

交易咨询资格号：  
证监许可[2012]112

金融工程  
专题报告

2024 年 3 月 13 日

### 分析师

杨旻

金融工程首席分析师

期货从业资格号：F03096114

交易咨询从业证书号：Z0020184

时翔宇

金融工程分析师

期货从业资格号：F03104321

交易咨询从业证书号：Z0019649

联系人：杜思嘉

期货从业资格号：F03103175

联系人：李开来

期货从业资格号：F03124866

联系电话：021-61625026

邮箱：ztqh\_sh@163.com

客服电话：400-618-6767

### 报告概述

- 现有通用大模型正解决人们真正意义上接受自然语言处理和机器学习应用于金融研究的关键问题——建立心理共识。本文对 Cao 和 Zhai (2023) 部分研究成果进行翻译，介绍了通用大模型建立这种心理共识的基础性逻辑，阐述通用大模型如何为人们深入研究金融问题逐步消减技术障碍。本文通过讯飞星火平台开展商品策略设计的文本交互，得到了平台关于商品套利品种、套利逻辑、套利策略、策略 Python 代码（需进行修正）及策略调整优化方向的解答。
- **风险提示：**基于历史研究总结的相关规律未来可能存在失效的风险；模型风险。

### 中泰期货研究所



### 中泰期货服务号



## 弥合鸿沟：大模型对金融研究的影响

### ——量化技术系列研究之六

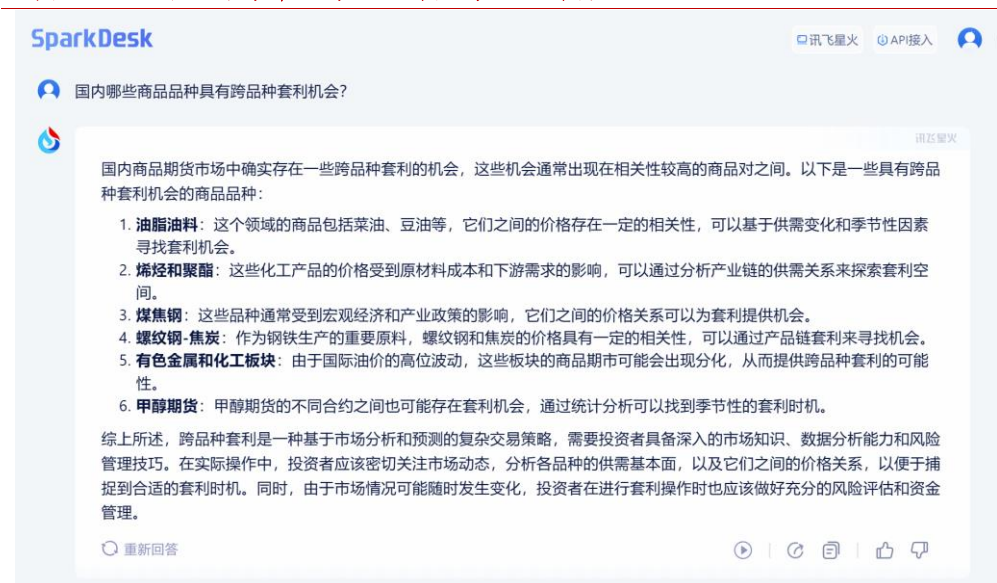
#### 一、建立心理共识

- 毫无疑问，无法解释的高性能和偶尔出现的幼稚错误都是研究中无法容忍的问题，特别是在风险敏感领域，如自动驾驶汽车、医疗保健和金融。在金融领域，这是经常被忽视但始终阻碍人们真正意义上接受机器学习或大模型应用于金融研究的关键问题——建立心理共识。
- Langer 等（1978）进行了著名的“复印机研究”，研究人员在图书馆排队复印的队伍中试图插队，并随机将以下 3 个提问中的 1 个抛给询问对象：（1）（请求）“对不起，我有 5 页，我可以先用复印机吗？”（2）（请求+真理由）“对不起，我有 5 页，我可以先用复印机吗？因为我很着急。”（3）（请求+假理由）“对不起，我有 5 页，我可以先用复印机吗？因为我必须复印一下。”很显然，第 3 个提问和给出的理由是较为荒谬的，然而令人惊讶的是，实验结果却是有利的：（1）在第 1 种提问场景下，60% 的询问对象让研究人员插队；（2）在第 2 种提问场景下，94% 的询问对象让研究人员插队；（3）在第 3 种提问场景下，93% 的询问对象让研究人员插队。
