

# 从主观经验到数字化智力资产架构开发\* ——论期货公司研究体系的转型升级之路

南华期货股份有限公司研究院 顾双飞 胡乐克

期货研究作为金融市场的核心驱动力之一，长期以来高度依赖分析师的个人经验与能力。然而，传统研究模式面临人才流动性高、知识传承困难、研究质量参差不齐等挑战。南华期货研究院通过构建南华期货投研一体化平台（Nanhua Trading Analysis System，以下简称NHTA），探索了一条将主观经验转化为数字化因子的创新路径，实现了从“人力资源”到“智力资产”的转型升级。这一变革不仅提升了研究效率与预测准确性，更通过数据资产、逻辑资产和模型资产的系统化沉淀，为期货公司研究体系的可持续发展提供了全新范式。本文将深入探讨这一转型的实践路径、核心价值及未来方向，为行业提供可借鉴的解决方案。

## 一、论期货公司研究部门的核心资产

### （一）期货公司研究部门的传统核心资产

一直以来，每当我们提到期货公司研究部门的核心资产，毫无疑问大家都会第一时间想到“人”。的确，一个高质量的研究团队，无论在期货行业，抑或证券行业，其核心关键要素都是一批高水平的分析师队伍。但是，我们

也应该深刻地意识到，不同于传统固定资产，人力资产本身就是一种高度流动性的资产，这种传统的人力资产结构，高度依赖分析师的成长性和稳定性。

首先，在一个传统的研究团队里，每一个分析师都需要做研究报告所需的所有流程任务，即从数据提取、数据清洗整理、指标生成、图表生成到最后的研报撰写。由于不同分析师的能力禀赋和个性特点都有差异，导致所呈现出来的研究成果也参差不齐。比如，我们为了产出一篇高质量的研究报告，分析师首先需要搜集并整理好对应品种的所有基本面信息及内在逻辑关系，然后再对这些基本面信息进行数据处理工作（依照方式的复杂程度不同，使用入门的excel到较为困难的python, matlab等编程语言），接着将做好的数据依照不同的用途绘制出不同的表格来阐述其因果关系，最后将整个逻辑推导的过程转换成研究报告。当然如果再进一步，某些分析师还熟练掌握了各种金融建模工具，就可以在报告中将研究经验和验证过的因果关系转换成具体的预测模型，提高研报的整体质量，使得报告更具说服力。

\* 本作品在 2025 年《期货与金融衍生品》征文活动中荣获二等奖。收稿时间为 2025 年 6 月。

简而言之，高质量的分析师需要掌握各个方面知识和技术，成为一个没有短板的全方位人才。而从一个初级分析师成长到这样的高度，需要一个漫长的培养过程。中间不单单是经验的累积，还需要对各种技能的学习掌握。这些技能和经验全部被赋予到了分析师这个“人力资产”上，而就像前文提到，资产一旦流动，将会对研究部门造成重大的影响。近年来期货公司研究部门向行业（产业机构、期现公司、私募基金等）输送了大批的优秀人才，但其本身却又很难沉淀下来一些具化的研究资产，这也是困扰期货公司研究部门多年的问题。

这就引发了我们进一步的思考：能否将这些人力资产，转变为一种智力资产，或者是一种更为明确的能看得见摸得着的研究型固定资产？本研究提出以下核心观点：期货公司研究部门的核心资产形态，必然从依托个体的人力资本，向体系化的智力资本演进。该过程旨在通过持续的技术积累与方法论工具化，将隐性知识显性化、体系化，最终固化为企业核心知识资产。此类资产通过整合宏观、中观及产业微观的分析框架，构成了持续为客户输出高效、稳定研究服务的坚实基础。

## （二）智力型固定资产-期货公司研究部门的新核心资产

智力资产结构是期货公司研究部门将研究能力和成果系统化、具象化的重要框架，旨在将抽象的研究力转化为可量化、可复用、可迭代的数据资产、逻辑资产和模型资产，从而提升研究效率、强化决策支持能力并增强机构的核心竞争力。

数据资产是智力资产的基础层，涵盖期货

公司研究过程中积累的原始数据、清洗后的结构化数据以及衍生指标库。例如，宏观经济指标、产业链供需数据、市场情绪指标、高频交易数据等，通过标准化存储和动态更新，形成可快速调用的数据资源池。数据资产的价值在于其时效性、准确性和多维关联性，为后续分析提供扎实的输入。

逻辑资产是研究方法的沉淀，包括分析框架、因果推理链条、市场规律总结等。例如，基于历史行情的周期规律提炼、事件驱动分析的标准化流程、产业链上下游传导逻辑的建模等。逻辑资产将分析师的经验转化为可复用的方法论，降低重复劳动成本，同时确保研究结论的一致性和可验证性。

模型资产是智力资产的顶层应用，通过算法或量化工具将数据与逻辑结合，形成可执行的策略或预测工具。例如，价格预测模型、套利策略模型、风险对冲模型等。模型资产需具备参数可调、场景适配和持续优化的特性，直接服务于交易、风控和客户服务。

期货公司通过构建智力资产结构，能够实现研究能力的积累与传承，推动从以人为本的“经验驱动”向平台优先的“数据驱动+逻辑驱动”转型，最终提升研究的商业价值与市场影响力。

