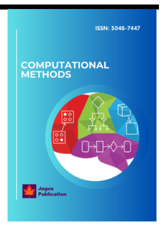
Computational Methods, 2(2):7–12, 2025. DOI: 10.58614/cm222

Computational Methods

Journal homepage: https://www.jpub.org/journal-details.php?journal-id=43



Research Article

Nonlinear Nonparametric Uncertain Autoregressive Time Series Model

Mengqin Tian[1] **and Yuanguo Zhu**[1*]

[1]*School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China.*
[*]*Corresponding author: ygzhu@njust.edu.cn*

Article Info

Keywords:
Uncertain Theory, Time Series Analysis, Legendre Polynomial Approximation.

MSC 2020: *60B99*

Received: *10.11.2025;*
Accepted: *01.12.2025;*
Published: *10.12.2025*

Abstract

Uncertain time series analysis provides a framework for modeling data shaped by human belief and cognitive limitations rather than randomness. However, existing models are largely parametric and assume uncertain normal residuals, which often fail in practical applications involving nonlinear and non-normal dynamics. To address these limitations, we propose a nonlinear nonparametric uncertain autoregressive (NNUAR) model based on multidimensional Legendre polynomial approximation. This approach leverages tensor-product Legendre polynomials to nonparametrically capture nonlinear relationships, with parameters estimated via least squares. A two-stage framework is developed to address residual autocorrelation and departures from uncertain normality, incorporating an uncertain hypothesis test and cross-validation for optimal lag selection. Numerical experiments on two years of weekly closing prices of Ping An Bank show that the NNUAR model effectively captures complex nonlinear dependencies and significantly reduces residual correlation.

1. Introduction

Time series analysis has been widely applied in diverse fields such as infectious disease control and financial market analysis. Conventional approaches rely on probability theory, treating observations as random variables governed by probability distributions. However, in many real-world settings, uncertainty often arises not from stochastic randomness but from incomplete information, cognitive limitations, subjective judgment, or unforeseen events, which do not satisfy frequentist assumptions.

To address such non-random uncertainty, Liu introduced uncertainty theory in 2009 [1], which models belief degrees through uncertain measures rather than empirical frequencies. This framework has since been extensively developed and applied. Building on it, Yang and Liu pioneered uncertain time series modeling [2], prompting numerous extensions. For example, Yang and Ni proposed an uncertain moving average model via least squares estimation [3]; Xin et al. developed an uncertain autoregressive moving average model using maximum likelihood estimation and applied it to financial markets [4]; Li and Wang employed an uncertain AR(1) model for urban water demand forecasting [5]. These studies, however, focus exclusively on linear structures. More recently, Xie and Lio introduced the first nonlinear uncertain time series model [6], and Zhang and Gao proposed a nonparametric spline-based additive autoregressive model for oil price forecasting [7].

Despite substantial progress in nonparametric and nonlinear modeling for classical time series [8–11], analogous developments remain absent in uncertain time series analysis. Furthermore, all existing uncertain time series models assume that residuals follow an uncertain normal distribution, an assumption frequently violated in practice. Critically, no existing method addresses how to proceed when this assumption does not hold.

To bridge these gaps, we propose a two-stage **N**onlinear **N**onparametric **U**ncertain **A**utoreg **R**essive (NNUAR) modeling framework based on multidimensional Legendre polynomial approximation, drawing inspiration from recent Legendre-based approaches to uncertain regression and differential equations [12–14]. Using tensor products, we extend one-dimensional Legendre polynomials to the multidimensional case to model the primary time series component. The lag order and polynomial truncation degree are selected via the cross-validation method of Liu and Yang [15]. Following this, an uncertain normality hypothesis test is applied to the first-stage residuals. If the test indicates a departure from uncertain normality, a second-stage model based on Legendre polynomials is fitted to the residuals to effectively remove autocorrelation and temporal dependencies.

We apply the NNUAR model to two years of weekly closing prices of Ping An Bank. Hypothesis tests reject uncertain normality of the first-stage residuals, while the second-stage residuals conform to uncertain normality, confirming the model's adequacy.

The remainder of this paper is structured as follows. Section 2 presents the mathematical formulation of the NNUAR model and constructs multidimensional Legendre polynomials via tensor products. Section 3 details the full modeling framework, including nonparametric Legendre approximation, first and second order residual modeling, uncertain hypothesis testing, selection of lag and truncation orders, and point and interval forecasting. Section 4 validates the approach through numerical experiments on Ping An Bank data. Section 5 concludes with a summary of contributions.

