axioms

Article

Partial Derivatives Estimation of Multivariate Variance Function in Heteroscedastic Model via Wavelet Method

Junke Kou * and Hao Zhang

School of Mathematics and Computational Science, Guilin University of Electronic Technology, Guilin 541004, China; zhanghao19990207@163.com
* Correspondence: kjkou@guet.edu.cn

Abstract: For derivative function estimation, conventional research only focuses on the derivative estimation of one-dimensional functions. This paper considers partial derivatives estimation of a multivariate variance function in a heteroscedastic model. A wavelet estimator of partial derivatives of a multivariate variance function is proposed. The convergence rates of a wavelet estimator under different estimation errors are discussed. It turns out that the strong convergence rate of the wavelet estimator is the same as the optimal uniform almost sure convergence rate of nonparametric function problems.

Keywords: nonparametric estimation; partial derivative; strong convergence rate; heteroscedastic model

MSC: 62G07; 62G20; 42C40

1. Introduction

This paper considers the following heteroscedastic model:

$$Y_i = g(\boldsymbol{X_i}) + f(\boldsymbol{X_i})U_i, i \in \{1, \cdots, n\}. \tag{1}$$

In this model, $(\boldsymbol{X_1}, Y_1), \cdots, (\boldsymbol{X_n}, Y_n)$ are independent and identically distributed random vectors. The functions $g(\boldsymbol{x})$ and $f(\boldsymbol{x})$ all are defined on $[a, b]^d$. $U_1, \ldots, U_n$ are identically distributed random variables, which satisfy $\mathbb{E}[U_i] = 0$ and $Var[U_i] = 1$. Furthermore, the random vector $\boldsymbol{X_i}$ and random variable U_i are uncorrelated for any $i \in \{1, \cdots, n\}$. This paper is devoted to estimating the partial derivatives $(\partial^{\boldsymbol{m}} r)(\boldsymbol{x})$ of the variance function $r(\boldsymbol{x})(r(\boldsymbol{x}) := f^2(\boldsymbol{x}))$ from the observed data $(\boldsymbol{X_1}, Y_1), \cdots, (\boldsymbol{X_n}, Y_n)$. The partial derivative $(\partial^{\boldsymbol{m}} r)(\boldsymbol{x})$ is defined by

$$(\partial^{\boldsymbol{m}} r)(\boldsymbol{x}) = \left(\frac{\partial^{\boldsymbol{m}} r}{\partial x_1^{m_1} \cdots \partial x_d^{m_d}} \right)(x_1, \cdots, x_d),$$

with $\boldsymbol{m} = (m_1, \cdots, m_d) \in \mathbb{N}^d$, and $|\boldsymbol{m}| = \sum_{i=1}^{d} m_i$.

In practical applications, the heteroscedastic model is widely used for monitoring the signal-to-noise ratios in quality control [1,2], measuring the reliability of time series prediction [3], evaluating the volatility or risk in financial markets [4] and so on. Hence, many significant results have been obtained by [5–8] and others. For this heteroscedastic model, Fan and Yao [9] propose a residual-based estimator of the variance function, and study the asymptotic normality properties of an estimator. A class of difference-based kernel estimators of a variance function are constructed by [10]. Moreover, the asymptotic rates of convergence and the optimal bandwidth of kernel estimators are discussed. Wang et al. [11] consider the minimax convergence rates of a kernel estimator over pointwise squared

check for updates

Citation: Kou, J.; Zhang, H. Partial Derivatives Estimation of Multivariate Variance Function in Heteroscedastic Model via Wavelet Method. *Axioms* **2024**, *13*, 69. https://doi.org/10.3390/axioms13010069

Academic Editors: Jiujun Zhang and David Katz

Received: 28 November 2023
Revised: 14 January 2024
Accepted: 18 January 2024
Published: 20 January 2024

error and global integrated mean squared error, respectively. The optimal estimation of a variance function with random design is discussed by [12]. Zaoui [13] studies the variance function estimation with model selection aggregation and convex aggregation.

Derivative estimation plays a crucial role in nonparametric statistics estimation, big data processing, and other practical applications. For example, some companies predict the profit growth rate when the research and development investment increases. Many financial and fund institutions evaluate the volatility of stock market prices and so on. Many important and interesting results of derivative estimation are obtained using different methods. Zhou and Wolfe [14] propose a spline derivative estimator, and discuss the asymptotic normality and variance property. A spatially adaptive derivative estimator is constructed by [15]. Chaubey et al. [16] consider the upper bound over L^p-loss for wavelet estimators of density derivative functions. A convergence rate over the mean integrated error of a derivative estimator for mixing sequences is proved by [17]. For the estimation problem (1), the derivatives estimation of the variance function via the wavelet method is proved by [18]. However, it should be pointed out that those above results all focus on the estimation of the derivatives of a one-dimensional function. There is a lack of partial derivatives estimation of a multivariate variance function. Hence, in this paper, we construct a partial derivatives estimator using the wavelet method, and discuss the convergence rates of the wavelet estimator under different mild conditions.