这样的一种研究型固定资产，需要一个综合投研平台为载体，去承载这样的知识库框架，这亦是期货公司研究体系的转型升级必经之路。

## （三）投研平台类固定资产的功能和价值

我们已经提到，这个投研平台类固定资产的核心价值是将分析师的研究类资产沉淀。这

类研究类资产不仅仅包括简单的指标和数据，更包含了分析师的思考过程和逻辑关系提炼。这就对投研平台的功能提出了以下几个方面的要求：

- 1.全面的数据搜集能力（是否可以整合来自各个渠道的数据资源）
- 2.完善的数据资源归类体系（目录架构设置）
- 3.完善的数据加工处理能力（不仅包含简单的常规加减乘除计算，更包含数据对齐、拼接、相关性、拟合方程、标准差计算、残差分析等复杂的统计量化功能）
- 4.数据预测模型（根据不同的预测规则对数据做量化预测）
- 5.数据表格（可以输入一些主观经验类数据，以及呈现一些混合表格）
- 6.文本类输出功能（编辑研报、制作PPT）
- 7.将分析师计算推导过程记录沉淀（计算公式，逻辑图表）

投研平台完善的底层功能设计，以及更贴近分析师研究体系的架构设计一定是项目的基础，否则分析师就缺乏持续使用的驱动；研究部门在投研平台项目推动时也要有持续且充分的执行力，在分析师研究方式转化期，势必会增加分析师的工作量，且部分分析师出于“私有资产”的保护意识，也会缺乏将自己研究资源上传投研平台的驱动，需要不断加以引导。

## 二、当前期货公司研究体系转型升级难点

### （一）期货分析师研究能力层次不齐

中国证监会及中国期货业协会官方公布的数据显示，截至2025年6月，中国共有150家

期货公司。而按照业内研究院平均人员配置，期货分析师的人数大约在3000-4000人左右。如果算上其他机构及间接的分析师，人数也仅在8000人左右。如此少的基数再加上期货分析师只会流出到其他机构而不会流入，使得大部分期货公司的新手分析师的占比居高不下。

而占比最高的新手分析师，只具备初级的收集和整理基础信息的能力，一般缺乏研究逻辑框架而容易直接从基础数据跳到结论。

初级分析师因从业年龄和成长环境的不同，导致数据处理和加工能力差异大。在如今这样信息爆炸又传播极快的时代，公开数据已经不能提供额外信息优势，而加工过的数据优势更明显。但是缺乏IT技术的分析师可能因为编程能力不足导致数据加工度不够深入。

中级分析师大多数的研报和路演PPT里，图表只用来展示基础信息或相关性，而不是因果关系。大多数分析师没有因果关系库，采用拿来主义，套用热门的因果关系再结合基础信息就开始推导结论，显得逻辑不连贯，说服力不强。

占比最少的高级分析师对品种的前因后果已经有了理解，也累积了足够的因果关系库，但是无法准确抓住当前的主要矛盾，大部分时候随机选其中的一些因果关系，因此其预测准确度也会比较随机。典型的标签就是逻辑连贯，思维有序，具有较强的说服力，但是纵观职业生涯的行情判断胜率，并没有达到超过平均的水平。而能在大多数时间抓住主要矛盾并准确判断的高级分析师则是凤毛麟角的存在。

而这样初中级分析师占比大的研究部门，加上没有统一的投研标准会产生诸多问题，影响研究质量、决策效率和公司整体竞争力。研

究质量不稳定会导致市场分析可信度下降。初中级分析师经验和能力有限，容易依赖历史数据或简单逻辑，难以挖掘深层次的市场驱动因素，甚至不同分析师对同一品种的观点可能都会有矛盾。那么为什么优秀的分析师占比会如此之低呢？

## （二）期货分析师的成长困境

期货研究行业正面临严峻的人才困境和发展瓶颈。全行业仅有3000-4000名专职分析师分散在150家期货公司，平均每位分析师需要覆盖3-5个品种，这种人力配置难以支撑高质量的研究需求。更棘手的是，高校培养体系与实务需求严重脱节，新人分析师往往需要从头学起，而3-5年的成长周期与超过20%的年流失率形成恶性循环。在资深分析师都已疲于应付多品种研究的情况下，新人除非进入头部机构的热门品种组，否则只能从零开始摸索，依靠前任留下的零散资料艰难学习，这种低效的传承方式严重制约了行业整体研究水平的提升。

研究体系标准化缺失是另一个关键痛点。宏观、产业和量化三大研究流派各自为政，即便在同一家公司内部也难以形成统一框架。超过60%的机构缺乏系统的研究知识库建设，核心研究方法完全依赖分析师的个人经验积累。更令人担忧的是，据调研显示离职分析师会带走85%的隐性知识，这种知识资产的流失对研究连续性造成致命打击。缺乏标准化不仅导致研究质量参差不齐，更使得新人培养和团队协作变得异常困难，每位分析师都不得不在前人基础上重复造轮子。

激励机制扭曲进一步加剧了这些问题。当前考核体系过度强调报告数量而忽视研究深

度，80%的机构将报告产出作为主要KPI，导致分析师疲于应付短期任务。价值分配也存在严重失衡，热门与冷门品种分析师的待遇差异显著。更关键的是，知识共享缺乏实质性奖励，分析师既无动力分享核心成果，也缺乏帮助培养新人的积极性。这种激励机制直接导致了“敝帚自珍”现象普遍存在。

