

基于样条估计的非参数方差模型的区间预测

武新乾, 于雪*, 王军香, 姬静怡

河南科技大学数学与统计学院, 河南 洛阳

收稿日期: 2026年5月23日; 录用日期: 2026年6月17日; 发布日期: 2026年6月23日

摘要

为了探寻固定设计下非参数方差模型的区间预测问题, 基于方差函数的多项式样条估计和线性外推法, 构建了响应变量数据的预测区间。数值模拟算例表明: 随着样本容量的增加, 平均区间覆盖率逐渐接近给定的置信水平, 预测的平均区间宽度逐渐减小; 在节点个数选择上, AIC准则优于BIC准则。此外, 对河南省第三产业贡献率数据进行了应用分析, 得到了“十五五”时期的预测区间。

关键词

方差函数, 正态过程, 样条估计, 区间预测

Interval Prediction of Nonparametric Variance Models Based on Spline Estimation

Xinqian Wu, Xue Yu*, Junxiang Wang, Jingyi Ji

School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang Henan

Received: May 23, 2026; accepted: June 17, 2026; published: June 23, 2026

Abstract

In order to solve the problem of interval prediction of nonparametric variance models with fixed design, a prediction interval of response variable is constructed by the polynomial spline estimate of variance function and the linear extrapolation in this paper. The results of a simulation example show that as the sample size increases, the average interval coverage rate gradually approaches the given confidence level, and the average width of the prediction intervals gradually decreases. In terms of knot number selection, the AIC criterion outperforms the BIC criterion. Furthermore, an

*通讯作者。

application analysis is conducted on the contribution rate of the tertiary industry to the increase of the GDP in Henan province. And the prediction intervals for the 15th Five-Year Plan period are obtained.

Keywords

Variance Function, Normal Process, Spline Estimation, Interval Prediction

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在计量经济学和金融时间序列等领域中，常会遇到异方差现象。考虑一类反映异方差现象的非参数方差模型

$$Y_t = \sigma(x_t) \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (1)$$

其中 $x_t = t/n$ 为固定设计点， Y_t 为响应变量， $\sigma(x) > 0$ 为未知光滑函数，称 $\sigma^2(x)$ 为方差函数， $\{\varepsilon_t\}$ 是严平稳 α 混合序列，满足 $\varepsilon_t \sim N(0, 1)$ 。

对于非参数方差模型，学者们主要集中于方差函数的估计方法和理论研究[1]-[3]。近年来，一些学者对非参数回归模型的预测问题产生了兴趣。非参数回归模型是模型(1)的推广，即

$$Y_t = m(x_t) + \sigma(x_t) \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (2)$$

其中 $m(x)$ 为未知回归函数或均值函数。Cheng 等(2025)在同方差($\sigma(x)$ 为常数)下提出了一个基于核方法的多步非参数预测回归模型[4]。Li 和 Zhao (2020)应用基于核方法的非参数回归模型预测水力发电价格[5]。李威(2022)在同方差下基于核方法的非参数组合预测模型对江西省 GDP 进行了实证分析[6]。Xiao 等(2024)使用非参数回归预测 MU-MIMO 无线信道，并对 k 近邻方法和局部线性方法进行了比较分析[7]。武新乾和张刚(2016)在残差自回归模型下讨论了基于多项式样条的两种预测方法[8]。武新乾和程芳(2017)对模型(2)讨论了基于多项式样条的三种预测方法[9]。王彩霞等(2010)基于同方差非参数回归核技术讨论了短期风电功率的点预测和区间预测[10]。李晋云和武新乾(2021)基于样条方法构建了同方差非参数回归模型的区间预测[11]。然而，对于异方差非参数回归模型的预测研究相对滞后，尤其是区间预测问题尚未见有文献研究。

