

基于变系数模型的原油价格实证研究

漆艳霞, 刘艳萍 *

南宁师范大学数学与统计学院, 广西 南宁 531000

DOI:10.61369/ASDS.12188

摘 要 : 基于2023年1月3日至2025年3月18日 WTI原油期货收盘价格数据, 选取原油价格滞后3期作为解释变量建立变系数模型。本文采用B样条方法估计模型的未知系数函数, 运用Rolling方法进行多步向前预测。另外, 通过和传统的ARMA模型以及Holt-Winters - Additive模型比较, 本文构建的变系数模型具有较好的拟合和预测效果, 明显优于对比模型。

关 键 词 : 原油价格; 变系数模型; B样条

Empirical Study of Crude Oil Prices based on Variable Coefficient Modeling

Qi Yanxia, Liu Yanping*

School of Mathematics and Statistics, Nanning Normal University, Nanning, Guangxi 531000

Abstract: Based on the WTI crude oil futures closing price data from January 3, 2023 to March 18, 2025, a variable coefficient model is established by selecting crude oil price lagged by 3 periods as the explanatory variable. The unknown coefficient functions of the varying coefficient model are estimated by B-spline and the multi-step-forecasts are given through a rolling procedure. The empirical results show that the effect of crude oil price on itself is nonlinear. In addition, by comparing with the traditional ARMA model and Holt-Winters - Additive model, the variable coefficient model constructed in this paper has a better fitting and forecasting effect, which is obviously better than the comparison model.

Keywords: crude oil price ; varying coefficient models; B-spline

引言

作为全球金融市场重要的一种衍生品,原油期货在金融投资和风险控制中起着重要的作用^[1]。1983年,纽约商业交易所首次推出了轻质低硫的西得克萨斯州中间基原油(West Texas Intermediate, WTI)期货合约拉开了原油期货交易的序幕^[2]。原油兼具金融与能源双重属性,其价格往往具备着剧烈的波动性,严重影响全球军事环境与国际经济形势。我国作为全球最大的原油进口国,对海外原油依赖程度不断加深,提高国际原油价格预测精度、研究油价波动的潜在影响因素,对保障我国经济稳健运行、优化宏观战略调整以及有效规避交易风险等具有极为重要的理论价值和现实意义^[3]。实际上,在市场供需博弈、美元汇率波动、投机交易活跃、地缘政治冲突以及自然灾害等多重因素的综合作用下,国际原油价格呈现出高噪声、非线性和非平稳的复杂特征,导致对国际原油价格精确预测变得十分艰难。

当前的原油价格预测模型主要依赖于传统的时间序列模型或者多元统计模型^[4]。这些模型因其在处理时间序列中短期波动和趋势的能力而被广泛应用。这些模型大多基于线性假设,而原油期货价格通常表现出复杂非线性波动^[5]。因此,这些模型存在显著的局限性,往往在进行预测原油期货价格时常不够准确,也难以全面解释价格的波动。为解决上述局限,Huang和Shen提出的变系数模型提供了一种更为灵活的解决方案^[6]。变系数模型能够有效地处理数据中的非线性特征,从而提升了对复杂系统的预测能力。变系数模型已经广泛应用于社会、经济和科技等领域,研究表明它能够有效地捕捉数据的内在结构,以及提高预测的准确性和解释力^[7,8]。本文基于2023年1月3日至2025年3月18日 WTI原油收盘价格,建立变系数模型,深入探究原油期货价格的未来走势。

基金项目: 国家自然科学基金资助项目(12361049), 广西自然科学基金项目(2025GXNSFBA069559), 广西高校中青年教师科研基础能力提升项目(2023KY0389)。

作者简介: 漆艳霞(1991—),女,湖北荆州人,硕士生,主要研究非参数估计。Email: 1127450166@qq.com;