分析师自我保护心态的形成有其深层原因。在高度竞争的行业环境中，研究方法和策略逻辑是分析师的核心竞争力，公开分享可能削弱个人价值。现行考核制度只关注个体输出，不考量团队贡献和知识沉淀，进一步强化了这种封闭倾向。此外，缺乏有效的知识产权保护机制，使分析师担心成果被误用或剽窃，同时也将核心方法视为重要的“跳槽资本”。这些因素共同造就了当前知识流动阻滞的困境。

这种状况造成了严重的组织能力退化。研究传承仅靠工作文件交接的原始方式，新人接手时往往要耗费数月时间梳理凌乱的资料。当核心分析师离职时，相关品种的研究可能陷入长达半年的真空期。更深远的影响是，长期无法形成可迭代的研究体系，使机构在日益激烈的市场竞争中逐渐丧失优势。要打破这一困局，需要从人才培养、知识管理和激励机制等多个维度进行系统性改革，构建更开放、更可持续的研究生态。

而NHTA就构建了这样一个以智力资产为核心的研力沉淀传承平台。数据层面，所有公司采集的数据库资源所需要的研究数据，统一落地到NHTA投研平台。分析师无需花费大量时间采集处理数据，平台会统一定时采集更新。其规模效应不但降低采购成本，还提高了

数据使用率和生产效率，赋予分析师更多的时间去做深入研究。KPI和平台研究共享挂钩并附带奖励机制，提高分析师的分享动力。部门领导也可以通过此平台去评价一位分析师的研究能力，智力资产贡献度以及资产的优秀程度，比如预测模型的评估指标好坏。

### 三、NHTA投研平台沉淀传承研究资产的实践路径

#### （一）基础信息提炼和归类

期货研究的信息来源可分为基础信息层和深度加工层，其价值密度随加工深度递增。分析师需建立系统化的信息管理框架，从原始数据中提炼有效认知差，以支撑决策优势。基础信息越来越公开透明，已经基本不具备指导性，不能给未来走势的判断做出贡献。机构内部应该至上而下养成数据深加工的习惯，业内经验表明，精细加工的数据提供更多市场认知差，进而取得超额收益。

一个完善的投研综合平台，庞大、且系统归类的底层基础数据库底座必不可少。目前NHTA接入了包括但不限于彭博、路透、万得、同花顺、钢联、卓创红桃三、SMM、Kpler、克拉克森、百川盈孚、涌益、CCF等资讯终端，并按照要求的目录标准化管理（供需、库存、进出口、估值基差等）完成了NHTA投研平台的底层数据架构搭建。平台中系统设计了宏观和各产业链数据底座目录。宏观指标目录涵盖了国内外经济增长、通胀、就业、财政货币政策等维度数据，以及外汇方面的系统数据。

另外，在大宗商品产业链层面，按照“能化”“有色新能源”“黑色”“农产品”“航运”五大板块分类，对其中涉及到的每个具体

品种，大致按照“价格利润”“供应”“需求”“进出口”“库存”五个维度做指标分类。每个指标目录可以进一步做细项分类，比如在价格利润的目录下，设置了以下几个二级子目录：国内现货、期货盘面、上游成本、加工利润、进口成本利润、市场情绪。每位分析师根据自己负责品种的特性，同时也为了方便后期检索，合理地设置二级目录。

以上庞大的基础研究数据底座，是研究工作的必要条件。基于此，分析师制作混合图表、制作研究图表、进行研报撰写以及PPT制作，并推送给公司客户。

#### （二）数据加工和逻辑关系提炼

##### 1.对原始数据加工处理计算衍生指标

一个完整的综合投研平台，必须具备完善的数据加工处理能力。要照顾到分析师在研究过程中数据处理的方方面面，细节问题还有很多，比如不同周期、不同日期的数据对齐，取交集并集、空缺值处理等基础数据处理功能；又比如计算相关性、方差、残差、百分位等略微复杂的数据处理功能。

##### 2.预测指标的加工处理

NHTA平台提供了预测指标的生成机制，这在宏观需求端预测、或者一些季节性比较强的供需领域的应用场景比较广泛。

在NHTA投研平台的预测指标模块中，设置了多种预测规则，基本涵盖了分析师日常研究中预测方法论的各个场景，以下是一些预测指标的定义：

（1）固定值：预测值默认全部等于常数值；

（2）同比：去年同期值乘同比增速得到预测值；

- (3) 同差：去年同期值加同比增加得到预测值；
- (4) 环比：上期值乘环比增速得到预测值；
- (5) 环差：上期值加同比增加得到预测值；
- (6) N期移动均值：过去 N 期值的平均值；
- (7) N 期段线性外推值：过去 N 期值生成线性回归方程 ( $Y=aX+b$ , 将未来的期数代入得到预测值)；
- (8) 动态环差：选择几个指标进行指标计算得到结果值作为动态环差值，预测值=上期值+动态环差值（可以用来推演库存变化）；
- (9) 给定终值后插值：计算最新数据和预测终值的期数差T和数据差S，环差值=S/T，预测数值=前一期数值+环差值（本期-上期）；
- (10) 季节性：计算过去N年同期的环差（本期-上期）均值，预测值=上期值+环差均值；
- (11) 移动平均同比：计算过去N期平均值的同比值（本期/上期），选择同比年份，预测值=同比年份同期值\*同比值；
- (12) 同比增速插值：计算最新数据的同比增速=（本期数值-去年同期数值）/去年同期数值\*100%，输入同比增速终值，通过期数差值，计算每一期同比增速，预测值=去年同期值\*（1+该期同比增速）；
- (13) 年度值倒推：设定年度值，余额=年度值-年初至今累计值（算法参考累计值），进行余额分配，均值法分配时保证每期数值相