本文针对非参数方差模型，基于多项式样条估计方法，构造响应变量数据 Y_t 的预测区间，并且进行模拟比较和应用分析。

2. 样条估计与预测区间

由模型(1)可知

$$Y_t \sim N(0, \sigma^2(x_t)), \quad \frac{Y_t}{\sigma(x_t)} \sim N(0, 1),$$

于是，响应变量数据 Y_t 的置信水平为 $1 - \alpha$ ($0 < \alpha < 1$) 的置信区间为

$$\left[-u_{\alpha/2} \sigma(x_t), u_{\alpha/2} \sigma(x_t)\right], \quad (3)$$

其中 $u_{\alpha/2}$ 为标准正态分布的上侧 $\alpha/2$ 分位数。

鉴于 $\sigma(x_t)$ 未知, 无法直接应用(3)式计算置信区间。为此, 需构造 $\sigma(x_t)$ 的相合估计。

将区间 $[0,1]$ 等距剖分, 其内节点有 k 个, 与之相应的 ν 次样条函数空间 $S_{k,\nu}$ 的基函数记为 $B_s(x)$ ($s=1,2,\dots,k+\nu+1$)。又记 $K=k+\nu+1$, $(\cdot)'$ 表示向量或矩阵的转置。注意到

$$Y_t^2 = \sigma^2(x_t) + \sigma^2(x_t)(\varepsilon_t^2 - 1), \quad t=1,2,\dots,n, \quad (4)$$

且光滑函数 $\sigma^2(x)$ 可由样条函数空间 $S_{k,\nu}$ 中的函数 $\sum_{s=1}^K \beta_s B_s(x)$ 所近似。回归模型(4)可近似表示为线性回归模型, 即

$$Y_t^2 \approx \sum_{s=1}^K \beta_s B_s(x_t) + \sigma^2(x_t)(\varepsilon_t^2 - 1), \quad t=1,2,\dots,n. \quad (5)$$

由最小二乘法可得参数 $\beta = (\beta_1, \beta_2, \dots, \beta_K)'$ 的估计为

$$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_K)' = (X'X)^{-1} X'Z, \quad (6)$$

其中

$$X = \begin{pmatrix} B_1(x_1) & B_2(x_1) & \cdots & B_K(x_1) \\ B_1(x_2) & B_2(x_2) & \cdots & B_K(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ B_1(x_n) & B_2(x_n) & \cdots & B_K(x_n) \end{pmatrix}, \quad Z = (Y_1^2, Y_2^2, \dots, Y_n^2)'$$

从而可得方差函数 $\sigma^2(x)$ 的样条估计为

$$\hat{\sigma}^2(x) = \sum_{s=1}^K \hat{\beta}_s B_s(x). \quad (7)$$

在文献[2]的条件下, $\hat{\sigma}^2(x)$ 是 $\sigma^2(x)$ 的相合估计。

将(3)式中的 $\sigma(x_t)$ 替换为 $\hat{\sigma}(x_t) = \sqrt{\hat{\sigma}^2(x_t)}$, 可得响应变量数据 Y_t 的置信水平为 $1-\alpha$ 的置信区间为

$$[-u_{\alpha/2} \hat{\sigma}(x_t), u_{\alpha/2} \hat{\sigma}(x_t)]. \quad (8)$$

对于样本外预测, 记预测步数为 d ($d=1,2,\dots,l$), 根据(8)式可知 Y_{n+d} 的置信水平为 $1-\alpha$ 的置信区间为

$$[-u_{\alpha/2} \hat{\sigma}(x_{n+d}), u_{\alpha/2} \hat{\sigma}(x_{n+d})]. \quad (9)$$

但是, $\hat{\sigma}(x_{n+d})$ 的值无法由(7)式直接计算得到。由文献[9]知, 在模型(2)预测的非外推法、线性外推法和非线性外推法中, 线性外推法是简单可行的方法。鉴于此, 采用线性外推法计算 $\hat{\sigma}(x_{n+d})$ 的值, 其递推公式为

$$\hat{\sigma}(x_{n+d}) = 2\hat{\sigma}(x_{n+d-1}) - \hat{\sigma}(x_{n+d-2}). \quad (10)$$

