

文章标题：基于随机森林方法的甲醇期货价格预测与交易策略研究

作者 1 姓名：季晨洋（1994-），单位：同济大学经济与管理学院（200092），电话号码：19121806127，电子邮件：13151087817@163.com；（性别：男；籍贯：江苏淮安人，硕士研究生，研究方向：数据挖掘）

作者 2 姓名：林杰（1967-），单位：同济大学经济与管理学院（200092），电话号码：13524636367，电子邮件：linjie@tongji.edu.cn，通讯作者，通讯地址为：上海市四平路1239号同济大学经济与管理学院；（性别：男；籍贯：四川渠县人；职称：教授；博士生导师；主要研究方向：管理信息系统、数据挖掘）

基于随机森林方法的甲醇期货价格预测与交易策略研究

季晨洋, 林杰

(经济与管理学院, 同济大学, 上海 200092)

摘要: 甲醇是重要的新型清洁能源, 其期货产品上市已近 10 年, 本文采用机器学习技术研究甲醇期货价格的预测模型和交易规则。论文采用随机森林算法构造甲醇期货价格预测模型, 使用甲醇产业链上下游产品的基本面特征作为输入变量, 并对输入变量进行归一化处理构造对照模型, 结合使用基于 Aberration 策略思想构建交易策略集合, 使用夏普比率对交易策略进行筛选, 构造有效的甲醇期货量化交易模型。研究结果显示, 本文模型在保证交易策略良好泛化能力的情况下可以实现高出同期 10 年期国债收益率 1.2 倍以上的年化收益率。也表明利用甲醇产业链上下游产品的基本面特征能够很好地解释甲醇期货价格, 并结合使用夏普比率筛选的, 基于 Aberration 交易策略思想的交易规则能够构建有效的甲醇期货量化投资模型。

关键词: 甲醇期货; 价格预测; 随机森林; 交易策略泛化

JEL 分类号: G11、G13、G19

一、引言

期货交易是市场经济发展到一定阶段的产物, 成熟的期货交易市场在一定程度上可以认为是完全竞争市场, 是经济学中最理想的市场形式。我国第一个商品期货交易所的出现要追溯到上世纪 90 年代, 而世界上最早的期货交易市场出现在 19 世纪 50 年代, 相比于这些成熟国外期货市场, 我国的期货交易市场还不成熟, 期货理论对现实期货交易的支持还不够。国内甲醇期货产品于 2011 年 10 月 28 日在郑州商品交易所上市, 到今年已近十年之久, 由于我国能源结构的逐步改善, 甲醇作为新型清洁能源而被广泛使用, 相关企业和投资者对甲醇期货的关注度也在不断上升, 分析甲醇期货价格对国内甲醇期现货市场的健康发展以及相关企业和投资者进行甲醇期现货交易具有重要意义。

在现有期货价格研究中, 研究方法主要可分为计量经济学方法和机器学习方法。其中计量模型包括自回归模型、差分自回归移动平均模型、广义自回归条件异常模型等。机器学习方法主要集中在决策树、支持向量机、随机森林、人工神经网络等(陈标金等人, 2019)。随着计算机技术的飞速发展, 机器学习也逐渐被应用到金融领域的研究中, 并表现出良好的研究性能。Einav 和 Levin (2014) 认为在经济学研究领域, 经济模型关注的重点主要放在能够代表工序变化的因素上, 研究人员需要对数据的异质性进行校正, 并尝试使估计参数的异方差与标准误差和协方差一致, 此外还需要通过不同模型来评估模型的鲁棒性。经典统计模型更强调模型的结构和理论性质, 具有一定的可解释性, 单对复杂数据结构的拟合不够(周亚, 2021)。与传统计量模型相比, 机器学习算法对数据的前提假设较少, 是以数据为驱动, 将重点放在模型的输出结果上, 进而揭示出特征对模型的解释能力(Alessandro Vespignani, 2009)。随机森林是众多算法中适应能力较强的算法之一。Breiman (2001) 指出随机森林算法不需要考虑多元回归模型中需要考虑的多重共线性问题。方匡南, 吴见彬, 朱建平和谢邦昌 (2011) 指出随机森林对异常值和噪声有很好的容忍度, 并且预测精度高。Pérez-Rave, Correa-Morales 和 González-Echavarría (2019) 也指出随机森林算法可以有效避免模型的过拟合, 是一个能够做出精确预测的稳健算法。