通信作者: 刘艳萍(1990—),女,广西钦州人,博士,讲师,主要研究非参数估计。Email: liuyanping_nnnu@163.com。

一、模型构建与估计

变系数模型的一般形式为：

$$Y_t = X_t^T \alpha(U_t) + \varepsilon_t \quad (1)$$

其中 $X_t = (X_{t1}, \dots, X_{td})^T$, $\alpha(U_t) = (\alpha_1(U_t), \dots, \alpha_d(U_t))^T$ 。假设 $\alpha_j(\cdot)$, $j=1, \dots, d$ 是未知的光滑函数, 随着 U_t 的变化而变化, 称为系数函数。 X_t 和解释变量 U_t 既可以是 Y_t 的滞后变量也可以是外生变量。 ε_t 是模型的误差, 满足 $E(\varepsilon_t | X_t, U_t) = 0$ 。

(一) 变系数函数的 B 样条估计

系数模型的未知函数主要的非参数估计方法有核估计 [9, 10] 和样条估计 [11-13] 等。由于样条估计方法是基于全局的逼近, 具有灵活性和计算快捷的优点, 本文采用 B 样条法来估计模型中的系数函数。

假设 $U = (u_1, \dots, u_k)$, $U = (u_1, u_2, \dots, u_k)$, 其中 $u_0 = a < u_1 < \dots < u_k < b = u_{k+1}$, $u_0 = a < u_1 < u_2 < \dots < u_k < b = u_{k+1}$, U 为 $[a, b]$ 上的 k 个内部节点序列, 在空间 $S(p, U)$ 上, B 样条基函数的一个递归形式如下:

其中, 0 次 B 样条函数基为:

$$B_{i,0}(x) = \begin{cases} 1, & x \in [u_i, u_{i+1}) \\ 0, & x \notin [u_i, u_{i+1}) \end{cases},$$

P 次 B 样条函数基的递推公式为:

$$B_{i,p}(x) = \frac{x - u_i}{u_{i+p} - u_i} B_{i,p-1}(x) + \frac{u_{i+p+1} - x}{u_{i+p+1} - u_{i+1}} B_{i+1,p-1}(x),$$

$[u_i, u_{i+1})$ 是 B 样条函数的第 i 个节点区间, $i=1, 2, \dots, k+1$ 。为了方便, 上式中上标 P 忽略, 记 B 样条基向量为 $B(t) = (B_1(t), B_2(t), \dots, B_N(t))^T$, $N = p + k + 1$ 。

对于任意的 $s(t) \in S(p, U)$, 存在 $\beta \in R^{p+k+1}$, 使得 $s(t) = B^T(t)\beta$ 称为 P 次的 B 样条函数。本文中系数函数的 B 样条估计均使用相同的基函数阶数和相同的节点个数 [14]。

假设式 (1) 的系数函数 $\alpha_j(u) (j=1, 2, \dots, d)$ 是 $\beta_j(u) (j=1, \dots, p)$ 光滑的, 它可以通过 B 样条函数 $\alpha_j^*(u) \beta_j^*(u)$ 近似。因此, 存在基函数 $B_s(\cdot) B_{js}(\cdot)$ 和常数 $\beta_{js}^*, \beta_{js}^*$, $s=1, 2, \dots, K$, $K = 1, 2, \dots, K_j$ 使得

$$\alpha_j(u) \approx \alpha_j^*(u) = \sum_{s=1}^K \beta_{js}^* B_s(u).$$

下面给出推导 α_j 估计量的具体计算过程。

设 $i_1, i_2, \dots, i_d$ 为滞后阶数, 可以不连续, 则变系数模型 (1) 可变形为如下形式:

$$\begin{aligned} Y_t &= \alpha_1(U_t) Y_{t-i_1} + \alpha_2(U_t) Y_{t-i_2} + \dots + \alpha_d(U_t) Y_{t-i_d} + \varepsilon_t \\ &= \sum_{j=1}^d \alpha_j(U_t) Y_{t-i_j} + \varepsilon_t \\ &\approx \sum_{j=1}^d \left(\sum_{s=1}^{K_j} \beta_{js}^* B_s(U_t) \right) Y_{t-i_j} + \varepsilon_t \end{aligned} \quad (2)$$

其中, $t = i_d + 1, i_d + 2, \dots, n$ 。

记 $\beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jK_j})^T$, $W_j = (B_1(U_j), B_2(U_j), \dots, B_{K_j}(U_j))^T$,