- 这项研究揭示了影响人类行为的最具影响力的词之一：“因为”。Langer 等（1978）研究表明，当人们面对不熟悉的提问、行为或概念时，他们往往只是需要对方有一个理由或借口，就容易接受。当人们普遍接受一种类型“理由”来作为解释时，对行为的隐性接受共识就建立起来了。很明显，这种共识尚未在金融机器学习研究领域得到广泛建立。
- 以 GPT-4 为代表的大模型已逐步搭建机器学习技术应用于金融研究领域的心理共识，有以下原因：一方面，当我们向以 GPT-4 为代表的大模型询问金融研究问题时，往往会得到并非特别荒谬的“理由”来作为解释，支持和丰富了我们对于未知问题的探索，这搭建了人们对于机器学习应用于金融研究的心理共识。另一方面，以 GPT-4 为代表的大模型在金融专业和学术测试中表现出色，这是人类也很难达到的水准，一场无法阻止的新智能工业革命正在发生，并已产生了广泛的商业应用。

## 二、消减技术障碍

- **第一**，对于金融研究人员来说，自然语言处理可以帮助他们在许多领域解决困难的研究问题，但对于缺乏专业计算机编程背景的研究人员来说，应用自然语言处理通常是困难的。对此，大模型可能是较为合适的工具，它在语言环境中交互对话，并为人们深入研究主题消减技术障碍。
- 当人们给出金融领域适当的提问指令，包括情绪分析（从新闻、报道、社交媒体中提取情感价值）、关键词分析、事件解读（探寻某类型事件历史上对资产价格影响）、ESG 分析等，通用大模型应用将给出相应的定性或定量结论，感兴趣的读者可以参考 Cao 和 Zhai（2023）文中示例进行提问。
- **第二**，通用大模型应用可以辅助开展量化研究设计。下面基于讯飞星火（SparkDesk）探讨商品期货策略的研发思路和方案设计，仅供参考，不构成任何交易建议。
- 一是，我们询问国内哪些商品品种具有跨品种套利机会，从回复来看，平台主要是从商品基本面关联角度出发给出跨品种套利思路。

图表 1：提问：国内哪些商品品种具有跨品种套利机会？



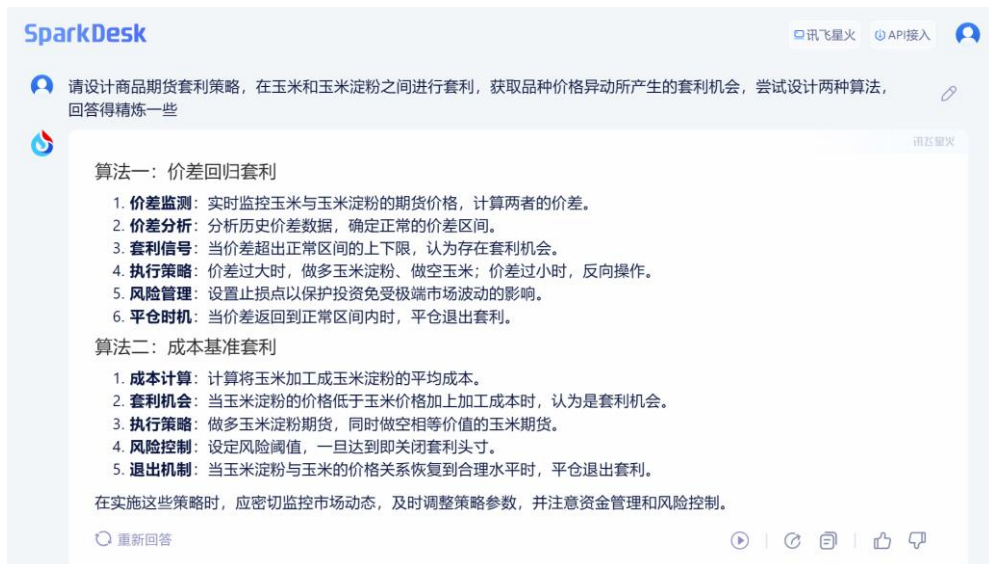
来源：讯飞星火，中泰期货研究所整理

- 二是，我们结合第一个问题，提问除了已给出的基于基本面的套利逻辑，有基于量价的跨品种套利逻辑吗？从回复来看，回复较为笼统但的确有一定的启发性，例如给出了基于时间序列方法构建套利策略的具体思路。