关于期货价格预测模型的输入特征变量的选择, 已有众多实证研究指出产业链上下游相关产品的基本面特征对预测期货价格具有重要作用。姬培森 (2020) 通过基本面分析对商品期货价格的走势进行充分预测; 李顺平和朱昂明 (2019) 使用螺纹钢产业链上下游相关产品的基本面特征变量很好地解释了螺纹钢期货价格; 金志峰, 崔楠楠和宋歆欣 (2021) 利用

SVAR 回归发现供需会对铁矿石期货价格变化产生显著影响。

鉴于此,本文利用 2016 年 1 月 2 日至 2021 年 5 月 31 日期间甲醇期货主力合约收盘价序列作为被预测变量,基于经济学生产理论以及供需关系准则选取产业链上甲醇上下游产品的基本面特征作为备选预测变量,利用相关性分析筛选甲醇期货价格预测模型的预测变量;分别以选取的特征变量、以及去除量纲后的特征变量作为预测变量,利用随机森林算法构造甲醇期货价格预测模型对照组,选取预测能力强的模型作为甲醇期货价格预测模型;使用夏普比率进行基于 Aberration 交易策略思想的回测交易规则筛选及该交易策略的泛化能力检验,进而间接检验甲醇期货价格预测模型的预测能力。

二、预测变量筛选方法

期货价格预测效果取决于预测变量的选择和预测模型的构建。在预测变量的选择上,众多实证研究表明产业链上下游产品的基本面数据对金融资产价格预测具有重要意义(马郑玮等, 2021)。

(一) 甲醇产业链结构分析

甲醇是合成气生产的重要化学品之一,是一种化工原料和燃料,目前市场上交易的甲醇以工业甲醇为主,其生产原材料主要为煤、焦油、天然气等,其生产过程不会出现任何副产品,消耗的过程中也没有污染物的排放。甲醇自工业化生产以来,生产的方式也发生了很大的改变,由原先的以煤、焦油作为主要生产原料转变为了以天然气为生产原料。由于我国独特的能源结构分布,我国的甲醇生产依旧是以煤作为主要原料,目前我国仍以煤作为原料的甲醇生产产能约占总产能的 65%。

甲醇的产业链下游产品也很多,包括醋酸、甲醛、二甲醚等,还可用作动力燃料,涉及行业诸多,包含化工、建材、能源等。

(二) 备选预测变量选取

根据上述上下游产业链结构分析可知,甲醇产业链的上下游产品主要有煤、天然气、焦炉气,下游产品主要有甲醛、二甲醚、醋酸等。根据生产理论,上游产品作为下游产品的生产原料,其销售价格将作为下游产品生产的直接费用,并直接反应到下游产品的销售价格当中;从供需关系角度出发,当产品供给充足时,价格会出现下降趋势,反之则会出现上升的趋势,呈现出负相关关系,而产品的消费量、出口量和产能、进口量将直接影响产品的库存量,并分别表现出正、负相关性;从商品期货市场交易过程来看,期货交易最终将以现货的方式进行交割,张宗成,王骏(2005)通过对上海期货交易的实证分析中已验证我国期货与现货价格之间存在着双向因果关系。

根据上述分析,并结合我国甲醇产业链的实际情况以及数据的可获性,本文选择的上下游产品分别为煤和醋酸,备选预测变量为:煤现货价格、产量、进口量、库存量;甲醇现货价格、产量、进口量、出口量、库存量、毛利;醋酸现货价格、进口量、出口量、库存量、消费量、毛利。

(三) 备选预测变量预处理及预测变量筛选方法

在上述选取的备选预测变量中,部分特征变量是日序列数据,部分是周、月序列数据,需要对数据进行预处理。本文对周以及月序列数据的处理方式是当前样本观察日期与上一样本观察日期区间内的各观察样本值均设定为上一样本观察值;将备选预测变量样本数据滞后一个交易日后与被预测变量进行配对形成样本向量。