$\beta = (\beta_{11}, \dots, \beta_{1K_1}, \dots, \beta_{d1}, \dots, \beta_{dK_d})^T$,

$$X = \begin{pmatrix} Y_{i_d+1-i_1} W_{i_d+1-i_1} & Y_{i_d+1-i_2} W_{i_d+1-i_2} & \dots & Y_{i_d+1-i_d} W_{i_d+1-i_d} \\ Y_{i_d+2-i_1} W_{i_d+2-i_1} & Y_{i_d+2-i_2} W_{i_d+2-i_2} & \dots & Y_{i_d+2-i_d} W_{i_d+2-i_d} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n-i_1} W_{n-i_1} & Y_{n-i_2} W_{n-i_2} & \dots & Y_{n-i_d} W_{n-i_d} \end{pmatrix},$$

$Y = (Y_{i_d+1}, Y_{i_d+2}, \dots, Y_n)^T$, $\hat{\alpha} = (\varepsilon_{i_d+1}, \varepsilon_{i_d+2}, \dots, \varepsilon_n)^T$, 则模型 (2) 的矩

阵形式为

$$Y = X\beta + \varepsilon.$$

通过最小二乘法来估计 β , 即最小化 (3) 式

$$l(\beta) = (Y - X\beta)^T (Y - X\beta) \quad (3)$$

由此可得到 β 的最小二乘估计 $\hat{\beta} = (X^T X)^{-1} X^T Y = (\hat{\beta}_{11}, \dots, \hat{\beta}_{1K_1}, \dots, \hat{\beta}_{d1}, \dots, \hat{\beta}_{dK_d})^T$, 则 $\alpha_j(u)$ 的估计量为 $\hat{\alpha}_j(u) = \sum_{s=1}^{K_j} \hat{\beta}_{js} B_s(u)$, $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_d)^T$ 。

二、数据来源及参数确定方法

(一) 数据来源

本文研究选取 WTI 原油期货价格的日收盘价 (单位为美元 / 桶) 作为原油价格, 时间跨度均为 2023 年 1 月 3 日至 2025 年 3 月 18 日, 原油期货会因为各种节假日和周末等进行休市和停牌, 本文将这些缺失的日期予以删除, 共剩余 575 个数据。其中, 2023 年 1 月 3 日至 2025 年 2 月 25 日的 560 个数据作为模型建立的样本, 2025 年 2 月 26 日至 2025 年 3 月 18 日的 15 个样本作为预测样本。数据分别来源于英为财经网站 (<https://cn.investing.com>)。

(二) 节点数的确定和变量选择

为了提高计算效率高和数值稳定性, 本文系数函数的 B 样条估计均采用三次样条和等间距放置节点方式, 并采用 AIC 、 BIC 和 AIC_c 信息准则确定最优的节点个数。定义为:

$$\begin{aligned} AIC &= \log\left(\frac{RSS}{n}\right) + 2 \times \frac{p}{n}, \quad BIC = \log\left(\frac{RSS}{n}\right) + \log(n) \times \frac{p}{n}, \\ AIC_c &= AIC + \frac{2(p+1)(p+2)}{n(n-p-2)}, \end{aligned}$$

其中 n 为样本容量, p 是待估计参数个数, RSS 为模型残差平方和。

在本例中, 将原油价格滞后数据作为解释变量, 假定模型显著变量个数 $p_{\max} = 12$, $\max(1, 2, \dots, j) \leq p_{\max}$, $1 \leq d \leq p_{\max}$ 分别采用前进法和后退法来确定变量集合。考虑如下候选模型:

$$Y_t = \sum_{j \in S_d} \alpha_j(Y_{t-d_j}) Y_{t-j} + \varepsilon_t,$$

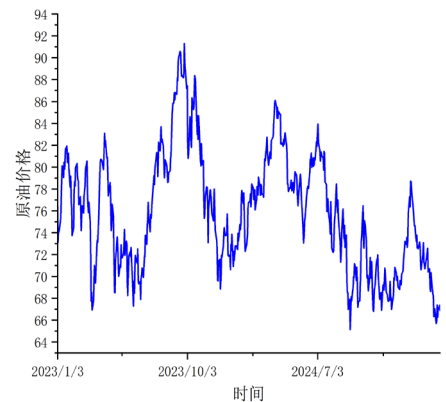
其中, 变量集合为 $\{(Y_{t-d_1}, S_{d_1}^*), (Y_{t-d_2}, S_{d_2}^*), \dots, (Y_{t-d_{p_{\max}}}, S_{d_{p_{\max}}}^*)\}$ 。

