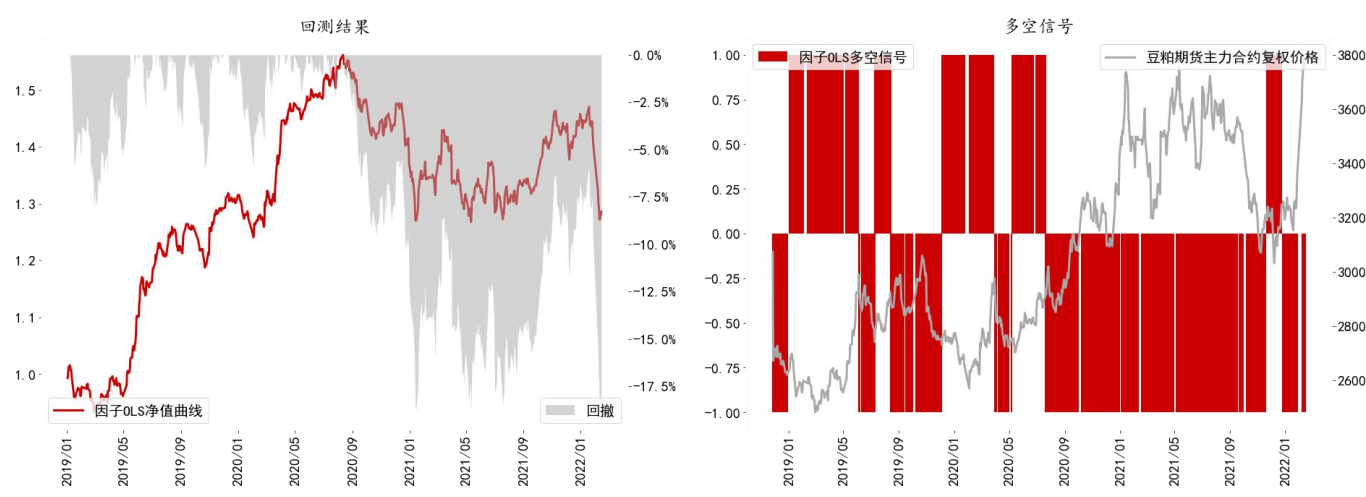
资料来源：东证衍生品研究院

图表 19：豆粕：成交量：前 5 名会员合计单因子回测结果



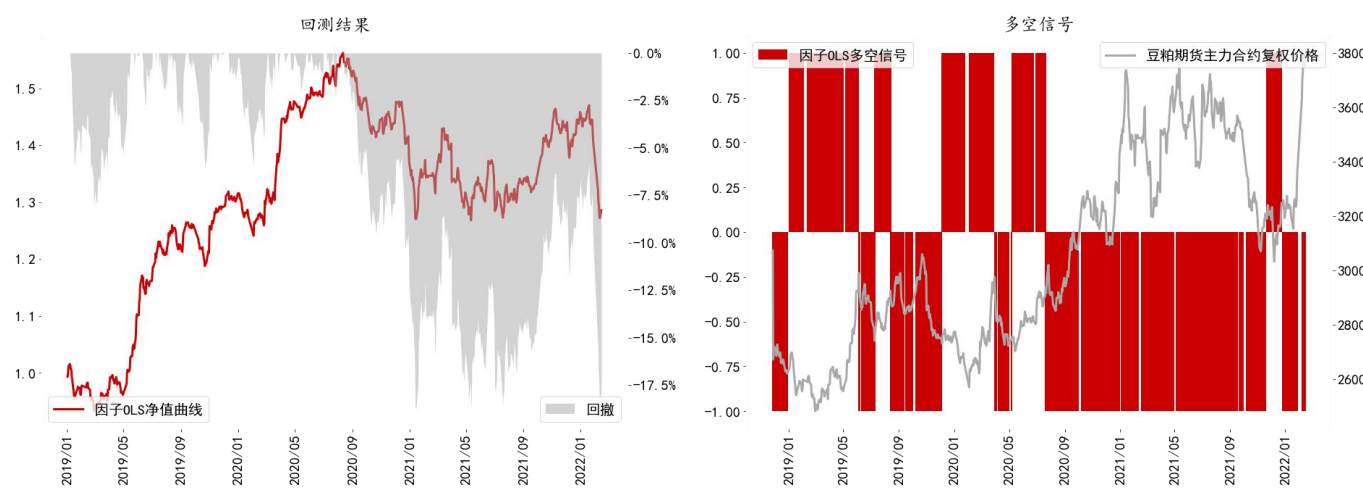
资料来源：东证衍生品研究院

图表 20：现货价:豆粕:哈尔滨单因子回测结果



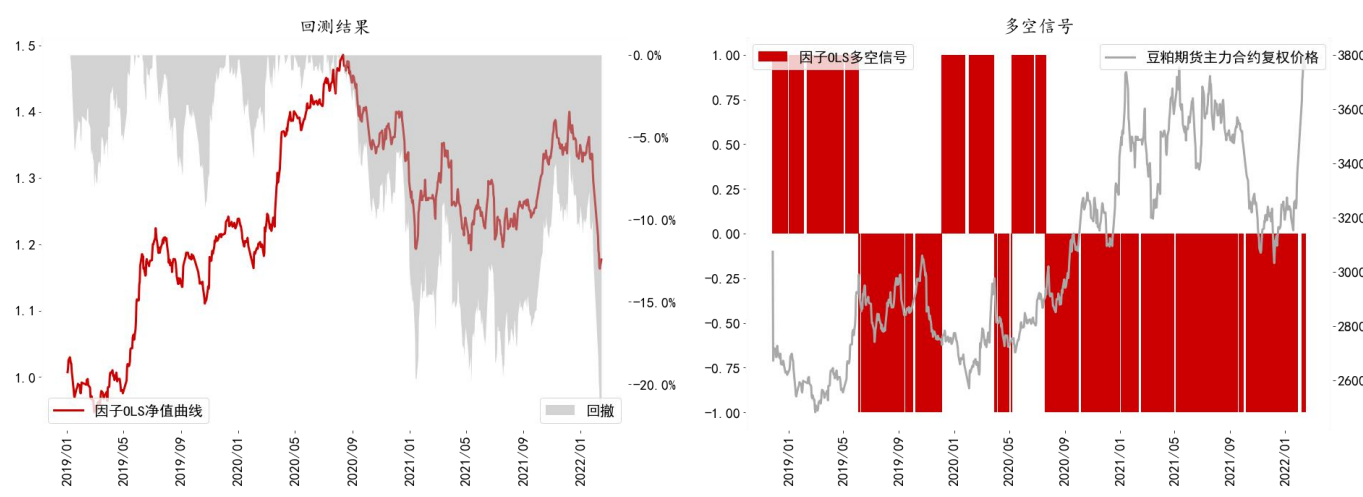
资料来源：东证衍生品研究院

图表 21: 现货价:豆粕:石家庄单因子回测结果



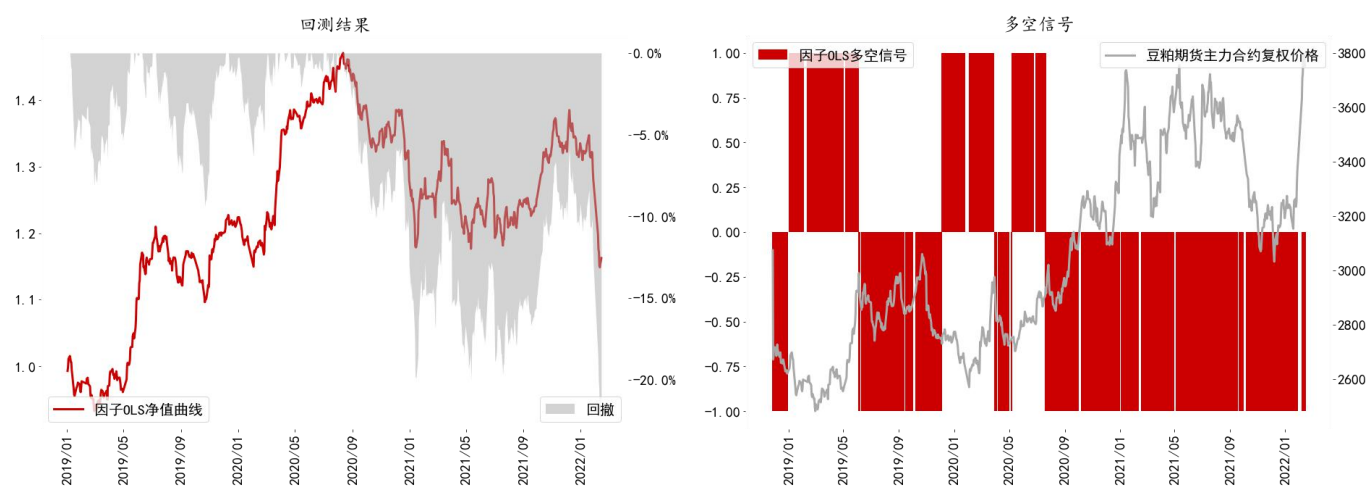
资料来源: 东证衍生品研究院

图表 22: 进口数量:豆饼、豆粕:累计值单因子回测结果



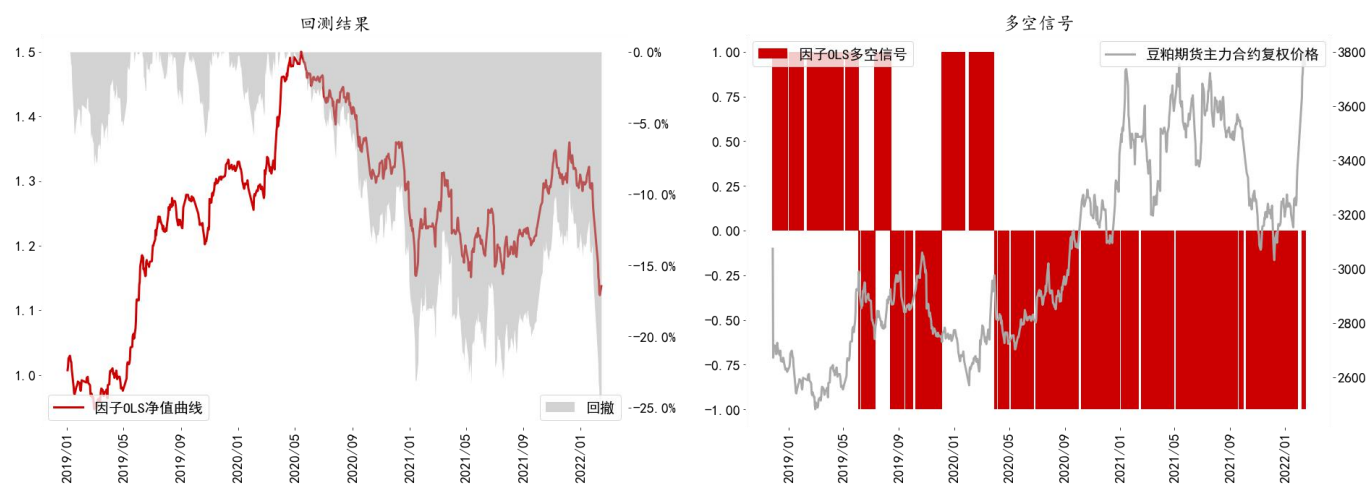
资料来源: 东证衍生品研究院

图表 23: 进口数量:豆饼、豆粕:累计同比单因子回测结果



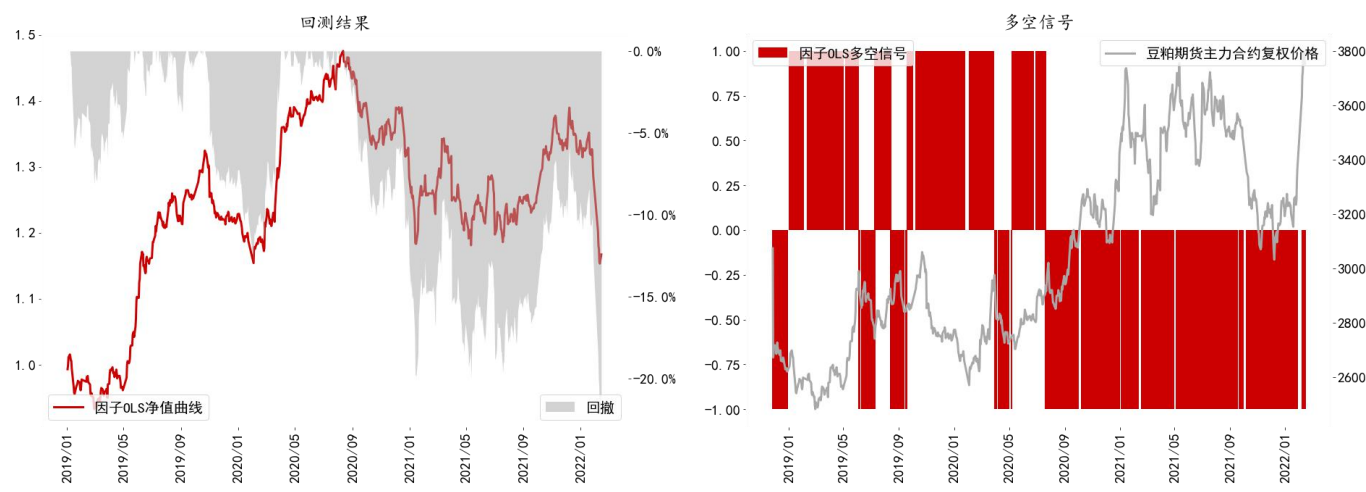