将上述备选预测变量与甲醇期货价格进行相关性检验,选取与甲醇期货价格显著相关的备选预测变量作为模型的预测变量。

三、随机森林预测模型构建及回测检验规则设计

（一）随机森林预测模型构建

随机森林是决策树模型的集成算法，它的实用性以及高预测精度已在大量实证分析中得到论证。它是将 n 棵决策树的预测结果结合起来以做出比单棵决策树更为精确的预测，随机森林在构造单棵决策树时除了在选择样本是使用随机抽样，在构建单棵决策树的输入特征时也采用随机抽样的方式从原有特征空间中选子特征空间，此做法有效提高了模型的稳定性。随机森林回归模型的预测结果是由众多弱评估器的预测结果进行平均而得，即众多决策树预测结果的平均值作为整体模型的回归值。本文选用拟合优度 (R^2) 和平均绝对百分比误差 (MAPE) 作为选取预测模型的标准，并设计交易规则进行回测交易以检验选定的预测模型的跟踪交易效果。

1. 随机森林预测模型构建方法

随机森林回归模型的构建方法是：首先，从 N 个交易日时间序列样本集中，按 7:3 比例有序导出 70% 的样本量作为模型的训练集 D ，30% 样本量作为模型的测试集；其次，使用 bootstrap 重抽样方法从训练集 D 中抽取同样数量的子样本集，重复此步骤 n 次生成子样本集 $D_i, i = 1, 2, 3, \dots, n$ ；接着，以随机抽样的方式对上述生成的各个子样本集进行特征抽样，抽样的特征数量保持在总特征数量的 70% 以上，各个样本集根据特征抽样的结果进行训练子集数据的重新生成；再次，根据抽样的各子样本集构建决策树，并将 n 棵决策树的算术平均值作为模型的预测结果；最后，对上述模型参数进行调整，选择预测效果较优的随机森林回归模型。

2. 对照模型构建

考虑到本文所选取的预测变量的量纲不同，可能会对模型的预测结果产生影响。为了避免预测变量的不同量纲对预测模型造成负面影响，本文对被预测变量和筛选出的预测变量进行归一化处理，使用上述相同的方式构建随机森林回归模型，形成对照模型，对比模型回归结果选取甲醇期货价格的最终预测模型。归一化处理的方法为：

$$X'_{ij} = \frac{X_{ij} - \text{Min}_i}{\text{Max}_i - \text{Min}_i}, i = 1, 2, 3, \dots, 14, j = 1, 2, 3, \dots$$
$$Y'_j = \frac{Y_j - \text{Min}}{\text{Max} - \text{Min}}, j = 1, 2, 3, \dots$$

其中， X_{ij} 为第 i 个特征变量的第 j 个样本值， Min_i 为第 i 个特征变量的样本最小值， Max_i 为第 i 个特征变量的样本最大值， Max_i 为第 i 个特征变量的第 j 个经过归一化后的样本值； Y_j 为被预测变量的第 j 个样本值， Min 为被预测变量的样本最小值， Max 为被预测变量的样本最大值， Y'_j 为被预测变量的第 j 个经过归一化后的样本值。

3. 模型评估指标

本文采用的模型评估指标为 MAPE 和 R^2 。设 n 为样本数， y_i 和 y'_i 分别表示为被预测变量的样本值和预测值，其计算方式为：

$$\text{MAPE} = \frac{100\%}{n} \sum_i^n \left| \frac{y'_i - y_i}{y_i} \right|$$
$$R^2 = 1 - \frac{\text{SSR}}{\text{SST}}, \text{SSR} = \sum_i^n (y_i - y'_i)^2, \text{SST} = \sum_i^n (y_i - \bar{y})^2$$

其中， \bar{y} 为被预测变量的样本均值。

（二）模型回测检验方法

依据两种预测模型各自评估指标结果，选定最终的甲醇期货价格预测模型。使用选定的

预测模型对样本训练集进行预测得到预测样本集，针对得到的预测样本集构建回测交易规则。本文选用的回测交易策略为基于 Aberration 策略思想所构建的策略集合，并使用夏普比率来确定交易策略的相关参数，进而确定最终的回测交易策略。