对于给定的 d , 前进法在添加阶段, 每次添加一个滞后变量, 该滞后变量使得模型的 AIC 、 BIC 或 AIC_c 准则函数值达到最小, 这里要添加的滞后变量来自未被选入模型中的滞后变量所组成的集合, 当模型中的滞后变量不再增加时, 添加过程结束, 从而得到滞后变量的一个添加顺序; 后退法在删减阶段, 每次删减一个滞后变量, 即去掉该显著滞后变量后, 模型的 AIC 、 BIC 或 AIC_c 准则函数值达到最小, 这里所删减的滞后变量来自模型中滞后变量所组成的集合, 当模型中不再删减滞后变量时, 删减过程结束, 从而得到显著滞后变量的一个删减顺序 [15]。选择 AIC 、 BIC 或 AIC_c 最小的为 $\{Y_{t-d}, S_d^*\}$ 。本文用 R 语言运行程序, 前进法和后退法得到变量集合为 $\{Y_{t-3}, Y_{t-5}\}$ 。

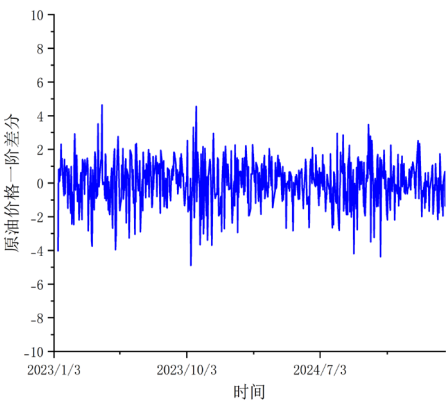
三、结果分析

(一) 数据平稳性分析

根据数据自身特点及建模的需要，文中采用一阶差分运算，差分后的原油价格记为 Y_t 。



> 图1 原油价格时序图



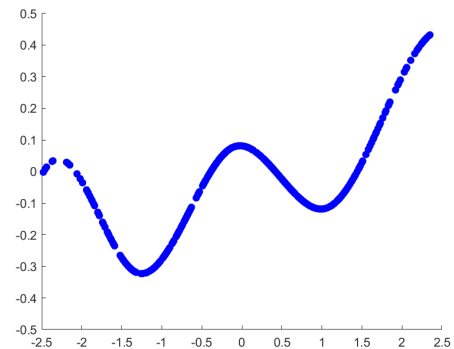
> 图2 一阶差分原油价格时序图

图1和图2分别绘制了原油价格和一阶差分变换的原油价格时序图，从图1可以看出原油价格波动性较大，数据呈现明显的非平稳性，对原油价格数据进行单位根检验，结果显示其检验值均大于0.05。图2中显示的一阶差分原油时序图呈现平稳趋势，在R语言中对差分后的数据进行单位根检验，结果显示其检验值均小于0.05，这表明经过差分变换后的数据是一个平稳序列。