等（日度/周度：剩余期数=剩余自然日历天数/今年指标最新日期自然日历天数\*今年至今指标数据期数；旬度/月度/季度/半年度：剩余期数=全年期数（36/12/4/2）-今年至今自然日历期数），同比法保证每期同比相等（同比增速=余额/同比年份相应日期的余额，预测值等于同比年份同期值\*同比增速）。

### 3.因子间的相关性和预测模型挖掘（识别市场“公开的隐藏信息”）

在开始说因果关系之前，有些情况需要先做下说明。金融建模里，有很多相关性的模型和分析方法，但是相关性不等于因果关系。我们的出发点必须是两个变量之间首先有逻辑关系，然后再去设法找到描绘两个变量之间关系的正确模型。而NHTA平台里有着诸多模型可以使用，比如拟合残差、因果强度检验、线性回归和动态相关性。虽然它们与数据间关系的分析密切相关，但它们各自关注的维度和应用场景存在显著差异。

(1) 线性回归是分析变量间线性相关性的基础工具，通过建立因变量与自变量的线性关系模型 ( $Y=\beta X+\varepsilon$ )，量化解释变量的边际效应。其核心输出回归系数 $\beta$ 直接衡量自变量对因变量的平均影响程度，属于静态相关性分析范畴。但需注意，回归分析仅能揭示统计关联，不能直接证明因果关系。

(2) 拟合残差反映的是模型未能解释的数据变异部分。通过分析残差序列的自相关性（如Ljung-Box检验）或异方差性（如White检验），可以评估模型是否充分捕捉了变量间的动态关系。显著的残差模式往往暗示存在未被建模的非线性效应或时变相关性，这为改进模型提供了方向。

(3) 动态相关性（如滚动窗口相关系数、DCC-GARCH模型）专门刻画变量间关系的时变特征。例如，在研究大宗商品与股票市场的联动效应时，2008年危机期间相关系数可能显著上升，揭示风险传染机制。这类方法能有效捕捉市场结构突变，比静态相关系数更具现实意义。

(4) 因果强度检验（如Granger因果、传递熵）试图突破相关性局限，通过统计方法推断变量间的引导关系。例如，用Granger检验分析美联储政策声明（X）是否显著改善对国债收益率（Y）的预测精度。但需注意，这些方法仍依赖统计准则，真正的因果识别需要结合理论框架和自然实验设计。

关键区别在于：线性回归和动态相关性侧重量化关联强度，而因果检验尝试确定作用方向（A影响了B还是B影响了A），拟合残差则用于诊断模型缺陷。实践中，这些方法常需结合使用——先用回归建立基准关系，通过残差

分析发现模型不足，再引入动态相关性捕捉时变特征，最终用因果检验验证逻辑合理性。例如在研究原油价格与通胀关系时，可能先建立线性模型，发现残差存在聚类波动后改用时变参数模型，最后通过因果检验确认价格传导机制。

还有一种非常规的模型，STL趋势/季节性分解，主要是用来分解时间序列的长期趋势成分，季节性成分和残差成分。常用于一些销售类的预测，可分别对长期趋势和季节性变量建模来提高综合的预测精度。或单独对其中的一个成分进行相关性揭示亦可。相比于其他的分解方法，比如HP滤波或美国官方的X11/X13标准，STL的优势在于可以灵活地处理季节性和非线性趋势，这点在期货行业尤其重要。分析师在选择模型的时候，需要重复考虑到模型的特性和适用的场景。