3. 模拟算例

考虑非参数方差模型(1), 其中

$$\sigma(x) = 0.4 \sin(\pi x) + 0.1, \quad \varepsilon_t = 0.5\varepsilon_{t-1} + e_t + 0.5e_{t-1},$$

这里 $\{e_t\}$ 为独立同分布序列, 且 $e_t \sim N(0, 3/7)$ 。

鉴于低阶样条(比如一次样条、二次样条)欠光滑和高阶样条易于过拟合, 选取常用的三次 B-样条基函数构造响应变量数据 Y_i 的置信区间, k 由 AIC 和 BIC 准则从 1 到 10 之间进行自动选择, 其中

$$AIC = \ln(RSS/n) + 2K/n, \quad BIC = \ln(RSS/n) + \ln(n) \cdot K/n, \quad (11)$$

这里 $RSS = \sum_{i=1}^n [Y_i^2 - \hat{\sigma}^2(x_i)]^2$ 为残差平方和。

为了评价区间预测方法的优劣, 选取平均区间覆盖率 MICP 和平均区间宽度 AIW 作为评价指标, 其中

$$MICP = \frac{1}{m} \sum_{i=1}^m ICP_i, \quad AIW = \frac{1}{m} \sum_{i=1}^m IW_i. \quad (12)$$

这里, m 为模拟次数; $ICP_i = n_{yi}/n$, n_{yi} 为第 i 次模拟中响应变量的观测值 Y_1, Y_2, \dots, Y_n 位于相应置信区间内的个数; $IW_i = \frac{1}{n} \sum_{i=1}^n (U_{ait} - L_{ait}) = \frac{1}{n} \sum_{i=1}^n 2u_{\alpha/2} \hat{\sigma}_{(i)}(x_i)$ 为第 i 次模拟中观测值 Y_1, Y_2, \dots, Y_n 的置信区间宽度的平均值, L_{ait} 和 U_{ait} 分别为第 i 次模拟中 Y_i 的置信区间的左端点和右端点, $\hat{\sigma}_{(i)}(x_i)$ 为第 i 次模拟中 Y_i 的标准差的估计值。

在模拟中, 选取 $e_0 = 0$, 置信水平 $1 - \alpha = 0.95$; 样本容量 $n = 50, 100, 200, 400$ 和 600 , 产生 $n + l$ 个数据, $l = 5$ 。利用前 n 个数据建立模型, 对剩余的 l 个数据进行预测。每种情况下重复模拟次数 $m = 1000$ 。

表 1 和表 2 分别给出了拟合的平均区间覆盖率 MICP 和平均区间宽度 AIW, 表 3 和表 4 分别给出了预测的平均区间覆盖率 MICP 和平均区间宽度 AIW。

由表 1 可知, 随着样本容量 n 的增加, 拟合的置信区间覆盖率的标准差逐渐递减, 平均区间覆盖率 MICP 稳定在 0.95 附近。由表 2 可知, 随着样本容量的增加, 拟合的置信区间宽度的标准差逐渐递减, 但平均区间宽度 AIW 有递增的情形, 反映了方差函数变化对估计精度的影响。此外, 由表 1 和表 2 还可以看出, AIC 准则下的平均区间覆盖率 MICP 更接近于置信水平 0.95, 标准差更小; AIC 准则下的平均区间宽度 AIW 和标准差整体上小于 BIC 准则下的值。在这个意义下, AIC 准则优于 BIC 准则。

由表 3 可知, 随着样本容量的增加, 预测的平均区间覆盖率 MICP 逐渐趋于置信水平 0.95, 而标准差有增加趋势。虽然如此, 但由表 4 可知, 随着样本容量的增加, 预测的平均区间宽度 AIW 及标准差均呈现递减态势, 说明了预测精度的提高。此外, 由表 3 和表 4 还可以看出, AIC 准则下的平均区间覆盖率 MICP 整体上更接近于置信水平 0.95, AIC 准则下的平均区间宽度 AIW 和标准差整体上小于 BIC 准则下的值, AIC 准则优于 BIC 准则。