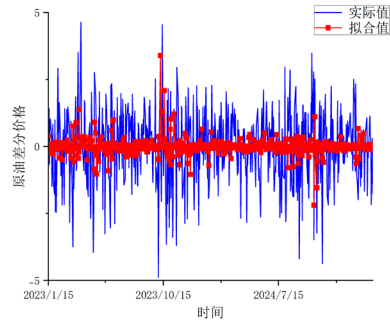
(二) 模型的建立

根据上述分析，本文所建立的变系数回归模型的具体形式为：

$$Y_t = \alpha(Y_{t-3})Y_{t-5} + \varepsilon_t \quad (4)$$

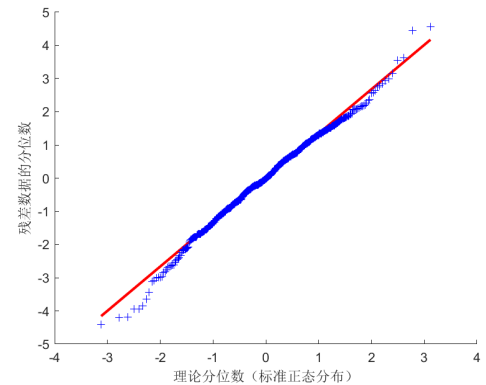


> 图3 $\beta(\cdot)$ 的估计效果图

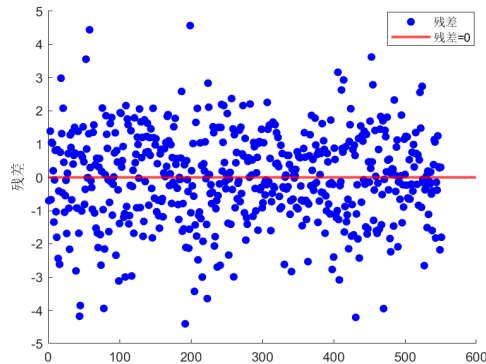


> 图4 原油价格数据的实际值和拟合值

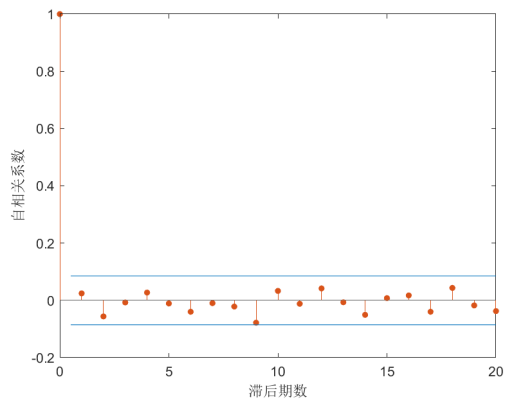
图3为变系数函数的系数估计效果图，可以看出系数呈现强烈的非线性结构，建模的结果显示第 t 期原油价格受上五期原油价格的影响，并且这种影响呈非线性关系。图4描绘了原油价格的实际值和模型(4)的拟合值，可以看出，模型(4)的拟合效果很好。



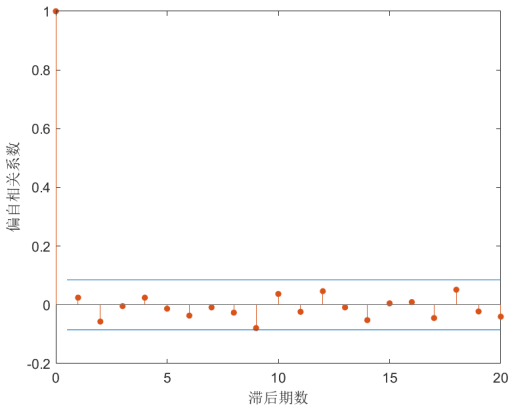
> 图5 残差 QQ图



> 图6 模型(4)残差散点图



> 图7 模型 (4)残差自相关图



> 图8 模型 (4)残差偏自相关图

对模型 (4)的残差进行分析,图5和图6分别给出了模型 (4)的残差 QQ 图和残差散点图,图7和图8分别给出了残差序列自相关系数图和偏自关系数图。图5中,残差分位点基本分布在45度对角线上,并且近似地在一条直线附近,表明残差序列服从正态分布。图6中,残差在0附近随机分布,大部分残差值集中在-2到2之间,表明模型的预测误差相对较小。图7和图8中,自相关和偏自相关条形都落在蓝色虚线内,说明残差序列没有显著的自相关性和偏自相关性,模型拟合较好。

表1 Ljung-Box 检验结果

自由度	检验统计量 X-squared	Ljung-Box 检验 P 值
3	2.0933	0.5533
6	3.4387	0.7521
9	7.1052	0.6262