Aberration 交易策略分为两个部分，该策略的第一部分是相关参数的计算：

$$UpperBand_t = MA_t + K_1 * \sigma_t, t = 1, 2, 3 \dots$$

$$LowerBand_t = MA_t - K_2 * \sigma_t, t = 1, 2, 3 \dots$$

$$MA_t = \frac{\sum_{i=1}^N Close_{ti}}{N}, t = 1, 2, 3 \dots$$

$$\sigma_t = \sqrt{\frac{\sum_{i=1}^N (Close_{ti} - \overline{Close})^2}{N}}, t = 1, 2, 3, \dots$$

其中 $UpperBand_t, t = 1, 2, 3 \dots$ 为上轨线在 t 时刻的向上突破值， $LowerBand_t, t = 1, 2, 3 \dots$ 为下轨线在 t 时刻的向下突破值， $MA_t, t = 1, 2, 3 \dots$ 为甲醇期货收盘价在 t 时刻的移动时长为 N 的移动平均值， $\sigma_t, t = 1, 2, 3 \dots$ 为甲醇期货收盘价在 t 时刻移动时长为 N 的移动标准差， K_1 和 K_2 为阈值。

第二部分是交易规则，Aberration 交易策略是基于布林格通道 Bollinger Bands 理论的策略，Bollinger Bands 理论是由 John Bollinger 在 20 世纪 60 年代提出，该理论指出股价将主要在由移动平均线以及在移动平均值基础上增加和减少一个单位的移动标准差形成的上轨线和下轨线之间进行波动，当产品价格触及上轨或者下轨线时，价格将有很大可能性重新回到移动平均线附近。Keith Fitschen 根据此理论进行了大量实验后发现布林格通道的上轨和下轨线其实是价格突破信号，即价格靠近上轨或下轨线附近的时候将有很大可能性超过上轨值或者跌破下轨值。本文采用的交易策略为基于 Aberration 交易策略结合甲醇期货价格预测模型进行改进，具体方法为，当使用选定的甲醇期货价格预测模型预测第二交易日价格上涨超过上轨线时进行买多操作，并在价格回落到移动平均线下方时进行平仓操作；当使用选定的甲醇期货价格预测模型预测的第二交易日的预测价格跌破下轨线时进行买空操作，并在价格上升到移动平均线上方时进行平仓操作。

由于夏普比率是能够同时对收益和风险进行考虑，能够很好地评估回测交易效果，所以本文将根据夏普比率来进行交易策略的筛选，其计算方法为：

$$SharpeRatio = \frac{E(R_p) - R_f}{\sigma_p}$$

其中 $E(R_p)$ 是投资组合预期年化报酬率， R_f 为年化无风险利率，本文采用同期 10 年期国债收益率代替， σ_p 是投资组合年化报酬率的标准差。根据回测交易策略的夏普比率确定基于 Aberration 交易策略的策略参数 N 、 K_1 、 K_2 ，再根据确定后的交易策略来评估基于基本面特征的甲醇期货价格预测模型的预测能力。

使用回测交易方法检验甲醇期货价格预测模型有效跟踪交易效果的具体步骤为：首先，根据两组对照模型的表现结果进行预测模型的选择；其次，根据选定的甲醇期货价格预测模型对价格预测模型的训练样本集进行预测得出甲醇期货价格预测序列；再次，根据不同 N 、 K_1 、 K_2 参数进行策略回测交易，本文中 N 的取值范围为 $[20, 60)$ ，步长为 1， K_1 的取值范围为 $[0.6, 2)$ ，步长为 0.1， K_2 的取值范围为 $[0.6, 2)$ ，步长为 0.1，并计算各交易策略的夏普比率；最后根据策略交易回测收益以及夏普比率选定交易策略作为甲醇期货价格预测模型所能带来的收益，进而评估甲醇期货价格预测模型的预测能力。为了提高交易策略的现实可用性，本文将使用甲醇期货价格预测模型预测前一段时间区间内的甲醇期货价格，并使用此时间区间内的样本进行交易策略的筛选，然后使用剩余时间区间的样本进行选定的交易策略的测试，以检验交易策略的泛化能力。