2. Preliminaries

This section defines a generalized nonlinear nonparametric uncertain autoregressive time series and outlines the construction of multidimensional Legendre polynomials via tensor products, along with relevant notation.

Definition 2.1. For a given lag order p, the nonlinear nonparametric uncertain autoregressive time series model is defined by the following form:

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-p}) + \varepsilon_t, \tag{1}$$

where $f : \mathbb{R}^p \to \mathbb{R}$ is an unknown (typically nonlinear) function and ε_t are disturbance terms following an uncertainty distribution.

Legendre polynomials $l_n(x)$ satisfy the following differential equation:

$$\frac{d}{dx}\left((1-x^2)\frac{d}{dx}l_n(x)\right) + n(n+1)l_n(x) = 0, \quad n = 0,1,2,\ldots, \tag{2}$$

and can be efficiently computed via the recurrence:

$$(n+1)l_{n+1}(x) = (2n+1)xl_n(x) - nl_{n-1}(x), \quad n \geq 1, \tag{3}$$

with $l_0(x) = 1$, $l_1(x) = x$. Legendre polynomials are orthogonal on $[-1,1]$:

$$\int_{-1}^{1} l_m(x)l_n(x)\,dx = \frac{2}{2n+1}\delta_{mn}, \tag{4}$$

where δ_{mn} denotes the Kronecker delta. The orthogonality property making them suitable for L^2-optimal function approximation. Following [14], multivariate Legendre polynomials are constructed via tensor products to fit high-order time series models:

$$l_\alpha(x) = \prod_{i=1}^{p} l_{\alpha_i}(x_i), \tag{5}$$

where $x \in [-1,1]^p$, $\alpha = (\alpha_1, \ldots, \alpha_p)$ is a multi-index, and l_{α_i} is the univariate Legendre polynomial of degree α_i. For total degree at most d, the number of basis functions is

$$K = \binom{p+d}{d}, \tag{6}$$

which grows rapidly with p and d, necessitating small truncation orders for computational feasibility.

To define the design matrix in Section 3, we map multi-indices α to a one-dimensional index via

$$\text{index}(\alpha) = \sum_{k=0}^{|\alpha|-1} \binom{p+k-1}{k} + \text{lex\_rank}(\alpha \mid |\alpha|), \tag{7}$$

where

$$\text{lex\_rank}(\alpha \mid m) = \sum_{i=1}^{p-1} \sum_{j=0}^{\alpha_i - 1} \binom{(m - s_{i-1} - j) + (p-i-1)}{p-i-1}, \tag{8}$$

with $s_0 = 0$ and $s_i = \sum_{k=1}^{i} \alpha_k$. The first term counts all basis functions of degree less than $|\alpha|$; the second gives the lexicographic rank among indices of equal total degree $m = |\alpha|$. This bijection ensures a deterministic enumeration of the polynomial basis.

3. Nonlinear Nonparametric Uncertain Autoregressive Time Series Model

In this section, we present a nonparametric approach for modeling nonlinear uncertain autoregressive time series using Legendre polynomial approximations. Section 3.1 constructs an estimator for the unknown regression function via multivariate Legendre bases and estimates coefficients using least squares. Section 3.2 analyzes residuals and proposes a second-stage model when residuals deviate from an uncertain normal distribution. Section 3.3 introduces the uncertain hypothesis test to validate the residual model. Section 3.4 describes a cross-validation procedure to select optimal lag orders and polynomial truncation degrees. Finally, Section 3.5 derives the predictive form of the NNUAR model and its associated confidence intervals.

3.1. Nonparametric Polynomial Approximation

We model a time series $\{x_t\}_{t=1}^T$ as $X_t = f(X_{t-1},\dots,X_{t-p}) + \varepsilon_t$, where f is an unknown nonlinear function and ε_t is an uncertain disturbance. To estimate f nonparametrically, we approximate it by a finite linear combination of multivariate Legendre polynomials:

$$f(X_{t-1},\dots,X_{t-p}) \approx \sum_{k=1}^{K} \theta_k \phi_k(X_{t-1},\dots,X_{t-p}),$$

where ϕ_k are product Legendre bases defined in Section 2. Since Legendre polynomials are orthogonal on $[-1,1]$, we normalize the data via global min-max scaling:

$$\tilde{x}_t = 2\cdot\frac{x_t - x_{\min}}{x_{\max} - x_{\min}} - 1.$$

Let $\mathbf{x}_t = (x_{t-1},\dots,x_{t-p})^\top$ and $\phi(\tilde{\mathbf{x}}_t) = [\phi_1(\tilde{\mathbf{x}}_t),\dots,\phi_K(\tilde{\mathbf{x}}_t)]^\top$. The least squares estimate of $\theta = [\theta_1,\dots,\theta_K]^\top$ minimizes $\|\mathbf{y} - \Phi\theta\|_2^2$, where

$$\Phi = \begin{bmatrix} \phi(\tilde{\mathbf{x}}_{p+1})^\top \\ \vdots \\ \phi(\tilde{\mathbf{x}}_T)^\top \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \tilde{x}_{p+1} \\ \vdots \\ \tilde{x}_T \end{bmatrix}.$$

Assuming $\Phi^\top\Phi$ is invertible, the solution is $\hat{\theta} = (\Phi^\top\Phi)^{-1}\Phi^\top\mathbf{y}$, yielding the estimated function $\hat{f}(\mathbf{x}_t) = \phi(\mathbf{x}_t)^\top\hat{\theta}$.

3.2. Residual Analysis and Second-Stage Modeling

The residuals $\hat{\varepsilon}_t = x_t - \hat{f}(\mathbf{x}_t)$ for $t = p+1,\dots,T$ often exhibit autocorrelation and departures from normality, while previous work does not account for these issues. To address this, we model the residuals as $\varepsilon_t = g(\hat{\varepsilon}_{t-1},\dots,\hat{\varepsilon}_{t-q}) + \xi_t$, where g is approximated using multivariate Legendre polynomials up to degree r:

$$g(\hat{\varepsilon}_{t-1},\dots,\hat{\varepsilon}_{t-q}) \approx \sum_{m=1}^{M} \eta_m \psi_m(\hat{\varepsilon}_{t-1},\dots,\hat{\varepsilon}_{t-q}).$$

Input residuals are normalized to $[-1,1]$. Let $\hat{\varepsilon}_t = (\hat{\varepsilon}_{t-1},\dots,\hat{\varepsilon}_{t-q})^\top$, and construct design matrix $\Psi \in \mathbb{R}^{(T-p-q)\times M}$ and response vector $\hat{z} \in \mathbb{R}^{T-p-q}$. The coefficient estimate is $\hat{\eta} = (\Psi^\top\Psi)^{-1}\Psi^\top\hat{z}$, giving the refined fit:

$$\hat{x}_t^{(\text{final})} = \hat{f}(\mathbf{x}_t) + \hat{g}(\hat{\varepsilon}_t) + \xi_t.$$

3.3. Uncertain Hypothesis Test

We test whether residuals $\hat{\varepsilon}_t$ follow an uncertain normal distribution $\mathcal{N}(\hat{e},\hat{\sigma})$ using the method in [16], with

$$\hat{e} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{\varepsilon}_t, \quad \hat{\sigma}^2 = \frac{1}{T-p}\sum_{t=p+1}^{T}(\hat{\varepsilon}_t - \hat{e})^2.$$

The hypotheses are:

$$H_0 : \varepsilon_t \sim \mathcal{N}(\hat{e},\hat{\sigma}), \quad H_1 : \varepsilon_t \nsim \mathcal{N}(\hat{e},\hat{\sigma}).$$

At significance level α, the rejection region is

$$W = \left\{(\hat{\varepsilon}_{p+1},\dots,\hat{\varepsilon}_T) : \text{more than } \alpha(T-p) \text{ residuals lie outside } [\Phi^{-1}(\alpha/2), \Phi^{-1}(1-\alpha/2)]\right\},$$

where $\Phi^{-1}(\alpha) = \hat{e} + \frac{\hat{\sigma}\sqrt{3}}{\pi}\ln\left(\frac{\alpha}{1-\alpha}\right)$. If H_0 is rejected, the second-stage model in Section 3.2 is applied; otherwise, the model is deemed adequate.

3.4. Cross-Validation

We use fixed-origin cross-validation proposed by [15] to select optimal lag orders p,q and truncation degrees d,r. The series is split into training (1 to T) and testing ($T+1$ to n) sets. For each candidate (p,d), the average testing error is

$$\text{ATE}_T(p,d) = \frac{1}{n-T}\sum_{t=T+1}^{n}\left(x_t - \hat{f}^{(p,d)}(\mathbf{x}_t)\right)^2.$$

The optimal (p^*,d^*) minimizes $\text{ATE}_T(p,d)$. The same procedure selects (q^*,r^*) for the residual model.