**图表 2：提问：有基于量价的跨品种套利逻辑吗？**

来源：讯飞星火，中泰期货研究所整理

- 三是，我们提问设计玉米和玉米淀粉跨品种套利策略，策略思路基于品种价格异动所产生的潜在套利机会。讯飞星火给出了价差回归套利和成本基准套利两种算法，总体上看其给出的策略逻辑虽然简洁，但在理论上和操作上均具有一定可行性，且策略描述清晰。

**图表 3：提问：请设计玉米和玉米淀粉套利策略**

来源：讯飞星火，中泰期货研究所整理

- 我们来看讯飞星火给出的价差回归套利具体流程：（1）价差监测：实时监控玉米与玉米淀粉的期货价格，计算两者的价差。（2）价差分析：分析历史价差数据，确定正常的价差区间。（3）套利信号：当价差超出正

常区间的上下限，认为存在套利机会。(4) 执行策略：价差过大时，做多玉米淀粉、做空玉米；价差过小时，反向操作。(5) 风险管理：设置止损点以保护投资免受极端市场波动的影响。(6) 平仓时机：当价差返回到正常区间内时，平仓退出套利。

- 四是，我们提问请针对算法一（价差回归套利）给出 Python 回测代码，图表 4 是讯飞星火回复内容，可看出，所给出代码包括但不限于以下问题需进行修正：(1) 回测过程未进行样本滚动，即在“价差”计算过程中使用了未来数据；(2) 策略收益使用玉米淀粉收益率\*仓位，不符合策略逻辑；(3) t 期末信号生成后，回测执行交易的初始时点应为 t+1 期，而不是 t 期；(4) 回测过程未考虑总资金仓位和交易成本等。

图表 4：提问：针对算法一给出回测代码（仅为示例，需修正，不构成任何交易建议）

SparkDesk

讯飞星火 API接入

请针对算法一给出python回测代码

讯飞星火

以下是针对算法一的Python回测代码示例：

```
python
import pandas as pd
import numpy as np

# 读取历史数据
data = pd.read_csv('historical_data.csv')

# 计算价差
data['spread'] = data['corn_price'] - data['cornstarch_price']

# 设置正常价差区间
upper_limit = data['spread'].mean() + 2 * data['spread'].std()
lower_limit = data['spread'].mean() - 2 * data['spread'].std()

# 初始化交易信号和持仓状态
data['signal'] = 0
data['position'] = 0

for i in range(1, len(data)):
    if data['spread'][i] > upper_limit:
        data['signal'][i] = -1 # 做空玉米淀粉，做多玉米
    elif data['spread'][i] < lower_limit:
        data['signal'][i] = 1 # 做多玉米淀粉，做空玉米
    else:
        data['signal'][i] = 0 # 无交易信号

# 根据交易信号更新持仓状态
data['position'] = data['signal'].cumsum()

# 计算策略收益
data['strategy_return'] = data['cornstarch_price'].pct_change() * data['position']

# 计算累积收益
data['cumulative_return'] = (1 + data['strategy_return']).cumprod()

# 输出回测结果
print("策略收益率: ", data['strategy_return'].mean())
print("策略波动率: ", data['strategy_return'].std())
print("最大回撤: ", np.max(np.maximum.accumulate(data['cumulative_return']) - data['cumulative_return']))
```