Table 1. The fitted average interval coverage rate MICP and standard deviation (in the parentheses)

表 1. 模型拟合的区间覆盖率的均值 MICP 与标准差(括号内)

n	AIC 准则		BIC 准则	
	MICP	MICP - 0.95	MICP	MICP - 0.95
50	0.9562 (0.0335)	0.0062	0.9433 (0.0377)	-0.0067
100	0.9463 (0.0212)	-0.0037	0.9393 (0.0240)	-0.0107
200	0.9445 (0.0154)	-0.0055	0.9402 (0.0170)	-0.0098
400	0.9451 (0.0108)	-0.0049	0.9439 (0.0117)	-0.0061
600	0.9474 (0.0078)	-0.0026	0.9456 (0.0098)	-0.0044

Table 2. The fitted average interval width AIW and standard deviation (in the parentheses)
表 2. 模型拟合的区间宽度的均值 AIW 与标准差(括号内)

n	AIC 准则	BIC 准则
50	1.2244 (0.1802)	1.2273 (0.1966)
100	1.2869 (0.1478)	1.3378 (0.1528)
200	1.3296 (0.0982)	1.3542 (0.1033)
400	1.3681 (0.0831)	1.3785 (0.0773)
600	1.3722 (0.0625)	1.3837 (0.0655)

Table 3. The predicted average interval coverage rate MICP and standard deviation (in the parentheses)
表 3. 模型预测的区间覆盖率的均值 MICP 与标准差(括号内)

n	AIC 准则		BIC 准则	
	MICP	MICP - 0.95	MICP	MICP - 0.95
50	1 (0)	0.05	0.9980 (0.0198)	0.0480
100	0.9978 (0.0209)	0.0478	0.9989 (0.0150)	0.0489
200	0.9901 (0.0583)	0.0401	0.9925 (0.0517)	0.0425
400	0.9819 (0.0840)	0.0319	0.9861 (0.0734)	0.0361
600	0.9547 (0.1369)	0.0047	0.9614 (0.1334)	0.0114

Table 4. The predicted average interval width AIW and standard deviation (in the parentheses)
表 4. 模型预测的区间宽度的均值 AIW 与标准差(括号内)

n	AIC 准则	BIC 准则
50	0.9582 (0.3765)	1.0062 (0.3686)
100	0.8469 (0.2825)	0.9325 (0.3352)
200	0.7281 (0.3025)	0.8087 (0.3199)
400	0.6534 (0.2414)	0.7042 (0.2591)
600	0.5964 (0.2299)	0.6414 (0.2448)

4. 应用例子

选取 1981~2025 年黄河流域中河南省第三产业贡献率数据 $\{Y_{0t}\}$ (单位: %) (图 1), 其中 1981~2024 年的数据来自于《河南统计年鉴 2025》, 2025 年数据来自于河南省统计局官方网站发布的“2025 年河南省国民经济和社会发展统计公报”。图 1 中 $t=1, 2, \dots, 45$ 分别代表 1981 年、1982 年、……、2025 年。

为了便于建模分析, 对原始数据序列 $\{Y_{0t}\}$ 进行变换得到新的序列 $\{Y_t\}$ (图 2), 其中

$$Y_t = \ln Y_{0t} - 3.5985, \quad t = 1, 2, \dots, 45, \quad (13)$$

这里 3.5985 是序列 $\{\ln Y_{0t}\}$ 的均值。

基于模拟方法和算例结果, 本节采用 AIC 准则对均值为 0 的时间序列 $\{Y_t\}$ 进行建模分析。图 3 和图 4 分别给出了估计的方差函数 $\hat{\sigma}^2(x)$ 和误差序列 $\{\hat{\varepsilon}_t\}$, 其中

$$\hat{\varepsilon}_t = Y_t / \hat{\sigma}(x_t), \quad \hat{\sigma}(x_t) = \sqrt{\hat{\sigma}^2(x_t)}。$$