表2 3种模型对原油价格预测结果对比

日期	实际值	变系数模型		ARMA 模型		Holt-Winters - additive 模型	
		预测值	相对误差 (%)	预测值	相对误差 (%)	预测值	相对误差 (%)
2025/2/26	68.62	68.91	0.42	69.13	0.74	69.03	0.60
2025/2/27	70.35	68.83	-2.16	68.79	-2.22	69.04	-1.86
2025/2/28	69.76	68.87	-1.28	69.04	-1.03	69.05	-1.02
2025/3/3	68.37	68.84	0.68	69.19	1.20	68.88	0.75
2025/3/4	68.26	68.82	0.82	69.25	1.45	68.90	0.93
2025/3/5	66.31	68.79	3.73	68.98	4.03	69.00	4.06
2025/3/6	66.36	68.77	3.63	69.32	4.46	69.01	3.99
2025/3/7	67.04	68.75	2.54	69.36	3.46	69.02	2.95
2025/3/10	65.68	68.72	4.62	69.28	5.49	68.85	4.82
2025/3/11	65.93	68.70	4.20	69.13	4.85	68.86	4.45
2025/3/12	67.38	68.68	1.93	69.48	3.12	68.97	2.36

自由度	检验统计量 X-squared	Ljung-Box 检验 P 值
12	8.7436	0.7247
15	10.212	0.8062
18	12.293	0.8317

对残差进行白噪声检验, Ljung-box 检验结果为 p 值都大于 0.05,说明残差序列是白噪声序列,表明建立的模型是有效的。

(三) 模型的预测

对于变系数模型进行多步向前预测是很困难的^[6]。但由于样条估计提供了显式模型,我们采用 Rolling 方法^[6]进行多步向前预测,其思想是考虑30个子序列 $\{Y_1, \dots, Y_T\}, T = 531, \dots, 560$,对第 $k(k = 1, \dots, 30)$ 个子序列拟合变系数模型 (2) 并计算残差 $\hat{\varepsilon}_t$ 。利用拟合的模型,在 $\{Y_1, \dots, Y_T\}$ 之后,我们可以递归地生成1000个序列 $\{\hat{Y}_{T+1,j}, \dots, \hat{Y}_{T+10,j}\}, 1 \leq j \leq 1000$,其中 ε_t 通过对残差进行随机抽样得到。对这1000个序列样本取平均值,可以得到第 h 步平均预测值 $\hat{Y}_{T+h}$ 。

为了进一步对比模型预测效果,我们采用文献 [16,17] 的 ARMA 模型和 Holt-Winters - Additive 模型来拟合一阶差分变换后的原油收盘价格,模型形式为:

● ARMA(5,5)—GARCH(1,1)模型:

$$Y_t = -0.02 + 0.24Y_{t-3} + 0.77Y_{t-4} - 0.33Y_{t-5} + \varepsilon_t - 0.30\varepsilon_{t-3} - 0.78\varepsilon_{t-4} + 0.31\varepsilon_{t-5},$$

● Holt-Winters - additive 模型:

$$y_{t+m} = (S_t + mB_t) + C_{t-l+m},$$

其中: y_{t+m} 为第 $t+m$ 期的黄金价格预测值; S_t 为第 t 期的水平因子; B_t 为第 t 期的趋势因子; C_t 为第 t 期的季节因子。

表1给出了2025年2月26日至2025年3月18日的原油交易日不同模型的预测值。本文中,我们采用平均绝对误差 (MAE)、均方根误差 (RMSE)、平均相对误差 (MRE)、和均方误差 (MSE) 来说明模型预测效果的优劣:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|, \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2},$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|, \quad MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

其中, Y_i 表示第 i 个观测值的实际值, $\hat{Y}_i$ 表示第 i 个观测值的预测值, n 是观测值的个数。

日期	实际值	变系数模型		ARMA 模型		Holt-Winters - additive 模型	
		预测值	相对误差 (%)	预测值	相对误差 (%)	预测值	相对误差 (%)
2025/3/13	66.27	68.65	3.59	69.38	4.69	68.98	4.09
2025/3/14	66.91	68.63	2.57	69.26	3.51	68.98	3.10
2025/3/17	67.37	68.61	1.84	69.25	2.78	68.81	2.14
2025/3/18	66.90	68.58	2.51	69.53	3.93	68.83	2.89

其中，相对误差 = $(\hat{Y}_i - Y_i)/Y_i$ 。

表3 3种模型对原油价格预测效果优劣对比

模型名称	变系数模型	ARMA 模型	Holt-Winters - additive 模型
MAE	1.629552	2.094434	1.782557
RMSE	1.827843	2.300441	1.996895
MRE	0.024358	0.031309	0.026665
MSE	3.341012	5.292028	3.987590