3.5. Forecast Values and Confidence Interval

The final NNUAR model is $X_t = \hat{f}(\mathbf{x}_t) + \hat{g}(\hat{\varepsilon}_t) + \xi_t$, with $\xi_t \sim \mathcal{N}(\hat{e}, \hat{\sigma})$. The forecast for $T+1$ is

$$\hat{X}_{T+1} = \hat{f}(\mathbf{x}_{T+1}) + \hat{g}(\hat{\varepsilon}_{T+1}) + \xi_{T+1},$$

with point forecast $\mathbb{E}[\hat{X}_{T+1}] = \hat{f}(\mathbf{x}_{T+1}) + \hat{g}(\hat{\varepsilon}_{T+1}) + \hat{e}$. Since $\hat{X}_{T+1} \sim \mathcal{N}(\hat{\mu}, \hat{\sigma})$ with $\hat{\mu}$ as above, the α-confidence interval is

$$\left[\hat{\mu} - \frac{\hat{\sigma}\sqrt{3}}{\pi} \ln\left(\frac{1+\alpha}{1-\alpha}\right), \quad \hat{\mu} + \frac{\hat{\sigma}\sqrt{3}}{\pi} \ln\left(\frac{1+\alpha}{1-\alpha}\right)\right].$$

4. Numerical Experiments

We apply the proposed NNUAR model to weekly closing prices of Ping An Bank from October 2023 to October 2025, sourced from the East Money Choice Financial Terminal. Figure 1 displays the time series.

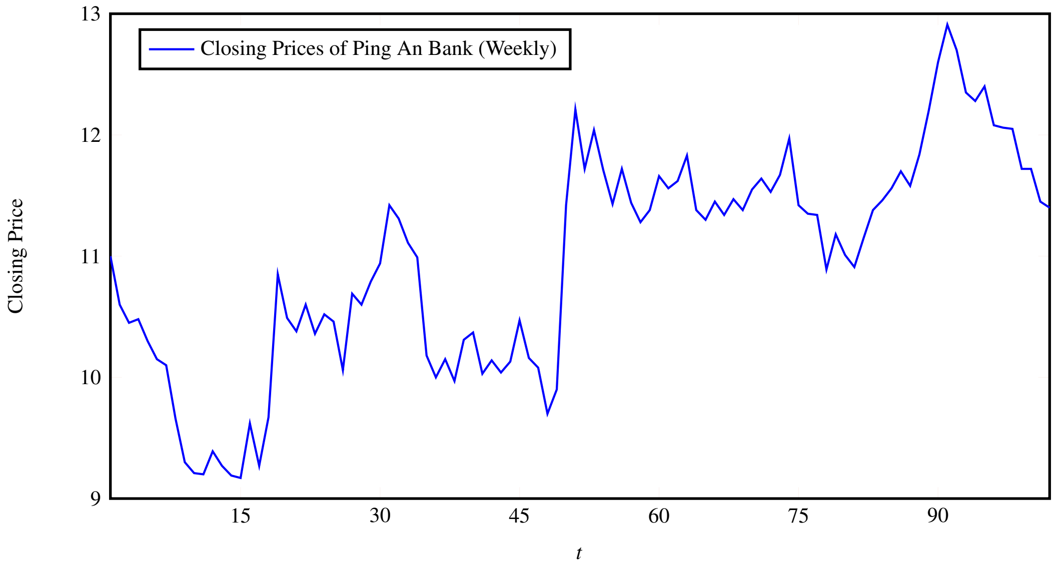


Figure 1: Closing Prices of Ping An Bank (Weekly)

4.1. NNUAR Modeling Results

Following the methodology in Section 3, the first-stage NNUAR model selects lag order 4 and truncation order 5, yielding 126 Legendre basis functions:

$$X_t = \sum_{k=1}^{126} \theta_k \phi_k(X_{t-1}, \ldots, X_{t-4}) + \varepsilon_t. \tag{9}$$

The second-stage model, applied to residuals $\hat{\varepsilon}_t$, uses lag order 3 and truncation order 3, producing 20 basis functions:

$$X_t = \sum_{k=1}^{126} \theta_k \phi_k(X_{t-1}, \ldots, X_{t-4}) + \sum_{m=1}^{20} \eta_m \psi_m(\hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-3}) + \xi_t. \tag{10}$$

Coefficient values are omitted due to their large number.