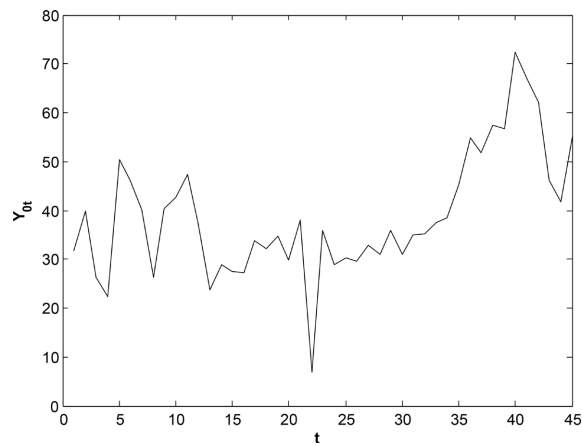


Figure 1. The sequence of the contribution rates of Henan's tertiary industry to the increase of the GDP $\{Y_{0t}\}$

图 1. 河南省第三产业贡献率序列 $\{Y_{0t}\}$

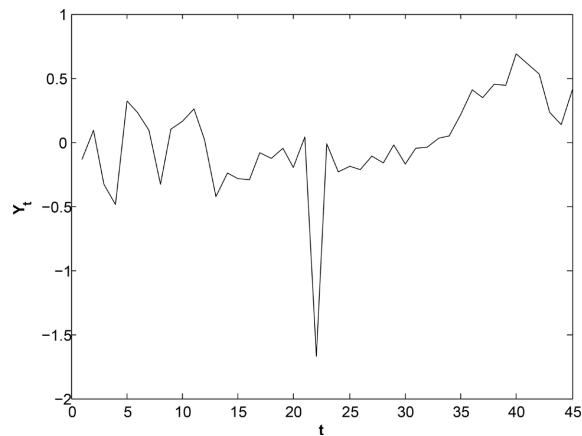


Figure 2. The transform sequence $\{Y_t\}$

图 2. 变换数据序列 $\{Y_t\}$

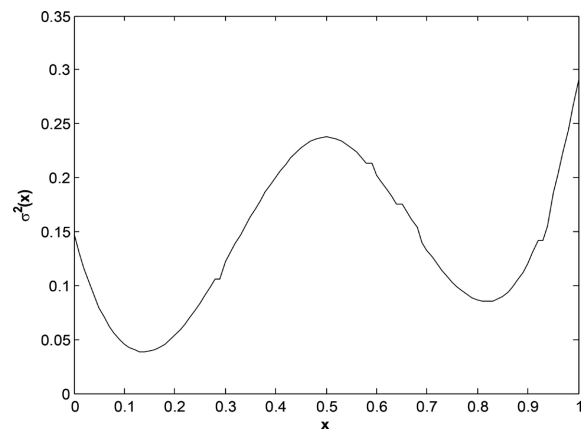


Figure 3. The estimated variance function $\hat{\sigma}^2(x)$

图 3. 估计的方差函数 $\hat{\sigma}^2(x)$

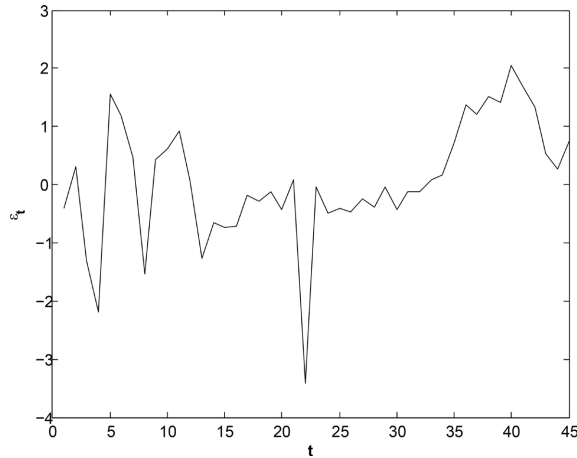


Figure 4. The estimated error sequence $\{\hat{\varepsilon}_t\}$

图 4. 估计的误差序列 $\{\hat{\varepsilon}_t\}$

经计算，误差序列 $\{\hat{\varepsilon}_t\}$ 的样本均值为 0.0619 和样本方差为 1.1021，可近似地看作 $\{\hat{\varepsilon}_t\}$ 是均值为 0 和方差为 1 的序列。