四、甲醇期货价格随机森林预测模型构建与检验

（一）样本数据说明

本文选取 2016 年 1 月 2 日至 2021 年 5 月 30 日时间区间内甲醇期货主力合约收盘价作为被预测变量。备选预测变量为焦煤现货价格、产量、进口量、库存量；甲醇现货价格、产量、进口量、出口量、库存量、毛利；醋酸现货价格、进口量、出口量、库存量、消费量、毛利，共 16 个特征变量。其中现货价格数据来源于生意社网站，期货价格数据来源于大智慧行情系统，其他特征变量来源于钢联客户端。备选预测变量中包含日序列、周序列、月序列数据，本文采取的处理方式是交易日若无前一交易日对应的特征变量样本值则选用已有的离本交易日最近的样本值与之匹配，形成日序列预测向量。使用向后偏移一个交易日的预测向量与交易日被预测变量进行匹配构成一条观测样本。

1. 备选预测变量预处理

在上述选取的备选预测变量中，部分特征变量是日序列数据，部分是周、月序列数据，需要对数据进行预处理。本文对周以及月序列数据的处理方式是将两个相邻样本日期之间所有交易日的的数据均取值为距现在较近样本相对应的取值；将备选预测变量样本数据滞后一个交易日后与被预测变量进行配对形成样本向量。

2. 备选预测变量相关性分析及预测变量筛选

对上述经过预处理后获得样本数据进行相关性检验。本文选用 SPSS 统计软件对样本数据进行相关性检验，检验结果表 1 所示。

表 1 甲醇期货价格与备选预测变量相关性检验

甲醇期 货价格	甲醇产 量	甲醇出 口量	甲醇进 口量	甲醇库 存量	甲醇现 货价格	甲醇毛 利	焦煤产 量	焦煤进 口量	焦煤库 存量	焦煤现 货价格	醋酸出 口量	醋酸进 口量	醋酸库 存量	醋酸消 费量	醋酸现 货价格	醋酸毛 利
Pearson	-.164**	.270**	-.506**	-.457**	.948**	.959**	-.122**	-.091	-.009**	.425**	.344**	-.483**	-.248**	-.052	.582**	.379**

**、在置信度（双测）为 0.01 时，相关性是显著的

本文仅将被预测变量与各备选预测变量之间的 Pearson 相关性检验结果呈现于表 1 上，根据相关性检验结果可以看出，在 0.01 显著水平（双侧检验）下，除了煤进口量、醋酸消费量之外的所有特征变量均与甲醇期货价格显著相关。此外，根据相关性检验结果，各备选预测变量两两之间，共 120 对的相关性检验结果中，有 109 对相关性检验结果是显著的，表明预测变量之间存在多重共线性。此外，将被预测变量与预测变量进行 Spearman 和 Kendall 相关性检验，结果和 Pearson 检验结果一样，本文未将 Spearman 和 Kendall 检验结果呈现出来。根据相关性检验结果，本文将煤进口量和醋酸消费量之外的共 14 个备选特征变量选为预测变量。

（二）价格预测模型实验结果

将上述筛选出的 14 个特征变量和甲醇期货价格分别作为随机森林算法模型的预测变量和被预测变量构造预测模型 1，并对这些变量进行归一化处理构造对照模型 2。根据本文选择的模型评估指标 R^2 和 MAPE 对两个模型的参数进行调整，选取相对最优模型。调整后的两个预测模型的实验结果见表 2：

表 2 随机森林模型评估指标结果

误差	R^2	MAPE
模型 1	0.990	0.012
模型 2	0.991	0.103

比较上述模型的结果可以看出，在经过归一化处理后的样本数据在 MAPE 评估指标上要模型 1 的结果相差 7.5 倍，而在 R^2 这一评估指标上也并未能表现的更好，而仅仅只提高了 0.1%左右，表明对样本数据进行归一化处理无法很好地提高模型的预测精度。因而本