资料来源: 东证衍生品研究院

图表 24: 进口数量:豆饼、豆粕:进口地:中国台湾:累计值单因子回测结果



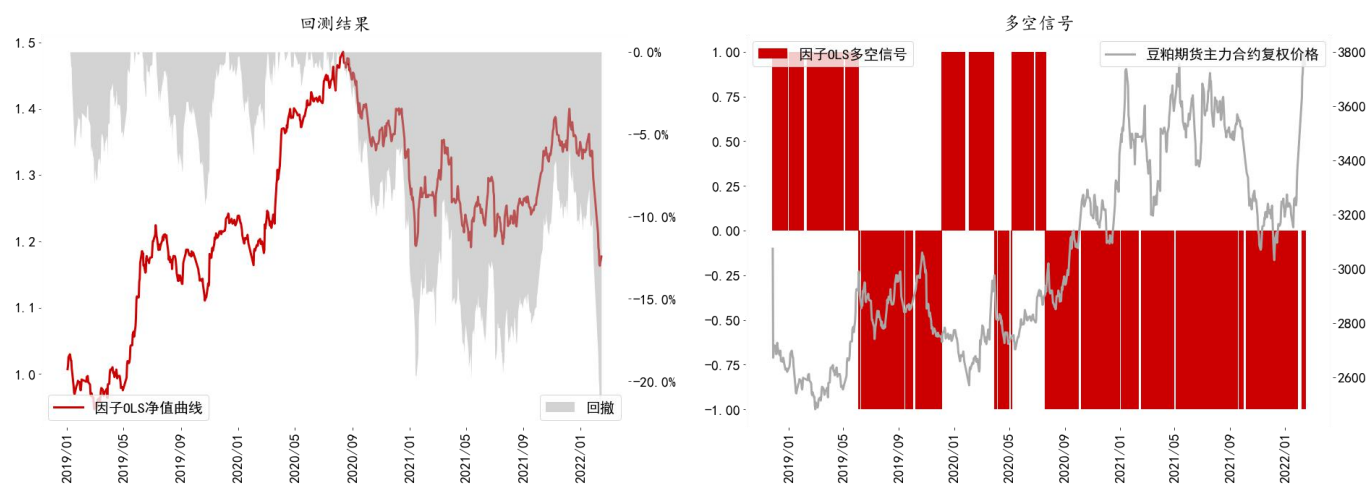
资料来源: 东证衍生品研究院

图表 25: 进口数量:豆饼、豆粕:进口地:中国台湾:累计同比单因子回测结果



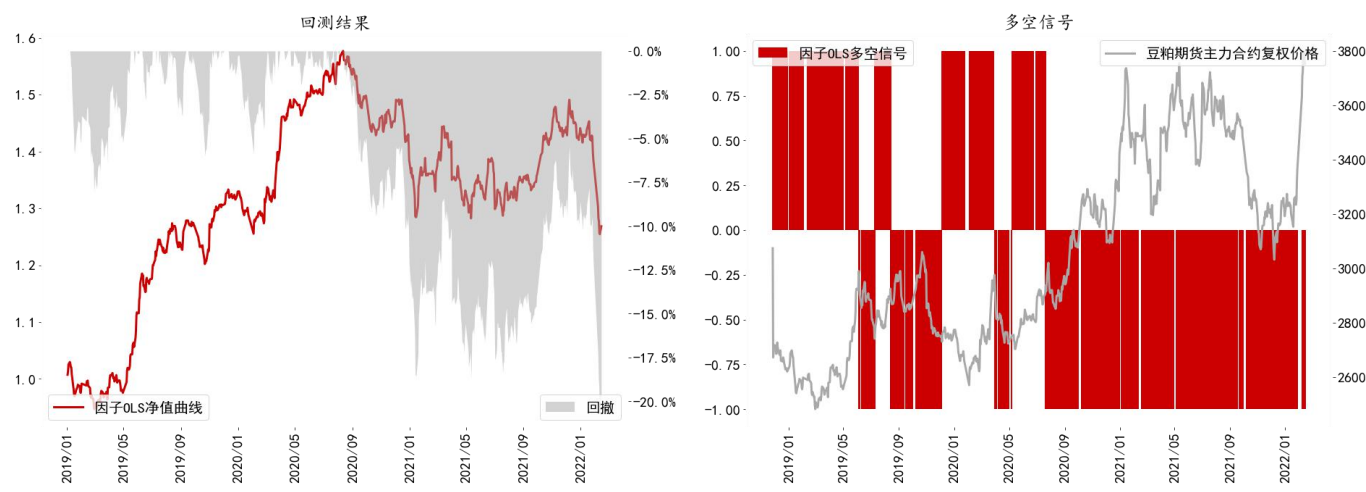
资料来源: 东证衍生品研究院

图表 26: 进口数量:豆饼、豆粕:进口地:印度:累计同比单因子回测结果



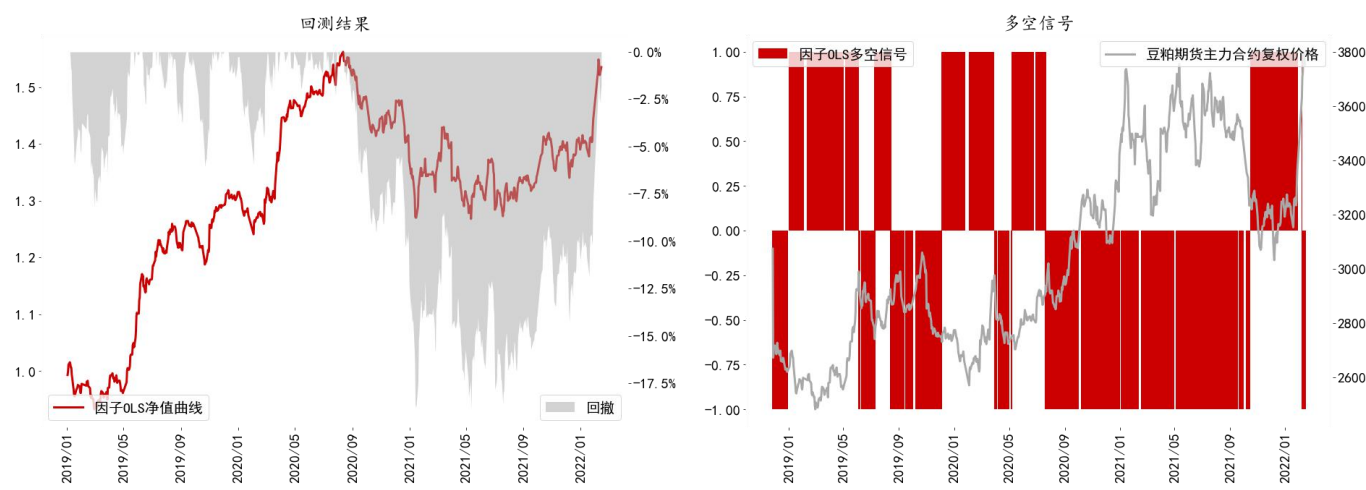
资料来源: 东证衍生品研究院

图表 27：进口到港数量:豆粕:华东区:当期值单因子回测结果



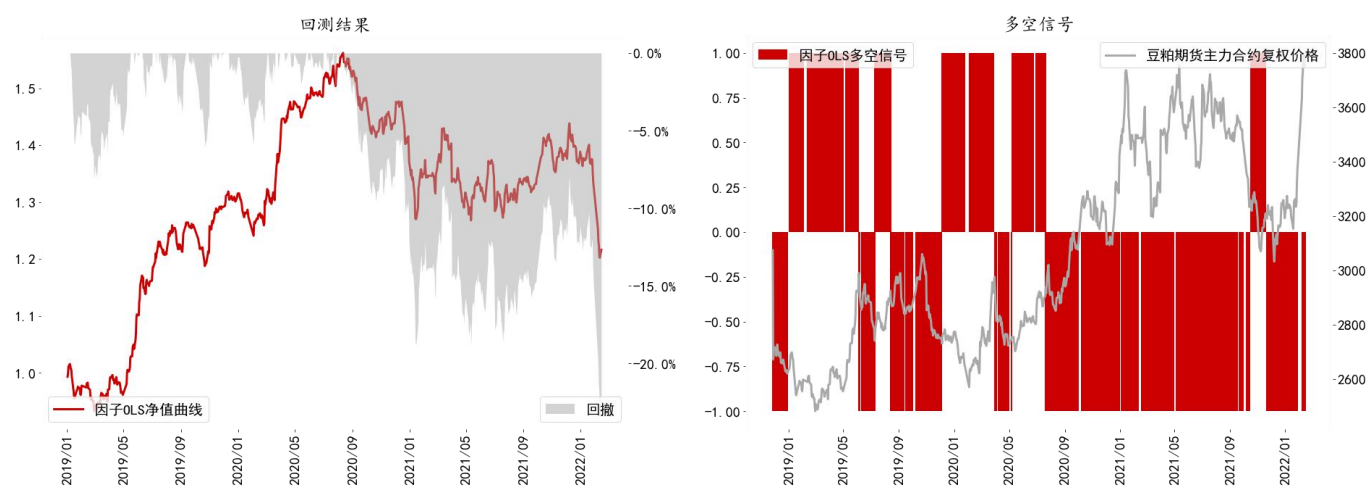
资料来源：东证衍生品研究院

图表 28：生猪存栏因子回测结果



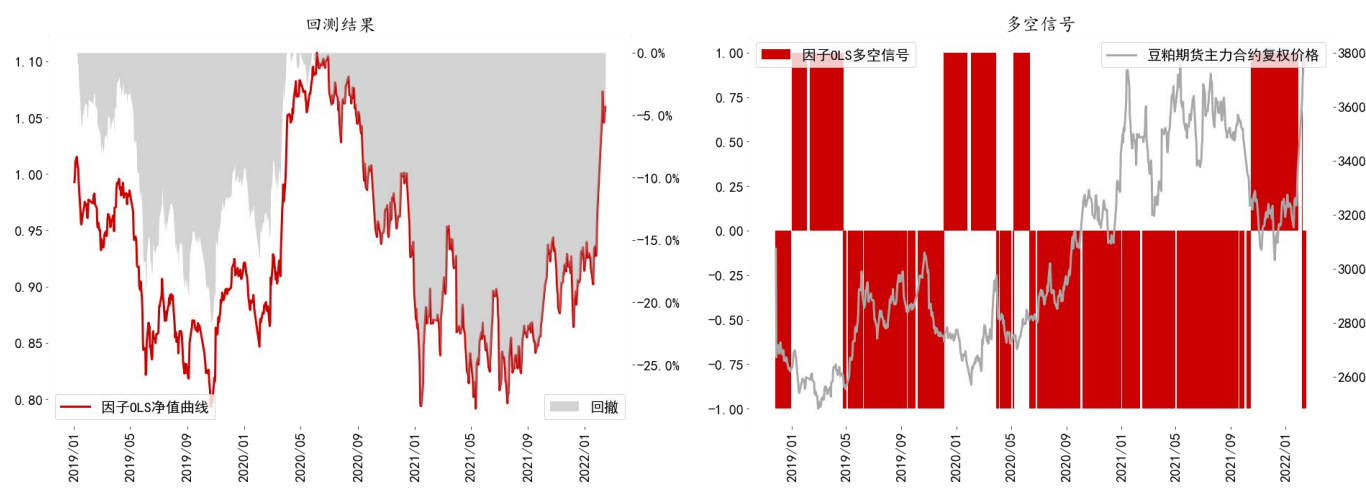
资料来源：东证衍生品研究院