为检验 $\{\hat{\varepsilon}_t\}$ 的平稳性，表 5 给出了它的 ADF 检验结果。由表 5 可知，ADF 检验的 P 值 $0.0370 < 0.05$ ，在显著性水平 0.05 下，拒绝存在单位根的原假设，说明 $\{\hat{\varepsilon}_t\}$ 为平稳序列。进一步地，图 5 给出了 $\{\hat{\varepsilon}_t\}$ 的自相关系数和偏自相关系数，可以看出延迟各期 Q_{LB} 统计量的 P 值均小于显著性水平 0.05，拒绝白噪声的原假设，说明 $\{\hat{\varepsilon}_t\}$ 是非白噪声序列。此外，图 6 给出了 $\{\hat{\varepsilon}_t\}$ 的 Q-Q 图，图中的点分布在一条直线附近，同时 KS-检验的 P 值 $0.085 > 0.05$ ，接受 $\{\hat{\varepsilon}_t\}$ 服从正态分布的原假设，近似地有 $\hat{\varepsilon}_t \sim N(0,1)$ 。

由图 5 还可以看出， $\{\hat{\varepsilon}_t\}$ 的自相关系数拖尾、偏自相关系数 2 阶截尾，可对 $\{\hat{\varepsilon}_t\}$ 建立 AR(2)模型，但在实际拟合中，AR(1)模型的信息准则较优，建立的 AR(1)模型为

$$\tilde{\varepsilon}_t = 0.40236\tilde{\varepsilon}_{t-1} \tag{14}$$

图 7 给出了模型(14)的残差序列 $\{\tilde{\varepsilon}_t\}$ 的检验结果，其中 $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \tilde{\varepsilon}_t$ 。由图 7 可见延迟各期 Q_{LB} 统计量的 P 值均大于显著性水平 0.05，可以认为 $\{\tilde{\varepsilon}_t\}$ 是白噪声序列。进一步地，表 6 给出了模型(14)的参数估计及其检验结果，显然，参数检验的 P 值 $0.0029 < 0.05$ ，可以认为模型参数显著，所建立的模型(14)是合适的。因而，序列 $\{Y_t\}$ 满足建立非参数方差模型(1)的条件。

图 8 给出了序列 $\{Y_t\}$ 的逐点估计置信区间，其中虚线和点虚线分别代表置信区间的上界和下界。估计的平均区间覆盖率 MICP 为 0.9333，平均区间宽度 AIW 为 1.4222。表 7 给出了预测未来 5 年 Y_t 的置信区间和河南省第三产业贡献率 Y_{0t} 的置信区间，其中

$$Y_{0t} = e^{(Y_t + 3.5985)} \tag{15}$$

Table 5. The ADF test results of $\{\hat{\varepsilon}_t\}$

表 5. $\{\hat{\varepsilon}_t\}$ 的 ADF 检验结果

ADF 检验统计量		t 统计量	P 值
		-3.6857	0.0370
检验的临界值	1%显著性水平	-4.15	
	5%显著性水平	-3.50	
	10%显著性水平	-3.18	