4.2. Residuals Analysis

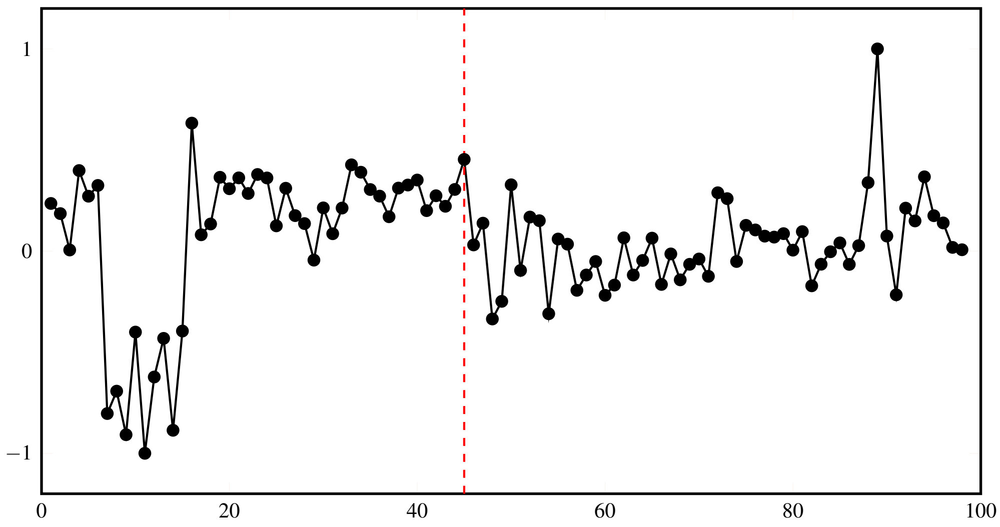


Figure 2: First-stage Residuals Plot in Closing Prices Modeling

We select the 45th residual as a splitting point, dividing the residuals into two segments. As shown in Figure 2, the two segments exhibit markedly different behaviors. To further substantiate this observation, we conduct a two-sample Kolmogorov–Smirnov (KS) test to assess the homogeneity of the two residual segments. The test yields a KS statistic of 0.4935 and a p-value of 6.16×10^{-6}, which is far below the 0.05 significance level. This provides strong statistical evidence that the left and right residual segments are drawn from distinct underlying distributions, thereby necessitating a time series modeling framework based on uncertainty theory.

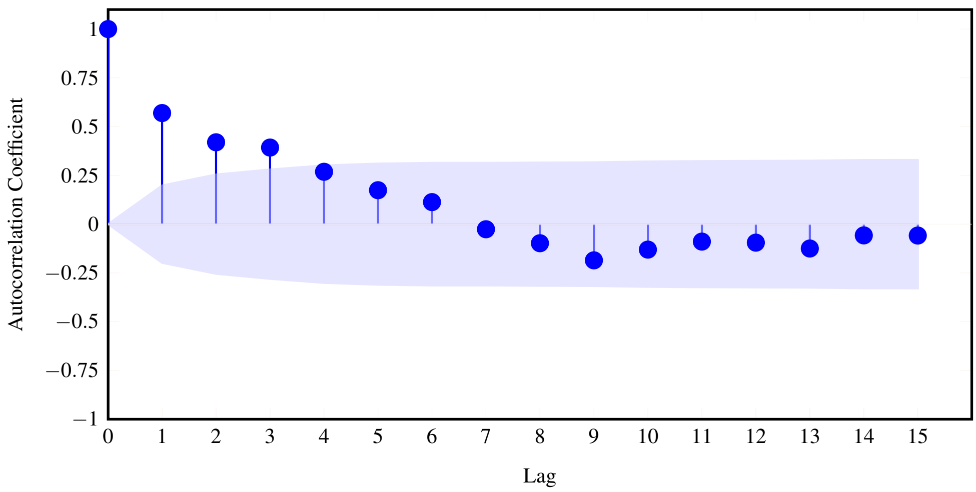


Figure 3: ACF Plot of First-stage Residuals

As shown in Figure 3, the autocorrelations of the first-stage residuals at lags up to and including 3 all lie outside the 95% confidence interval, indicating significant autocorrelation. This suggests that the residuals do not follow an uncertain normal distribution, as is commonly assumed in other approaches. Moreover, in our experiments, cross-validation selects a lag order of 3 as optimal for the second-stage model, which is consistent with the ACF analysis.