图表 29：生猪存栏变化率:环比增减单因子回测结果



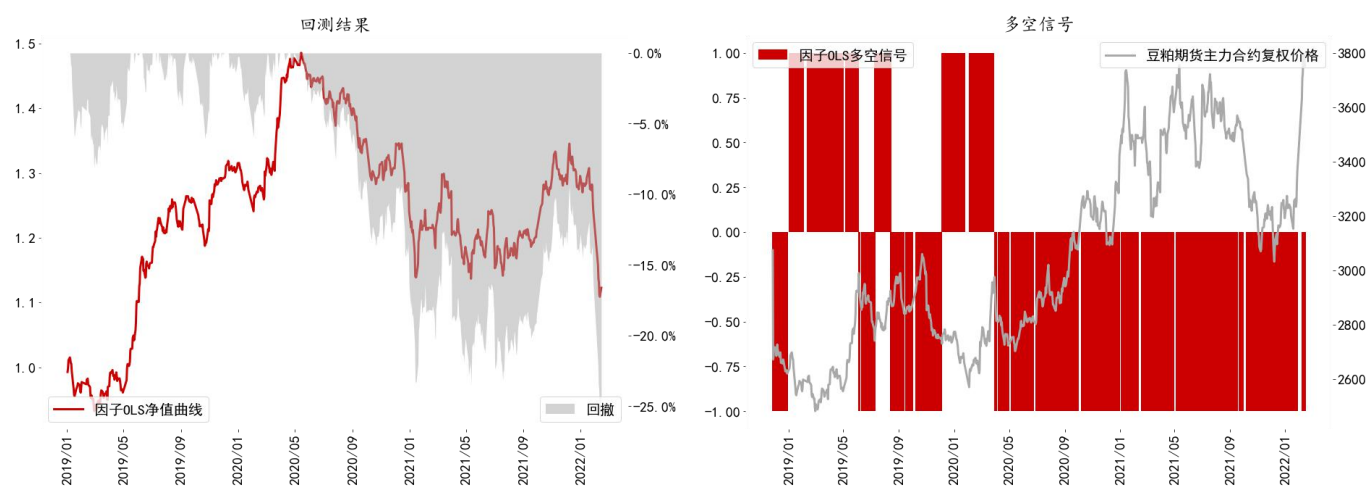
资料来源：东证衍生品研究院

图表 30：生猪存栏变化率:同比增减单因子回测结果



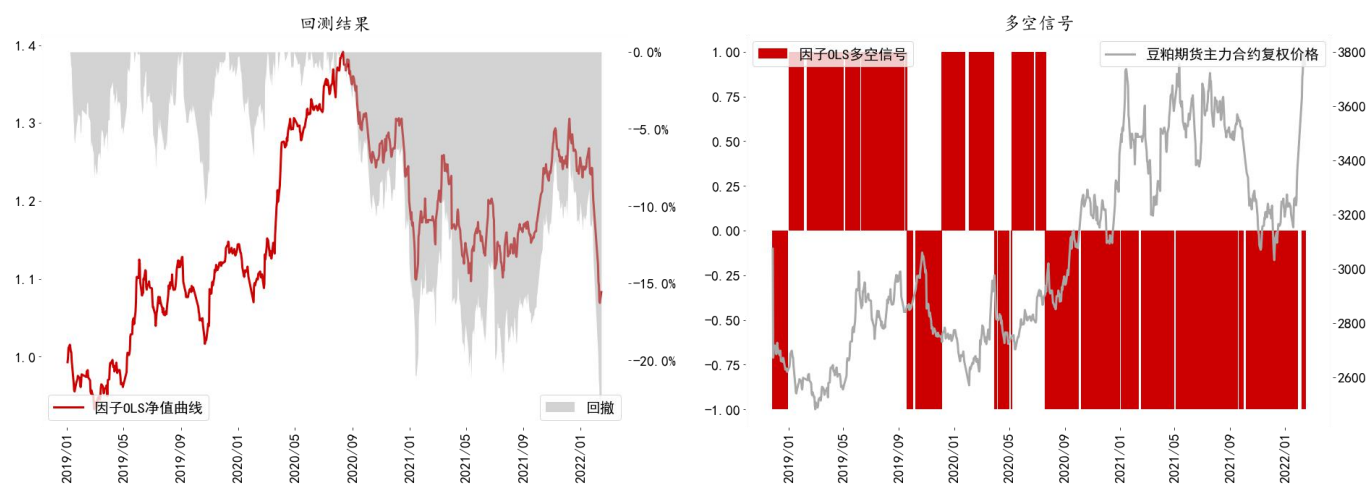
资料来源：东证衍生品研究院

图表 31：生猪出栏:同比单因子回测结果



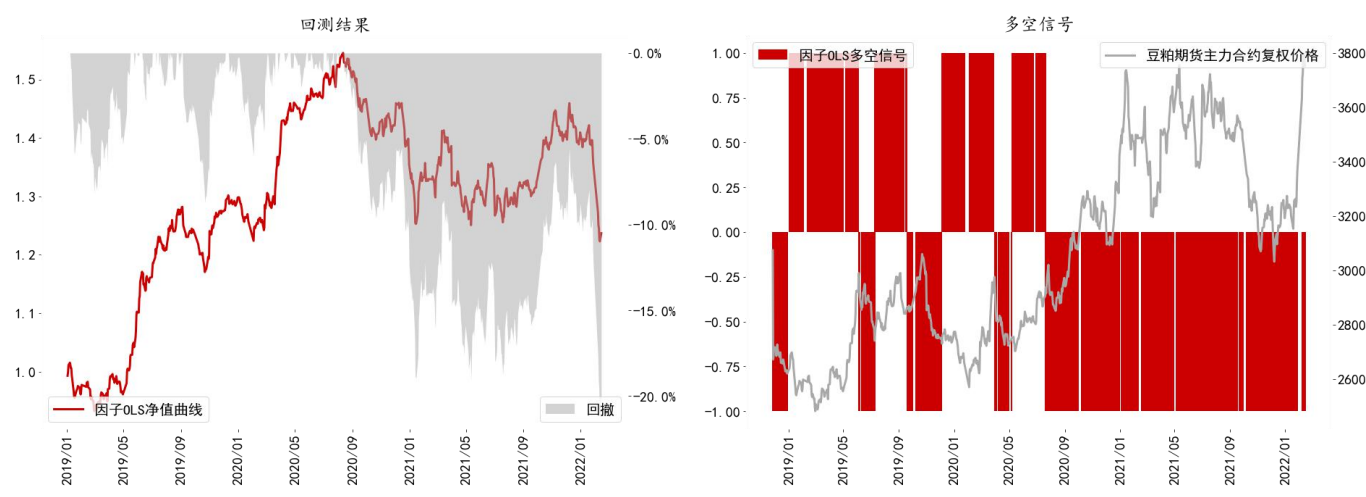
资料来源：东证衍生品研究院

图表 32：豆粕:库存:广西地区单因子回测结果



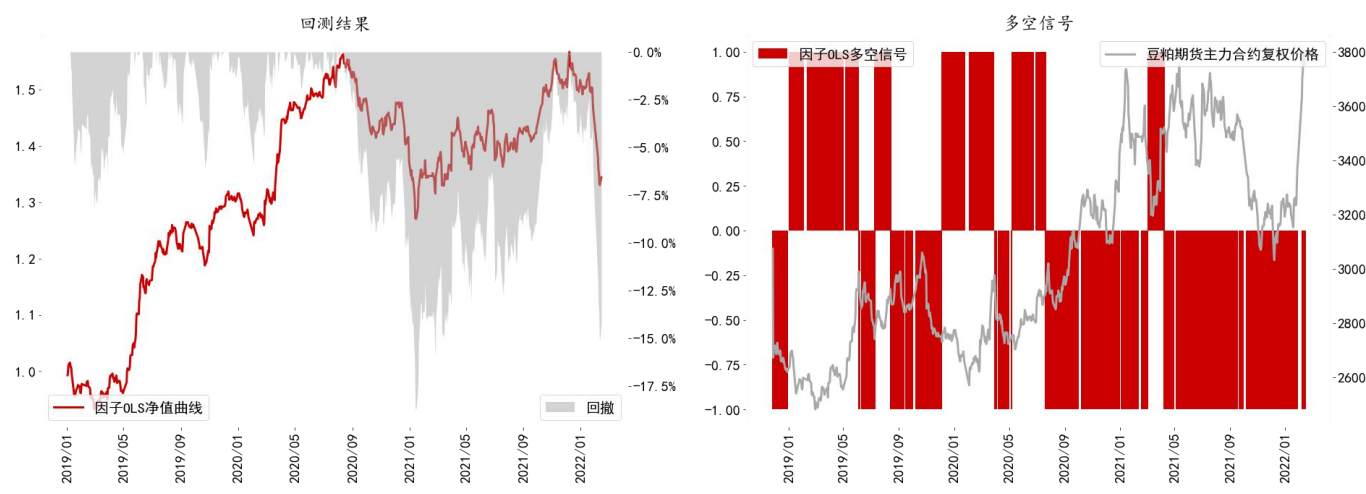
资料来源：东证衍生品研究院

图表 33: 豆粕:库存:东北地区单因子回测结果



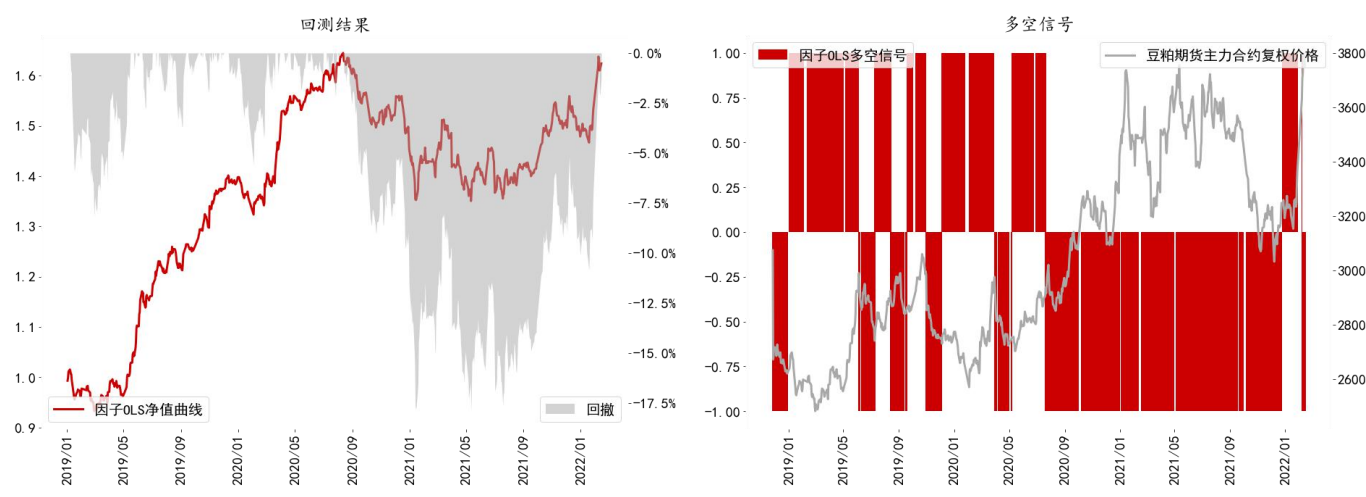
资料来源: 东证衍生品研究院

图表 34: 豆粕:卖持量:前10名会员合计(季度环比)单因子回测结果