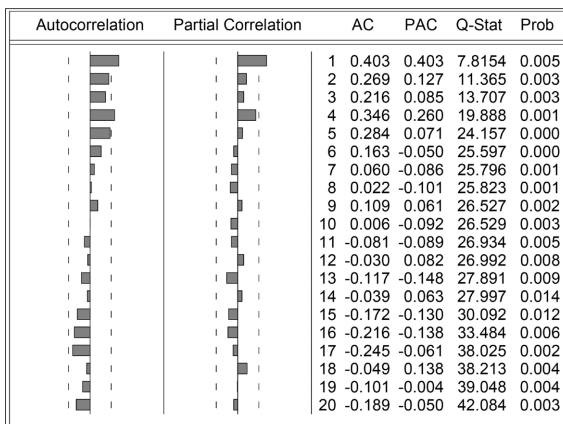


Figure 5. Autocorrelation coefficient and partial autocorrelation coefficient of $\{\hat{\varepsilon}_t\}$

图 5. $\{\hat{\varepsilon}_t\}$ 的自相关系数和偏自相关系数

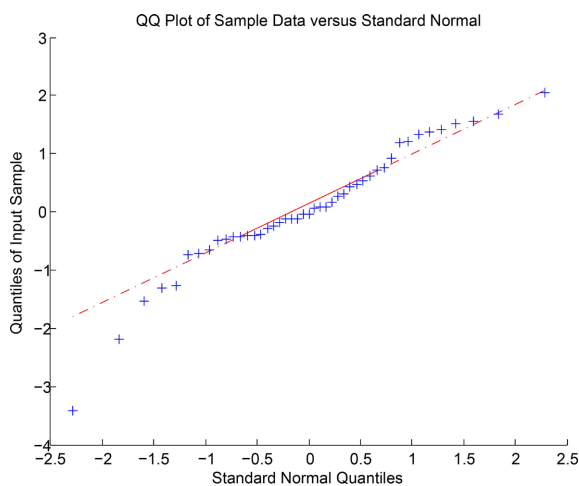


Figure 6. Q-Q plot of $\{\hat{\varepsilon}_t\}$

图 6. $\{\hat{\varepsilon}_t\}$ 的 Q-Q 图

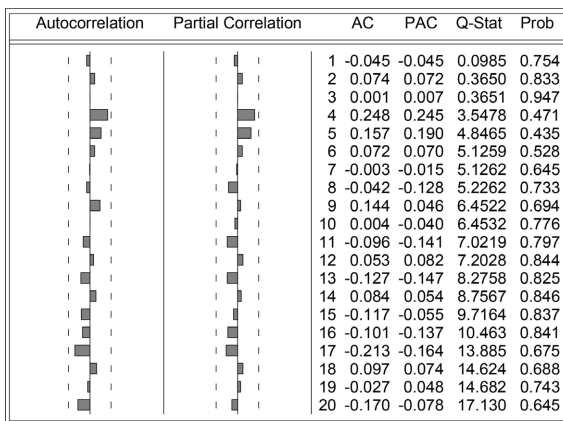
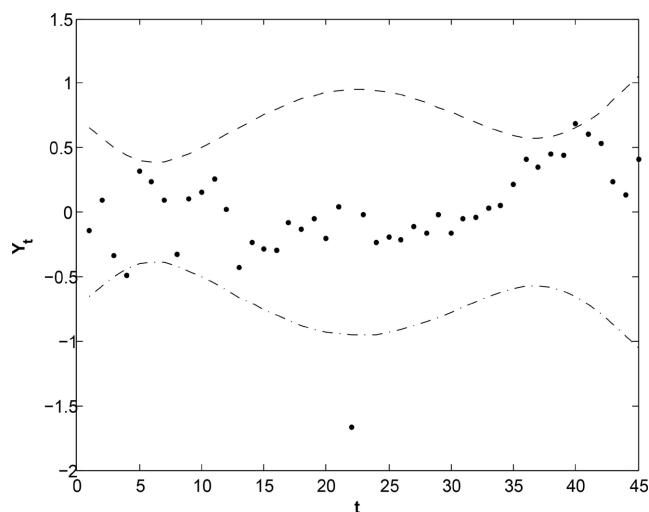


Figure 7. Autocorrelation coefficient and partial autocorrelation coefficient of $\{\tilde{\varepsilon}_t\}$