文选择未经归一化处理预测模型 1 作为甲醇期货价格预测模型。

（三）回测检验结果

使用模型 1 对所选样本时间区间的甲醇期货价格进行预测得出甲醇期货价格预测值，在 2016 年 5 月 1 日至 2020 年 5 月 1 日时间区间的甲醇期货价格模型预测值上使用前述的交易策略进行回测交易。由于不同的交易参数组合将产生不同的交易策略，进而可进行交易的次数也随之不同，本文除却那些在 5 年内交易次数小于 5 次以及交易次数大于 25 次并且买多或者买空次数总数低于 2 次的交易策略，以确保一年内可以实现至少一次的交易机会，并且摒弃交易次数过多的交易策略。交易策略的统计结果为：符合要求的交易策略共 274 个，其中能够带来正收益的策略共 139 个；在能够带来正收益的策略中，仅有 4 个策略的 K_2 值为 0.7，其余均为 0.6，并且这 4 个交易策略的收益分别为 455、432、364、298，属于收益较小的策略，并且夏普比率均小于 0.7，因而，将这 4 个策略排除作为最终策略模型的候选项，并将 K_2 的值设定为 0.6；将 K_2 其他取值的策略舍弃后只剩下 135 条策略，在所剩的 135 个策略中， K_1 取值为 0.9 时策略的收益和夏普比率均处于前列，并且两者的最大值均在其中。因而将 K_1 的值设定为 0.9；将 K_1 其他取值的策略舍弃后仅剩 30 条策略；再去除夏普比率小于 2 的策略后只剩下 4 条策略，筛选后的相对最优交易策略回测检验结果见表 3：

表 3 相对最优交易策略回测检验结果

N	K_1	K_2	收益	夏普比率
44	0.9	0.6	792	2.1884
43	0.9	0.6	766	2.0575
47	0.9	0.6	753	2.2465
48	0.9	0.6	731	2.1282

根据上表所示，当 N 的取值为 44 时，策略的夏普比率达到了 2.1884，为所选策略中的第二大取值，但是此时的收益为所有策略中的最大值，因而 N 的取值可设定为 44，即最终的交易策略中参数 N 、 K_1 、 K_2 取值分别为 44、0.9、0.6 时，模型的收益达到最优，策略的累积收益为 792，将此策略的收益作为甲醇期货价格预测模型的收益。以在交易过程中需要的最大本金作为本次交易的本金，为 2306，因而此次长期交易的收益率为 34.35%，平均年收益率为 6.87%，高于同期的 10 年期固定国债的利率 3.1% 一倍多。

为了检验上述回测交易策略的泛化能力，在 2020 年 5 月 1 日至 2021 年 5 月 1 日时间区间上应用该交易策略进行回测交易。回测交易的结果显示，依据此策略进行投资，投资收益可达 414，年化收益率为 21.25%，超过同期 10 年期固定国债利率 5.8 倍多。

可以看出，在结合本文提出的及与之相应的基于 Aberration 策略思想的交易规则后，基于产业链上下游产品基本面特征的随机森林甲醇期货价格预测模型具有很高的实用价值。

五、结论与启示

本文发现，以甲醇产业链上下游相关产品的基本面特征作为模型的预测变量来预测甲醇的期货价格具有很好的预测能力，结合基于 Aberration 交易策略思想构建的交易策略能够实现稳定且较高的交易收益。

首先，本文借助经济学原理中的生产成本理论以及供需关系准则初步选取了模型的备选预测变量，从提高模型实用性角度出发将周序列、月序列数据细化为日序列数据，将被预测变量与向前偏移一个交易日的预测变量数据进行匹配形成样本向量，通过相关性检验后筛选出模型的预测变量。

其次，考虑到变量的不同量纲可能会对模型造成负面影响，本文采用了归一化的方式去除量纲可能带来的影响，构造对照模型。根据两组对照模型的 R^2 和 MAPE 模型评估指标结果选择甲醇期货价格预测模型。结果发现，将变量进行归一化处理并没有提高模型的精确度，