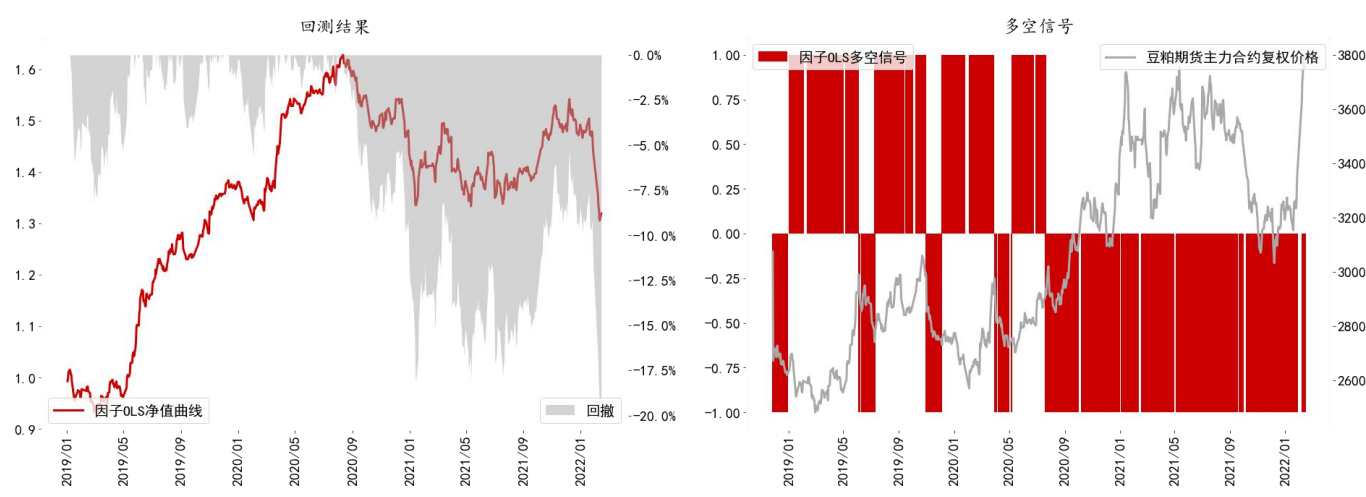
资料来源: 东证衍生品研究院

图表 35: 油厂豆粕报价:辽宁 (同比) 单因子回测结果



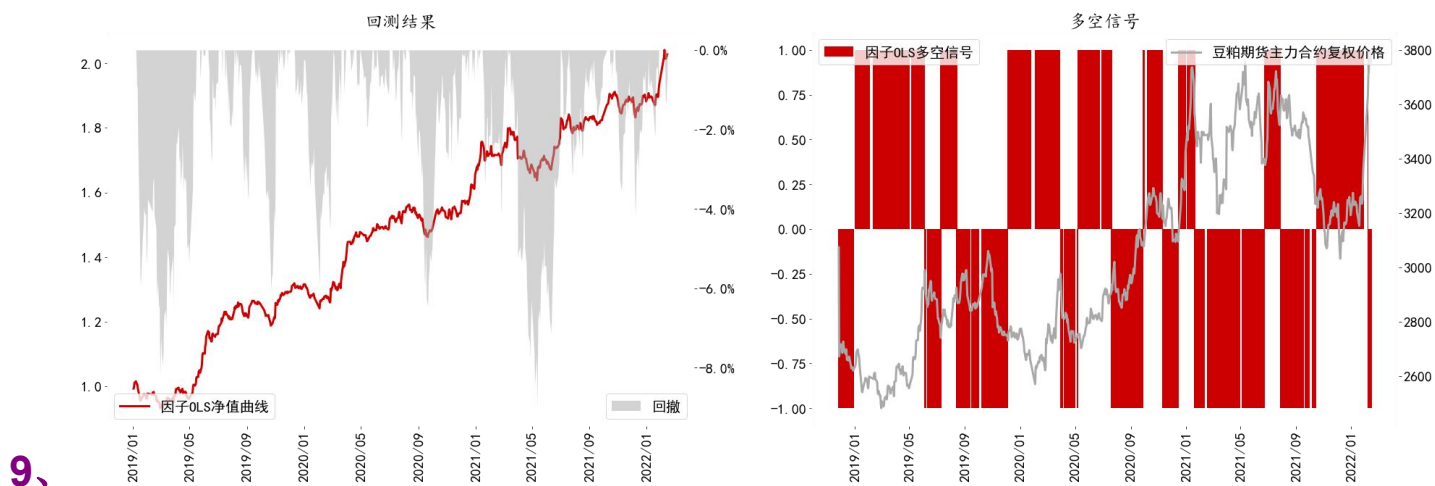
资料来源: 东证衍生品研究院

图表 36: 油厂豆粕报价:广东 (季度环比) 单因子回测结果



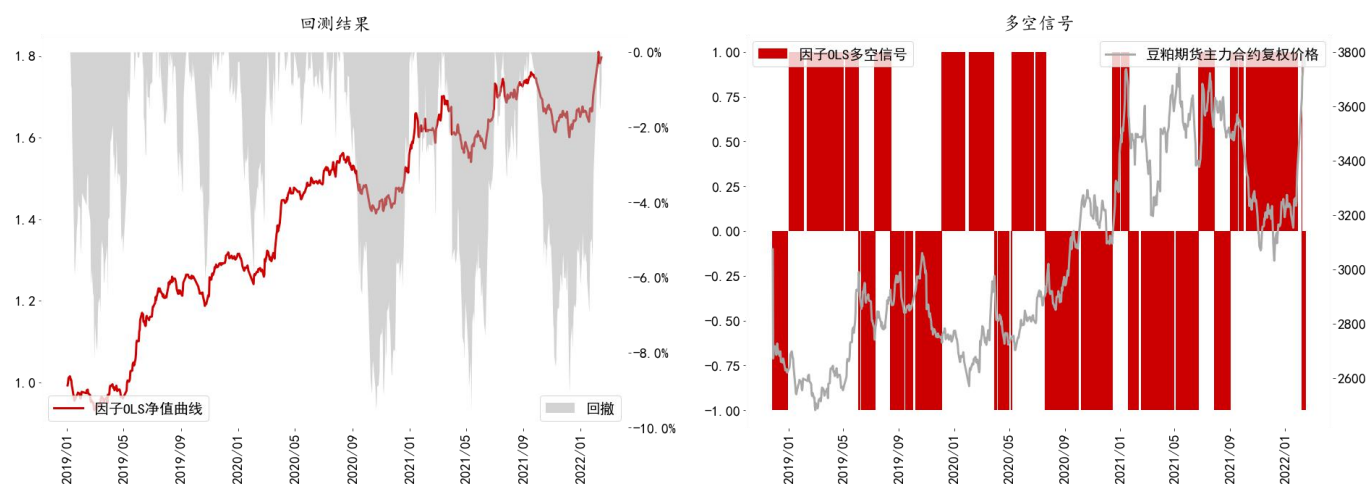
资料来源: 东证衍生品研究院

图表 37: 国内 1135 家饲料产量:总饲料 (月度环比) 单因子回测结果



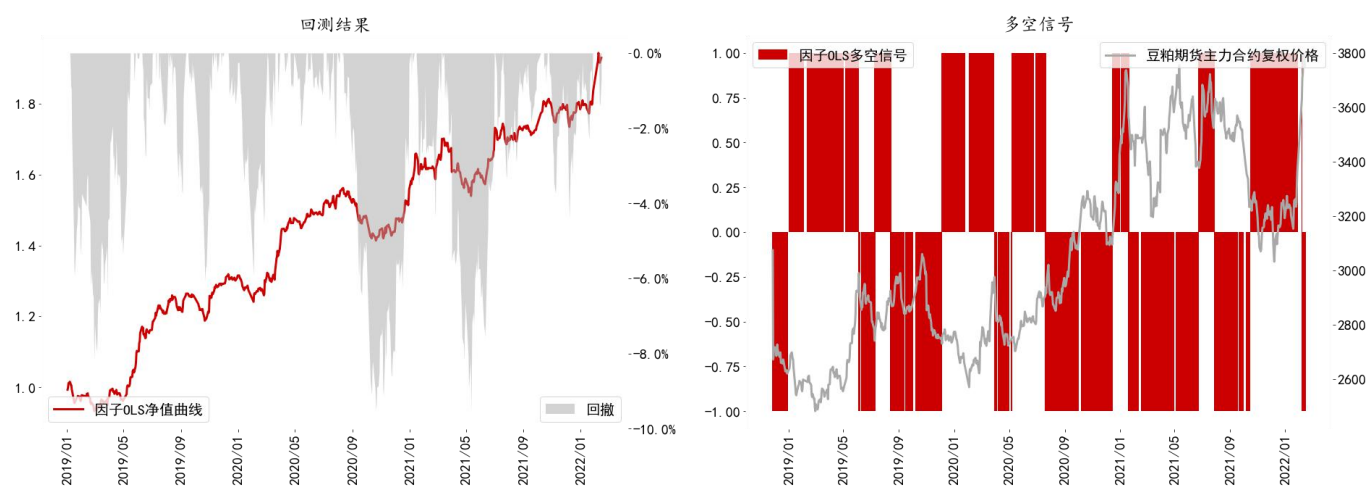
资料来源: 东证衍生品研究院

图表 38: 国内 1135 家饲料产量:总饲料 (季度环比) 单因子回测结果



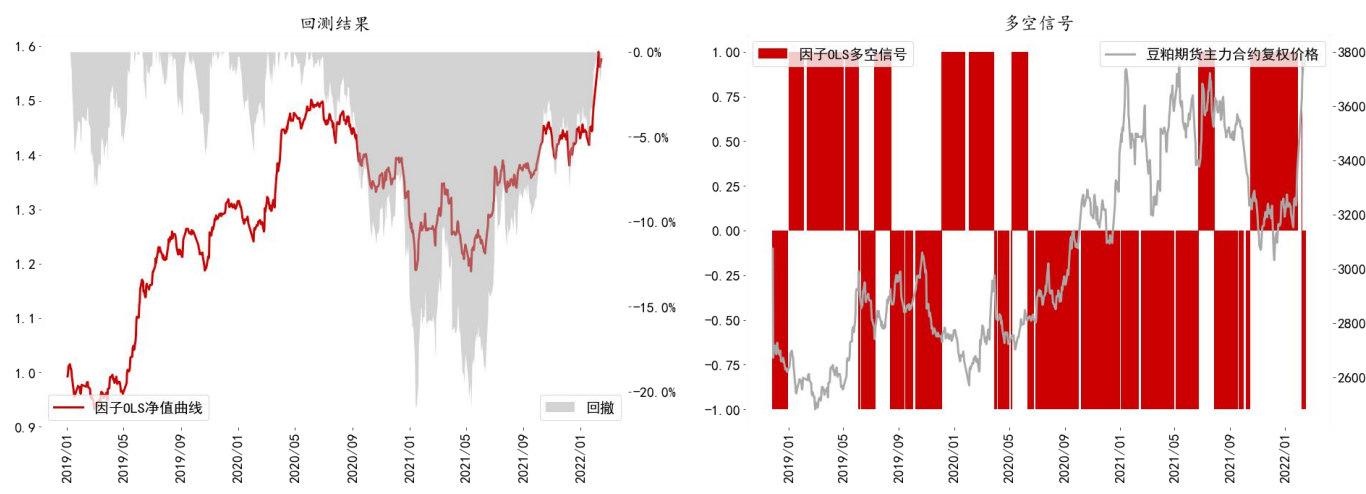
资料来源: 东证衍生品研究院

图表 39: 国内 1135 家饲料产量:禽料 (季度环比) 单因子回测结果