图 7. $\{\tilde{\varepsilon}_t\}$ 的自相关系数和偏自相关系数

Table 6. Parameter estimate in model (14) and its test results**表 6.** 模型(14)的参数估计及检验结果

变量	系数	标准误差	t 统计量	P 值
AR(1)	0.40236	0.1350	2.9795	0.0029

**Figure 8.** $\{Y_t\}$ and its pointwise confidence intervals**图 8.** $\{Y_t\}$ 及其逐点置信区间**Table 7.** Prediction intervals of the contribution rates of Henan's tertiary industry from 2026 to 2030**表 7.** 2026~2030 年河南省第三产业贡献率的预测区间

年份	t	Y_t 的预测区间	Y_{0r} 的预测区间及区间中点值(括号内)
2026	46	[-1.1503, 1.1503]	[11.5677, 115.4438] (63.5058)
2027	47	[-1.2353, 1.2353]	[10.6249, 125.6872] (68.1561)
2028	48	[-1.3148, 1.3148]	[9.8127, 136.0911] (72.9519)
2029	49	[-1.3898, 1.3898]	[9.1038, 146.6881] (77.8959)
2030	50	[-1.4609, 1.4609]	[8.4786, 157.5037] (82.9912)

5. 结束语

本文基于多项式样条估计和线性外推法，构建了固定设计非参数方差模型的预测区间，模拟算例表明了预测区间的可行性，并得到了“十五五”时期河南省第三产业贡献率的预测区间。

基金项目

河南省高校人文社会科学研究项目资助“基于包容性增长的黄河流域城市群高质量发展水平提升研究”(2024-ZZJH-281)。

参考文献

- [1] Dahl, C.M. and Levine, M. (2006) Nonparametric Estimation of Volatility Models with Serially Dependent Innovations. *Statistics & Probability Letters*, **76**, 2007-2016. <https://doi.org/10.1016/j.spl.2006.05.018>

-
- [2] 武新乾, 冯爱芬, 田萍. 相依数据下非参数方差模型的样条估计[J]. 河南科技大学学报(自然科学版), 2009, 30(5): 93-96+104+114.
- [3] 武新乾, 梅倩倩. 鞅误差下非参数方差模型的样条估计[J]. 河南科技大学学报(自然科学版), 2013, 34(2): 91-94+1.
- [4] Cheng, T.T., Gao, J., Linton, O., *et al.* (2025) Nonparametric Predictive Regression for Stock Return Prediction. *Econometric Reviews*, **44**, 1462-1493. <https://doi.org/10.1080/07474938.2025.2519389>
- [5] Li, J.J. and Zhao, L.F. (2020) Hydropower Price Prediction with the Nonparametric Statistics Regression Model. *Journal of Coastal Research*, **104**, 402-405. <https://doi.org/10.2112/jcr-si104-072.1>
- [6] 李威. 基于非参数组合预测模型的江西省 GDP 的实证分析[D]: [硕士学位论文]. 赣州: 江西理工大学, 2022.
- [7] Xiao, Z., Sun, J., Zhang, Z.Y., *et al.* (2024) Nonparametric Regression for MU-MIMO Channel Prediction: From KNN to Local Linear Regression. *IEEE Transactions on Wireless Communications*, **23**, 2784-2795. <https://doi.org/10.1109/twc.2023.3303024>
- [8] 武新乾, 张刚. 具有非参数趋势的残差自回归模型的预测方法[J]. 统计与决策, 2016(15): 84-86.
- [9] 武新乾, 程芳. 固定设计异方差非参数回归模型的预测方法[J]. 统计与决策, 2017(17): 76-79.
- [10] 王彩霞, 鲁宗相, 乔颖, 等. 基于非参数回归模型的短期风电功率预测[J]. 电力系统自动化, 2010, 34(16): 78-82+91.
- [11] 李晋云, 武新乾. 基于样条方法的固定设计非参数回归模型的区间预测[J]. 统计与决策, 2021, 37(4): 46-50.