表1为3种模型对原油价格预测值及相对误差，从表1中可以看出变系数模型预测相对误差整体上要小于 ARMA 模型和 Holt-Winters - additive 模型。表2给出了3种模型对原油价格预测效果优劣对比，可以得到变系数模型的平均绝对误差（MAE）、均方根误差（RMSE）、平均相对误差（MRE）、和均方误差（MSE）分别为1.629552、1.827843、0.024358、3.341012，都

分别小于另外两种模型。模型对应的评价指标的值越小，表明模型的预测效果越好，表1和表2的结果充分说明了变系数模型在预测原油价格时的优越性。

四、结束语

本文基于时间序列分析方法，在对原油期货价格数据建立变系数模型时，将原油价格滞后3期作为函数解释变量。综上分析，第 t 期原油价格受上五期原油价格的影响，并且影响呈非线性关系，系数函数是原油价格滞后3期的非线性函数。此外，建立的变系数模型比 ARMA 模型和 Holt-Winters —Additive 模型预测的误差小，精度更高，为原油价格预测提供了一种新方法。

参考文献

[1] 蒋锋, 胡成雨, 王辉. 基于多源多任务自动编码器的原油期货价格预测 [J]. 系统工程理论与实践, 2025, 45(02): 702-716.

[2] 张金良, 谭忠富. 基于混合模型的原油价格混沌预测方法 [J]. 运筹与管理, 2013, 22(05): 166-72.

[3] 徐进, 王雪莲. 国际原油现货价格预测指标构建及应用研究——以 WTI 原油价格为例 [J]. 价格理论与实践, 2022, (05): 118-121+206.

[4] 张小艺, 柴泳旭. 基于 ARIMA-GARCH 的原油期货价格预测研究 [J]. 北方经贸, 2024, (12): 88-92.

[5] 张晓, 成晟, 董睿, 等. 基于尺度分析和动态误差修正的油价选择性集成预测方法 [J]. 系统科学与数学, 2023, 43(10): 2451-66.

[6] Huang J Z, Shen H. Functional Coefficient Regression Models for Non-linear Time Series: A Polynomial Spline Approach [J]. Scandinavian Journal of Statistics, 2004, 31(4): 515-34.

[7] 邓金兰, 王彬寰, 樊仕利. 变系数模型的变量选择在股票数据中的应用 [J]. 四川大学学报 (自然科学版), 2009, 46(06): 1585-91.

[8] Xia X. Model averaging prediction for nonparametric varying-coefficient models with B-spline smoothing [J]. Statistical Papers, 2021, 62(6): 1-21.

[9] Cai Z, Fan J, Yao Q. Functional-Coefficient Regression Models for Nonlinear Time Series [J]. Journal of the American Statistical Association, 2000, 95(451): 941-56.

[10] 张巍巍, 萨如拉, 冯三营. 缺失数据下异方差半变系数模型的约束统计推断 [J]. 应用数学, 2022, 35(03): 617-27.

[11] Du J, Cui X. Semiparametric estimation in generalized additive partial linear models with nonignorable nonresponse data [J]. Statistical Papers, 2024, 65, 3235 - 3259 .

[12] 赵静. 空间计量单指标变系数模型的估计及其应用 [D]. 2021.

[13] 刘艳霞, 王芝皓, 芮荣祥, 等. 广义函数型部分变系数混合模型的估计 [J]. 系统科学与数学, 2021, 41(06): 1742-60.

[14] 卢一强, 曾林蕊. 变系数模型 B 样条 M 估计的收敛性 [J]. 应用概率统计, 2003, 19(04): 415-23.

[15] 姜爱平. 我国人口时间序列的变系数预测模型 [J]. 数理统计与管理, 2008, (05): 755-9.

[16] 李丽. 基于 ARMA-GARCH 模型的股市量价动态关系研究 [J]. 统计与决策, 2011, (04): 144-6.

[17] 伊力扎提·艾提提. 中国消费者价格指数预测模型的选择 [J]. 统计与决策, 2022, 38(04): 68-73.