下面通过NHTA里的示例说明一下，研究员如何快速地通过平台的功能模块去做相关性分析：

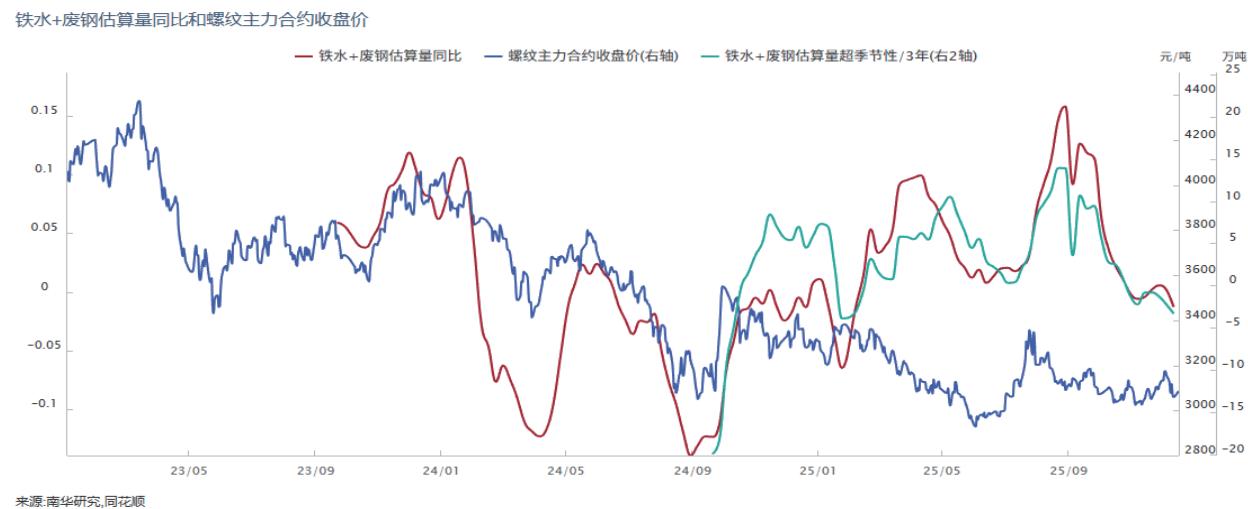


图 1：铁水 + 废钢估算量同比和螺纹钢主力合约收盘价相关性分析图

通过对图1的解读，分析师发现若钢厂复产幅度同比或超季节性水平，且遵循以销定产模式，则表明钢材需求较为旺盛。在此情况下，钢厂复产将推高原料成本，进而支撑成材价格上行。

热卷库存超季节性去化，通常表明其基本面表现强劲，从而对热卷价格形成上行支撑。

(见图2)