反而大大增大了模型的误差。

最后,结合基于 Aberration 交易策略思想构建交易策略集合,使用 2016 年 5 月 1 日至 2020 年 5 月 1 日时间区间内的甲醇期货价格模型预测的结果进行回测交易,并使用夏普比率对交易策略进行筛选,并根据选定的交易策略在价格预测模型的测试样本集上的表现结果来评估预测模型的实际价值。研究发现,在交易策略中周期取值为 44,上轨线中的系数取为 0.9,下轨线的系数取为 0.6 时,交易策略的累积平均收益可以达到较大值,平均年化收益率可以达到 6.87%,超过同期 10 年期国债利率 1.2 倍多。将此策略应用于 2020 年 5 月 1 日至 2021 年 5 月 1 日时间区间上的甲醇期货价格预测模型预测值所获得的收益达到 414,年化收益率达到 21.25%,超过同期 10 年期固定国债利率 5.8 倍多。策略的泛化结果表明,上文提出的甲醇期货价格预测模型在结合使用改进后的 Aberration 交易策略能够实现很好的投资收益。

RESEARCH ON METHANOL FUTURES PRICE FORECAST AND TRADING STRATEGY BASED ON RANDOM FOREST METHOD

Ji Chenyang, Lin Jie

(School of Economics & Management, Tongji University, Shanghai 200092, China)

Abstract: Methanol is an important new type of clean energy. Its futures products have been on the market for nearly 10 years. This article uses machine learning technology to study the prediction models and trading rules of methanol futures prices. The paper uses the random forest algorithm to construct a methanol futures price prediction model, uses the fundamental characteristics of the upstream and downstream products of the methanol industry chain as input variables, and normalizes the input variables to construct a control model, combined with the use of Aberration strategy ideas to build a set of trading strategies. Use Sharpe ratio to screen trading strategies and construct an effective quantitative trading model for methanol futures. The research results show that the model in this paper can achieve an annualized rate of return that is 1.2 times higher than the rate of 10-year Treasury bonds in the same period while ensuring a good generalization ability of the trading strategy. It also shows that the use of the fundamental characteristics data of the upstream and downstream products of the methanol industry chain can well explain the methanol futures price, and combined with the use of Sharpe ratio screening, the trading rules based on the Aberration trading strategy idea can construct an effective methanol futures quantitative investment model.

Keywords: methanol futures, price prediction, random forest, strategy generalization

参考文献

- [1] 陈标金,王锋.宏观经济指标、技术指标与国债期货价格预测——基于随机森林机器学习的实证检验[J].统计与信息论坛,2019,34(06):29-35.
- [2] 范俊明,刘洪久,胡彦蓉.基于 LSTM 深度学习的大豆期货价格预测[J].价格月刊,2021(02):7-15.
- [3] 周亚.基于 XGBoost 与弹性网络回归的集成模型对上证指数日极差的预测分析[J].中国物价,2021,{4}(05):68-71.
- [4] 方匡南,吴见彬,朱建平,谢邦昌.随机森林方法研究综述[J].统计与信息论坛,2011,26(03):32-38.

- [5] 姬培森.基本面和技术面在期货市场分析中的价值研究[J].时代金融,2020,{4}(13):93-94.
- [6] 李顺平,朱昂明.基本面数据对螺纹钢期货价格的指示作用研究[J].时代金融,2019,{4}(17):96-99.
- [7] 金志峰,崔楠楠,宋歆欣.铁矿石市场价格主导因素实证研究[J].冶金经济与管理,2021,{4}(01):8-13.
- [8] 马郑玮,琚心然,姚瑜,冯异立.基于产业链视角对石油期货与 PTA 期货价格关联性的分析[J/OL].现代化工,{3},{4}{5}:1-16[2021-07-20].<http://kns.cnki.net/kcms/detail/11.2172.tq.20200910.0845.005.html>.
- [9] 张宗成,王骏.世界铁矿石的生产与贸易和我国铁矿石供需的经济学分析[J].国际贸易问题,2005,{4}(09):28-33.
- [10] Liran Einav,Jonathan Levin. Economics in the age of big data[J]. Science,2014,346(6210).
- [11] Alessandro Vespignani. Predicting the Behavior of Techno-Social Systems[J]. Science, 2009, 325(5939).
- [12] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001).
- [13] Pérez-Rave,Correa-Morales,González-Echavarría. A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes[J]. Journal of Property Research, 2019,36(1).