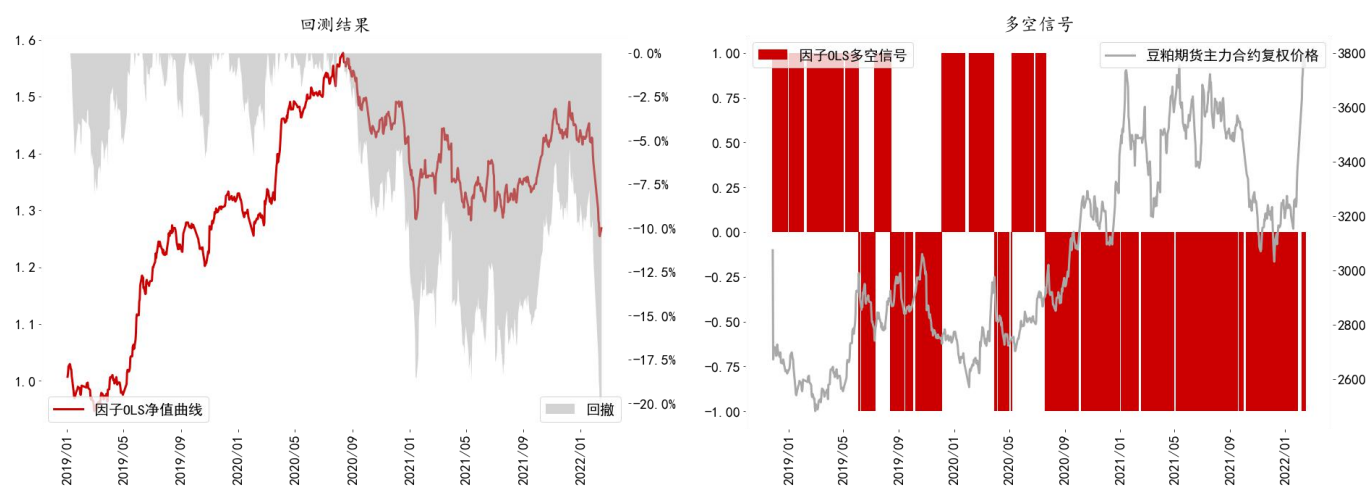
资料来源: 东证衍生品研究院

图表 40: 全国豆粕未执行合同 (同比) 单因子回测结果



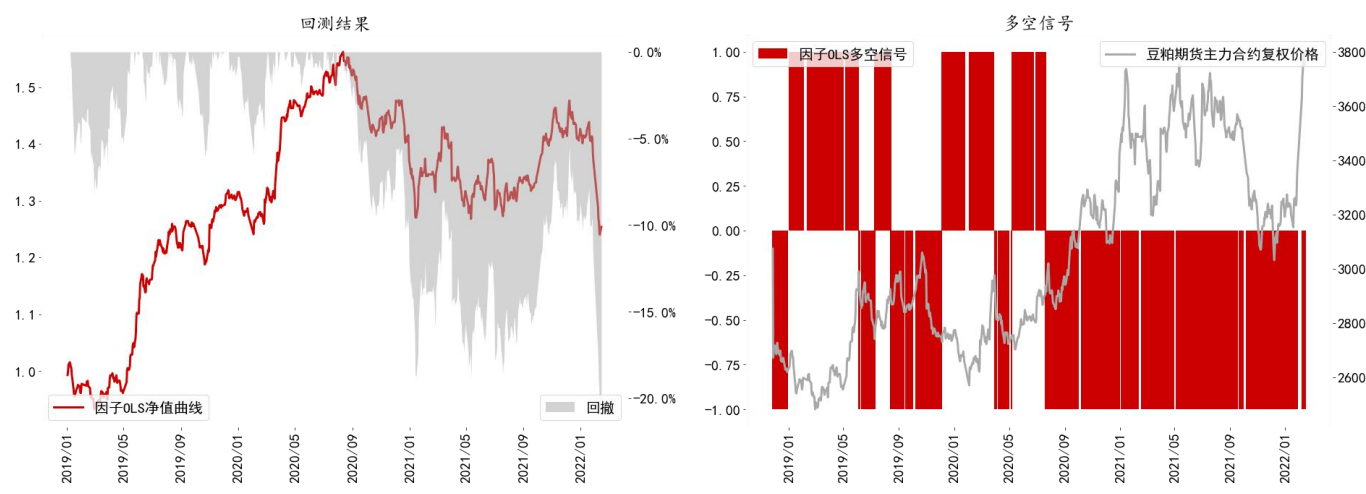
资料来源: 东证衍生品研究院

图表 41: 豆粕:库存:广西地区 (同比) 单因子回测结果



资料来源: 东证衍生品研究院

图表 42: 全国豆粕周度产量 (同比) 单因子回测结果



资料来源: 东证衍生品研究院

期货走势评级体系（以收盘价的变动幅度为判断标准）

走势评级	短期（1-3个月）	中期（3-6个月）	长期（6-12个月）
强烈看涨	上涨 15%以上	上涨 15%以上	上涨 15%以上
看涨	上涨 5-15%	上涨 5-15%	上涨 5-15%
震荡	振幅-5%-+5%	振幅-5%-+5%	振幅-5%-+5%
看跌	下跌 5-15%	下跌 5-15%	下跌 5-15%
强烈看跌	下跌 15%以上	下跌 15%以上	下跌 15%以上

上海东证期货有限公司