从图3中可以看出，螺纹钢需求具有较强季节性特征，因此用过去两年表观需求的季节性来推演今年螺纹钢需求的表现，具有一定代表意义。

可以看出，在NHTA平台上，分析师可以方便地调用自建指标去发现因子间的相关性并生成相应的图表。

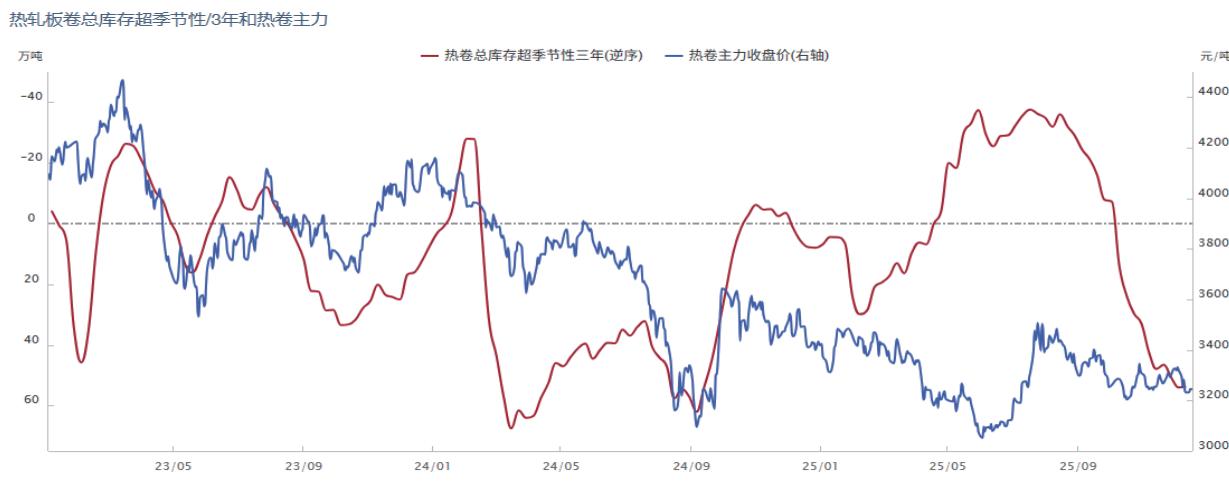


图 2: 热卷总库存超季节性三年和热卷主力价格相关性图

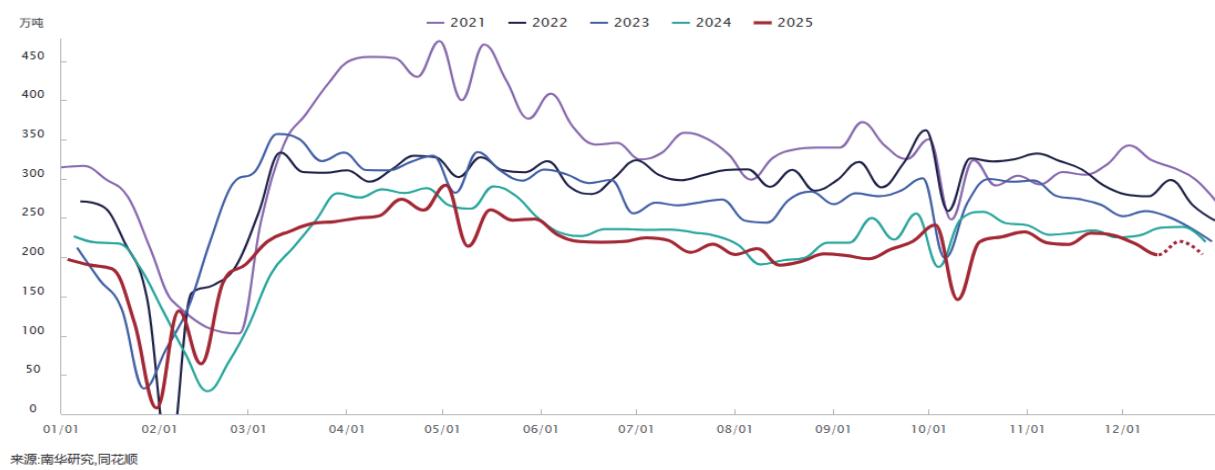


图 3: 螺纹钢表观需求预测图

## 四、进一步推动研究体系转型升级的思路和规划

### （一）因果关系库的深化沉淀和探索

期货研究的核心在于系统性地发现、理解和积累市场运行中的因果关系，而非简单地堆砌数据。通过前一章节例举的相关性模型再结合分析师对于底层逻辑的把握，分析出市场行为的本质是各种因素相互作用形成的动态因果网络，而构建因果关系库（CRD：Causal Relation Database）正是对这一认知体系的系统性沉淀。这个知识库包含了经过历史验证的市场驱动逻辑、事件传导路径和产业运行机制，为研究人员提供了坚实的分析基础。从初级分析师到高级分析师，再到整个研究团队，都需要参与因果关系库的建设与应用。初级人员通过学习现成的因果规律快速入门，中级人员负责发现和验证新的因果关系，高级人员则运用这些积累进行深度市场推演，整个团队通过共享这一知识体系来提升协作效率和研究传承质量，研究框架的清晰梳理，叠加数据完整，从中长期而言，有利于投研团队经验的无缝衔接，确保服务质量的一致性。

因果关系库由多个关键模块组成，共同构成完整的分析框架。逻辑时序图揭示多变量间的动态关联，帮助理解市场驱动因素的演变轨迹；动态平衡表不仅评估未来供需，其调整过程本身也隐含着市场逻辑的变化；情景推演则基于历史规律模拟未来可能的发展路径。研究人员还通过逻辑强弱分析监测市场主导逻辑的切换，运用事件冲击响应模型量化政策变化的影响，并借助数据逻辑图谱结构化地呈现行业因果关系。这些模块相互支撑，形成可迭代更新的研究体系。

建立完善的因果关系库能为研究工作带来多重优势。首先，它大幅提升了研究效率，新人可以快速掌握核心逻辑，资深分析师也能避免重复劳动。其次，基于历史规律进行推演显著增强了预判能力，使分析从主观猜测转向客观验证。在团队协作方面，不同领域的分析师可以在因果关系层面进行深度交流，而非停留在结论层面的简单讨论。最重要的是，这种系统化的知识管理有效降低了决策风险，避免因选择性使用因果关系导致的误判，同时为风险评估提供了科学依据。

作为期货研究的“底层操作系统”，因果关系库实现了从碎片化信息处理到系统化逻辑推演的质的飞跃。研究人员的工作重点不应局限于数据收集，而应着眼于发现、验证和优化各类因果关系，将其转化为可复用的知识资产。只有建立这样动态更新的知识体系，期货研究才能真正超越对市场的简单解释，实现对市场行为的准确预判，为投资决策提供可靠支撑。这种研究范式的转变，将推动整个行业向更专业、更高效的方向发展。

期货研究的终极目标是准确识别和把握市场的核心矛盾，这一过程常被视为高级分析师的个人能力甚至天赋，实则应当是一个严谨的系统化工作流程。市场运行具有典型的反身性特征，那些被广泛认知的观点往往已经充分反映在价格中，真正驱动市场的关键矛盾通常隐藏在日常认知范围之外。研究人员依靠个人经验构建的因果关系库往往存在局限性和主观偏好，容易导致核心矛盾的筛选过程过于随意，这正是多数判断失误的根本原因。

要系统性地解决这一问题，需要建立科学的分析框架。首先必须将历史积累的因果关系

库（包含宏观、行业和价格驱动逻辑）与当前市场信息集（构成选择约束条件）进行交叉验证。通过逻辑图谱沉淀、高阶数据分析和未来情景推演等系统方法，才能突破个人经验的局限，客观识别真正的主要矛盾。这一过程强调严谨的交叉验证，而非依赖直觉或经验判断，从而将看似“艺术化”的决策过程转化为可验证、可复用的科学方法。

## （二）价格预测模型的动态框架探索

因果库是尝试找到对于某个变量（比如价格）可能存在影响的因子有哪些，单个影响的程度又有多少。而本部分则是对价格变化构建一个建模流程框架，即如何组织这些因果关系，套用什么样的模型去做一个价格预测模型。而现在业内普遍采用的是机器学习算法去得到一个最终的结果。

在开始搭建预测模型之前，有些问题可能需要解决一下。比如宏观基本面的数据一般来说延迟性较大，所以需要采用一些领先指标（可用前面的相关性模型检测出）来提高时效性，同样的情况可能也需要应用到预测的目标上。而有些时候价格不可用的时候还需要用一批同步指标（按优先级排序）来替代。

为了构建有效的价格预测模型，需要建立系统化的动态框架，这一过程始于对价格驱动逻辑的深入理解。研究人员首先需要识别影响价格的核心因素，包括宏观经济指标、市场情绪、供需关系和技术指标等，并将这些因子分为单因子和因子组两个层级。单因子指能独立影响价格的变量，如成交量或波动率；而因子组则是多个相关因子的组合，如技术指标组或基本面因子组。基于这种分类，模型设计采用分层架构：底层是评估单个因子预测能力的单