4.3. Uncertain Hypothesis Test

We test whether residuals follow an uncertain normal distribution. For first-stage residuals ε_t, with estimated mean 0.0519 and variance 0.3173, the null hypothesis $H_0 : \varepsilon_t \sim \mathcal{N}(0.0519, 0.3173)$ is rejected due to 7 outliers exceeding the critical value of 4.9. For second-stage residuals ξ_t, with mean -1.8523×10^{-15} and variance 0.2191, only 3 outliers are detected, below the critical value of 4.75, so H_0 is not rejected. This confirms that the two-stage NNUAR model adequately captures the residual structure.

Finally, the NNUAR model yields a point forecast of 11.3493, with a 95% confidence interval of [11.3372, 11.3614], which is very close to the true value of 11.3400.

5. Conclusion

In this paper, we propose a nonlinear nonparametric uncertain autoregressive (NNUAR) model based on multidimensional Legendre polynomial approximation, which avoids pre-specified functional forms and distributional assumptions. A second-order modeling strategy is introduced to handle non-normal uncertain residuals, enhancing robustness. Experiments on the weekly closing prices of Ping An Bank over the past two years show that the NNUAR model effectively captures complex nonlinear dynamics.

Article Information

Disclaimer (Artificial Intelligence): The author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.), and text-to-image generators have been used during writing or editing of manuscripts.

Competing Interests: Authors have declared that no competing interests exist.

References

[1] Baoding Liu. Some research problems in uncertainty theory. *Journal of Uncertain systems*, 3(1):3–10, 2009.

[2] Xiangfeng Yang and Baoding Liu. Uncertain time series analysis with imprecise observations. *Fuzzy Optimization and Decision Making*, 18(3):263–278, 2019.

[3] Xiangfeng Yang and Yaodong Ni. Least-squares estimation for uncertain moving average model. *Communications in Statistics-Theory and Methods*, 50(17):4134–4143, 2021.

[4] Yue Xin, Jinwu Gao, Xiangfeng Yang, and Jing Yang. Maximum likelihood estimation for uncertain autoregressive moving average model with application in financial market. *Journal of Computational and Applied Mathematics*, 417:114604, 2023.

[5] Wei Li and Xiaosheng Wang. Analysis and prediction of urban household water demand with uncertain time series. *Soft Computing-A Fusion of Foundations, Methodologies & Applications*, 28, 2024.

[6] Jinsheng Xie and Waichon Lio. Uncertain nonlinear time series analysis with applications to motion analysis and epidemic spreading. *Fuzzy Optimization and Decision Making*, 23(2):279–294, 2024.

[7] Yi Zhang and Jinwu Gao. Nonparametric uncertain time series models: Theory and application in brent crude oil spot price analysis. *Fuzzy Optimization and Decision Making*, 23(2):239–252, 2024.

[8] Jianqing Fan and Qiwei Yao. *Nonlinear time series: nonparametric and parametric methods*. Springer, 2003.

[9] Elias Masry. Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599, 1996.

[10] Felipe Tobar, Thang D Bui, and Richard E Turner. Learning stationary time series using gaussian processes with nonparametric kernels. *Advances in neural information processing systems*, 28, 2015.

[11] Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. Deep state space models for time series forecasting. *Advances in neural information processing systems*, 31, 2018.

[12] Yue Xin and Jinwu Gao. Orthogonal series method for uncertain nonparametric regression with application to carbon dioxide emissions. *Communications in Statistics-Simulation and Computation*, 53(10):5018–5027, 2024.

[13] Liu He and Yuanguo Zhu. Nonparametric estimation for uncertain fractional differential equations. *Chaos, Solitons & Fractals*, 178: 114342, 2024.

[14] Yue Feng, Yuanguo Zhu, and Liu He. Nonparametric estimation of nonautonomous uncertain differential equations with application to temperature models. *Journal of Intelligent & Fuzzy Systems*, 46(3):7273–7281, 2024.

[15] Zhe Liu and Xiangfeng Yang. Cross validation for uncertain autoregressive model. *Communications in Statistics-Simulation and Computation*, 51(8):4715–4726, 2022.

[16] Tingqing Ye and Baoding Liu. Uncertain hypothesis test with application to uncertain regression analysis. *Fuzzy Optimization and Decision Making*, 21(2):157–174, 2022.