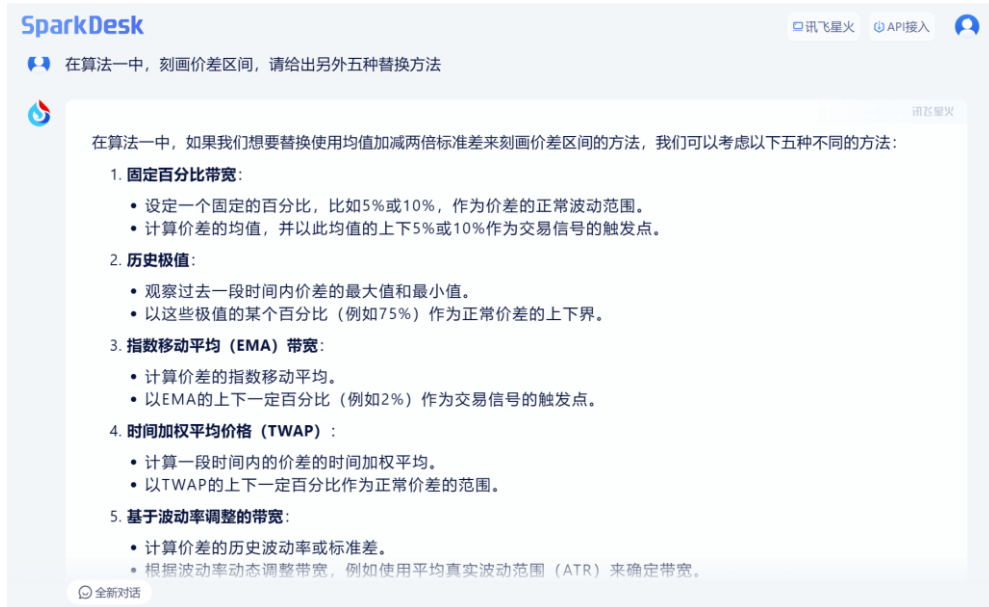
请注意，上述代码仅为示例，实际应用中需要根据具体情况进行修改和优化。

来源：讯飞星火，中泰期货研究所整理



- 五是, 细节优化和调整。算法一采用均值加减 2 倍标准差刻画价差区间, 我们针对这一算法细节, 提请给出五种刻画价差区间的替换方法, 讯飞星火所给出的建议具有一定参考价值, 包括固定百分比带宽、历史极值、指数移动平均带宽、时间加权平均价格、基于波动率调整的带宽。

**图表 5: 提问: 请给出另外五种刻画价差区间的替换方法**



来源: 讯飞星火, 中泰期货研究所整理

### 三、主要结论

- 一是人们向通用大模型应用询问问题后, 得到的回复比较容易接受, 因为回复中通常有较详细的“理由”, 这在一定程度上满足了人们的询问需求, 机器学习技术应用于金融研究领域的心理共识也逐渐建立起来了。
- 二是大模型可能是较为合适金融领域研究人员处理文本信息、提取有价值内容的工具, 它在语言环境中与人交互对话, 并为人们深入开展情绪分析、关键词分析、事件解读、ESG 分析等研究消减技术障碍。
- 三是通用大模型应用可以辅助开展量化研究设计。本文以讯飞星火平台为研究载体, 通过语言交互得到了关于商品套利品种、套利逻辑、套利策略的文本解答; 该平台根据作者提问, 给出了基于价差回归套利算法的玉米-玉米淀粉跨品种套利 Python 代码 (需进行修正); 此外, 在作者对细节算法优化要求下, 平台给出了 5 种刻画价差区间的替换方法。

■ 参考文献：

- [1] Cao Y, Zhai J. Bridging the gap: the impact of ChatGPT on financial research[J]. Journal of Chinese Economic and Business Studies. 2023, 21(2): 177-191.
- [2] Langer E J, Blank A, Chanowitz B. The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction[J]. Journal of Personality and Social Psychology. 1978, 36(6): 635-642.

■ 风险提示：基于历史研究总结的相关规律未来可能存在失效的风险；模型风险。

## 免责声明：

中泰期货股份有限公司（以下简称本公司）具有中国证券监督管理委员会批准的期货交易咨询业务资格（证监许可〔2012〕112）。本报告仅限本公司客户使用。

本公司不会因接收人收到本报告而视其为客户。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交易建议，本公司不对任何人因使用本报告中的内容所导致的损失负任何责任。市场有风险，投资需谨慎。

本报告所载的资料、观点及预测均反映了本公司在最初发布该报告当日分析师的判断，是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可在不发出通知的情况下发出与本报告所载资料、意见及推测不一致的报告，亦可因使用不同假设和标准、采用不同观点和分析方法而与本公司其他业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。本公司并不承担提示本报告的收件人注意该等材料的责任。

本报告的知识产权归本公司所有，未经本公司书面许可，任何机构和个人不得以任何方式进行复制、传播、改编、销售、出版、广播或用作其他商业目的。如引用、刊发、转载，需征得本公司同意，并注明出处为中泰期货，且不得对本报告进行有悖原意的引用、删节和修改。