因子模型，上层则是整合多因子的综合预测模型，可运用线性回归、XGBoost或神经网络等不同算法。这种分层设计既保证了分析的细致性，又确保了模型的综合性。

在模型评估阶段，需要采用多维度的检验方法。对于单因子，通过统计检验（如IC、RankIC、t检验）衡量其与价格的相关性，并借助滚动回测观察因子在不同市场环境下的稳定性。对于因子集模型，则需对比不同算法的预测能力，运用SHAP值或特征重要性分析识别关键驱动因子，并通过组合优化剔除冗余因子以提高泛化能力。这一阶段的重点在于建立量化指标体系，包括预测准确性（MSE、MAE、R<sup>2</sup>）、稳定性（收益波动、最大回撤）和市场适应性（不同行情下的表现），从而全面评估模型的有效性。

模型的动态维护是确保长期有效的关键环节。研究人员需要持续监测预测偏差，进行归因分析：检测因子失效（如IC值衰减）、分析市场结构变化（如政策影响）、检查模型过拟合（样本外测试）。基于这些分析，模型优化分为增量调整和根本性重构两个层面。增量调整包括增加新因子（如另类数据）、调整因子权重或参数调优；而当核心驱动逻辑失效或市场范式转变时，则需要进行算法升级或模型结构重构，如从线性模型转向深度学习。

实现模型的自动化运作是提升效率的重要手段。建立实时数据流和自动回测框架可以实现定期评估，设置预警机制（如预测误差阈值）能够及时触发模型更新。同时，要注重模型的可解释性，运用LIME或SHAP分析避免黑箱风险，并整合风险控制模块实现动态仓位调整。值得注意的是，虽然自动化程度越高越

好，但仍需保留人工干预机制，允许分析师对模型预测进行必要修正，或采用集成学习方法混合多个子模型以提高健壮度或鲁棒性（Robustness）。

这一动态框架的核心价值在于其适应性和进化能力。通过持续监测、评估和迭代，价格预测模型能够跟随市场变化而不断优化，既保持了科学的研究的严谨性，又具备了实战应用的

灵活性。从因子筛选到模型构建，从绩效评估到迭代优化，每个环节都形成闭环，确保预测能力随时间推移而不断增强。这种系统化的方法不仅提高了预测准确性，更重要的是建立了一套可持续完善的研究体系，为投资决策提供了更加可靠的预测模型支持。

下面举个简单的原油价格预测模型的例子：

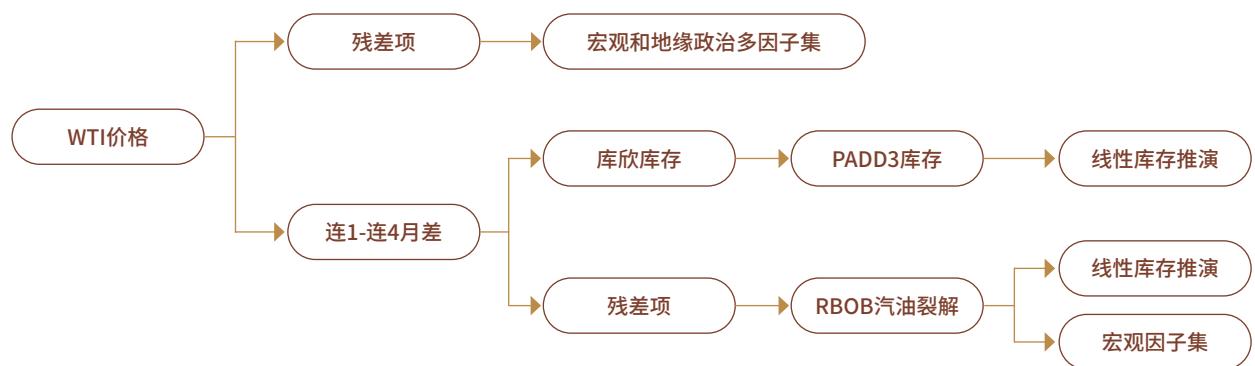


图 4：WTI 原油预测模型

从分析师的逻辑框架出发，先将WTI原油价格用线性回归做一下拆分，连1-连4的月差和残差项（未解释变量）。将提取出来的残差项进行单独建模，使用宏观和地缘政治多因子集。月差则先用库欣库存来预测，分解出来的残差项则用RBOB汽油裂解去解释。汽油裂解的解释变量则从成品油库存，汽油库存和汽油平衡产品等变量推演而出（一般使用线性外推）的因子再加上市场/宏观因子建模得出。而之前的库欣库存也是用的PADD3库存线性外推的因子来进行解释预测。简而言之，分析师先从主观逻辑角度将问题分解成更容易解释的子模型，然后选择最合适的模型对分解出来的变量进行解释验证，最后再还原成一个整体的预测模型。

预测模型不仅是高级分析师的深入研究产出，更是其投研框架的最终产物，毕竟研究的最后一步就是对标的进行预测并投资。其不仅对自己的研报和路演等输出起到极大的帮助，更是落地的智力资产。