上海东证期货有限公司成立于2008年，是一家经中国证券监督管理委员会批准的经营期货业务的综合性公司。东证期货是东方证券股份有限公司全资子公司，注册资本金23亿元人民币，员工近600人。公司主要从事商品期货经纪、金融期货经纪、期货投资咨询、资产管理、基金销售等业务，拥有上海期货交易所、大连商品交易所、郑州商品交易所和上海国际能源交易中心会员资格，是中国金融期货交易所全面结算会员。公司拥有东证润和资本管理有限公司，上海东祺投资管理有限公司和东证期货国际（新加坡）私人有限公司三家全资子公司。

东证期货以上海为总部所在地，在大连、长沙、北京、上海、郑州、太原、常州、广州、青岛、宁波、深圳、杭州、西安、厦门、成都、东营、天津、哈尔滨、南宁、重庆、苏州、南通、泉州、汕头、沈阳、无锡、济南等地共设有33家营业部，并在北京、上海、广州、深圳多个经济发达地区拥有134个证券IB分支机构，未来东证期货将形成立足上海、辐射全国的经营网络。

自2008年成立以来，东证期货秉承稳健经营、创新发展的宗旨，坚持市场化、国际化、集团化的发展道路，打造以衍生品风险管理为核心，具有研究和技术两大核心竞争力，为客户提供综合财富管理平台的一流衍生品服务商。

分析师承诺

王冬黎、谢怡伦

本人具有中国期货业协会授予的期货执业资格或相当的专业胜任能力，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。本人不曾因，不因，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

本报告由上海东证期货有限公司（以下简称“本公司”）制作及发布。

本研究报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本研究报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，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。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的报告之外，绝大多数研究报告是在分析师认为适当的时候不定期地发布。

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，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买投资标的的邀请或向人作出邀请。

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任何损失负任何责任，投资者需自行承担风险。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

如需引用、刊发或转载本报告，需注明出处为东证期货研究所，且不得对本报告进行任何有悖原意的引用、删节和修改。

东证衍生品研究院

地址：上海市中山南路318号东方国际金融广场2号楼21楼

联系人：梁爽

电话：8621-63325888-1592

传真：8621-33315862

网址：www.orientfutures.com

Email：research@orientfutures.com