而为了让研究团队能够专注于建模工作，算法工程师和IT开发人员也需要介入平台及模型模块化的开发工作，使得研究团队能够零代码实现各种模型的验证评估工作。分工上，研究员负责主观研究和预测模型的印证、解释模型的现实含义、预测偏差分析及决定迭代或重构模型。算法工程师负责细分模型的选取和算法实现及参数优化，建立模型运行的量化跟踪以及参与实现模型的迭代。IT开发人员则负责上线建模，模型跟踪的模块化，以及收藏历史

模型库，形成智力资产库。

### （三）量化研究过程要素，打造高质量分析师和研究团队

期货研究部门的转型升级，核心在于通过数字化投研平台（如南华期货的NHTA平台）将传统依赖个人经验的“人力资产”转化为可沉淀、可迭代的“智力资产”，从而系统性提升研究质量和团队效能。以下是关键实施策略：

#### 1. 构建标准化研究框架，降低新人培养门槛

知识沉淀：投研平台整合历史研究报告、因果关系库（CRD: Causal Relation Database）、逻辑图谱等核心资产，将资深分析师的隐性经验显性化。新人可通过平台快速掌握品种研究框架，避免从零摸索。

流程规范化：平台内置数据清洗、指标计算、模型构建等标准化工具（如动态相关性分析、STL分解），减少初级分析师在基础工作上的时间消耗，聚焦逻辑提炼。

案例学习：通过平台积累的典型市场案例（如政策冲击、周期切换），帮助新人理解因果关系的动态演变，缩短成长周期。

#### 2. 推动团队协作与知识共享

破除信息孤岛：平台强制要求分析师添加指标，深度加工数据、创建逻辑图表和预测模型，形成共享资源池。例如，NHTA平台通过KPI与知识贡献挂钩的机制，激励分析师开放私有研究成果。

跨品种协同：基于平台的“数据逻辑图谱”，不同领域分析师可在因果关系底层展开深度讨论（如黑色系与宏观研究的联动），而非仅交换结论。

传承保障：当分析师离职时，其研究逻辑和模型可通过平台完整保留，接任者能快速继承而非重建研究体系。

#### 3. 强化研究深度与预测能力

动态模型迭代：平台支持价格预测模型的自动化回测与优化（如因子失效预警、参数动态调整），推动研究从“描述性分析”向“预测性决策”升级。

因果关系验证：整合Granger因果检验、残差分析等工具，帮助分析师区分相关性与因果性，减少逻辑误判。例如，通过平台回溯历史事件冲击（如关税调整），验证当前逻辑的合理性。

核心矛盾识别：高级分析师利用平台的“情景推演”和“逻辑强弱分析”功能，系统筛选市场主要矛盾，避免主观随意性。

#### 4. 优化考核与激励机制

多维评价体系：平台记录分析师的模型准确性（如MSE）、知识贡献度（如上传的CRD数量）、团队协作（如被引用的研究报告）等数据，替代单一的“报告数量”考核。

差异化激励：对冷门品种研究（如农产品）和深度加工数据（如另类情绪指标）给予额外奖励，平衡资源分配。

职业发展通道：通过平台积累的智力资产（如独家预测模型），可作为分析师晋升的重要依据，强化长期价值导向。

#### 5. 实现研究-交易闭环

策略落地支持：平台对接模拟交易系统，将研究模型转化为可执行的量化策略（如套利信号生成），验证研究价值。

实时反馈优化：交易结果反哺研究端，动态修正因果关系库。例如，当预测模型连续失

效时，触发平台预警并启动归因分析。

## 五、总结

NHTA投研平台通过系统化的数字化转型，成功解决了传统期货研究体系中的过度依赖分析师和研究传承困难的核心痛点。

首先，平台将分散的“人力资产”转化为可沉淀、可复用的“智力资产”，包括数据资产（产业链指标）、逻辑资产（因果关系库CRD）和模型资产（动态价格预测框架）。这种结构化积累不仅降低了人才流动的负面影响，还显著提升了研究效率，使分析师从基础数据处理中解放出来，专注于高价值分析。

其次，NHTA平台通过标准化工具（如动态相关性分析、STL分解）和自动化流程（如模型回测与预警），实现了研究过程的科学化与规范化。初级分析师可快速掌握研究框架，资深分析师则能通过逻辑图谱和情景推演精准识别市场核心矛盾。

平台的初期智力资产累积需要研究员的倾力投入，因此期货公司需要通过KPI与知识贡

献挂钩的机制，对齐研究员与期货公司的长期共同利益，进而打破研究员“敝帚自珍”的壁垒，促进了团队协作与知识共享。

最后，南华期货的实践表明，投研平台的终极价值在于构建“研究操作系统”，形成从“经验驱动”到“数据与逻辑驱动”的范式变革。未来，随着AI技术的深度融合（如智能因子挖掘、自动化研报生成），这一平台将进一步推动研究体系向智能化跃迁，为期货公司打造可持续的核心竞争力。经实践得到启示：数字化转型不仅是技术升级，更是研究生态的重构，唯有系统性沉淀知识资产，才能实现研究能力的传承与提升。

（责任编辑：王建锋）

### 作者简介：

顾双飞，任职于南华期货股份有限公司，现任研究院副院长。

胡乐克，任职于南华期货股份有限公司，现任研究院策略分析师。