The structure of this paper is given as follows. Section 2 specifies some mild hypotheses for the estimation model (1), and constructs a wavelet estimator of the partial derivative function. Two important auxiliary results of the wavelet estimator are proved in Section 3. The estimation errors of the wavelet estimator under different assumptions are discussed in Section 4.

2. Wavelet Estimator

In this section, we will give some hypotheses of the estimation problem (1), and construct a partial derivatives estimator using the wavelet method. For the estimation model (1), the following mild assumptions are proposed, which are used in the later discussion.

A1: For the partial derivative $(\partial^{\boldsymbol{m}} r)(\boldsymbol{x})$ with $\boldsymbol{m} = (m_1, \cdots, m_d) \in \mathbb{N}^d$, if any $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d) \in \mathbb{N}^d$ and $\theta_i \leq m_i$, the partial derivative function satisfies $(\partial^{\boldsymbol{\theta}} r)(\boldsymbol{x}) = 0$ when $x_i \equiv a$ or b.

A2: The function $g(\boldsymbol{x})$ is known and bounded, i.e., there exists a positive constant c_1 such that $|g(\boldsymbol{x})| \leq c_1$.

A3: The density function $h(\boldsymbol{x})$ of the random vector $\boldsymbol{X}$ satisfies that $c_2 \leq h(\boldsymbol{x}) \leq c_3$, where $\boldsymbol{x} \in [a,b]^d$, c_2 and c_3 are two positive constants.

A4: The random variables $Y_1, Y_2, \ldots, Y_n$ are defined on $[c_4, c_5]$, where c_4 and c_5 are two constants.

In order to construct a wavelet estimator, some basic theories of wavelets are given in the following [19,20]. Let Φ be a scaling function and Ψ be a wavelet function such that

$$\left\{ \Phi_{\tau,\boldsymbol{k}},\ \Psi_{j,\boldsymbol{k},u}; j \geq \tau, u \in \{1, \cdots, 2^d - 1\},\ \boldsymbol{k} \in \mathbb{Z}^d \right\}$$

constitutes an orthonormal basis of $L^2(\mathbb{R}^d)$, where τ is a positive integer,

$$\Phi_{j,\boldsymbol{k}}(\boldsymbol{x}) = 2^{\frac{jd}{2}} \Phi(2^j x_1 - k_1, \cdots, 2^j x_d - k_d),$$
$$\Psi_{j,\boldsymbol{k},u}(x) = 2^{\frac{jd}{2}} \Psi_u(2^j x_1 - k_1, \cdots, 2^j x_d - k_d).$$

Then, for any integer j_0 such that $j_0 \geq \tau$, a function $f(\boldsymbol{x}) \in L^2([a,b]^d)$ can be expanded into a wavelet series as:

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \Lambda_{j_0}} \alpha_{j_0,\boldsymbol{k}} \Phi_{j_0,\boldsymbol{k}}(\boldsymbol{x}) + \sum_{j=j_0}^{\infty} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \beta_{j,\boldsymbol{k},u} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}), \quad (2)$$

where $\alpha_{j_0,\boldsymbol{k}} = \int_{[a,b]^d} f(\boldsymbol{x}) \cdot \Phi_{j_0,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x}$, $\beta_{j,\boldsymbol{k},u} = \int_{[a,b]^d} f(\boldsymbol{x}) \cdot \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})d\boldsymbol{x}$ and the cardinality of Λ_j satisfies $|\Lambda_j| \asymp 2^{jd}$. In this paper, we choose some compactly supported wavelets, such as the Daubechies wavelet [21]. In addition, this paper adopts the following symbol: $A \lesssim B$ denotes $A \leq cB$ for some constant $c > 0$; $A \gtrsim B$ means $B \lesssim A$; and $A \asymp B$ stand for both $A \lesssim B$ and $B \lesssim A$. For any $\boldsymbol{x} \in \mathbb{R}^d$, $\|\boldsymbol{x}\| := \sum_{i=1}^{d} |x_i|$.

Now we define a wavelet estimator of partial derivatives function $(\partial^{\boldsymbol{m}} r)(\boldsymbol{x})$ by

$$\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) := \sum_{\boldsymbol{k} \in \Lambda_{j_*}} \widehat{\alpha}_{j_*,\boldsymbol{k}} \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}). \tag{3}$$

In this definition,

$$\widehat{\alpha}_{j_*,\boldsymbol{k}} := \frac{(-1)^{|\boldsymbol{m}|}}{n} \sum_{i=1}^{n} \frac{Y_i^2}{h(\boldsymbol{X_i})} (\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_i}) - (-1)^{|\boldsymbol{m}|} \int_{[a,b]^d} g^2(\boldsymbol{x})(\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x})d\boldsymbol{x} \tag{4}$$

and

$$\begin{aligned}(\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x}) &= \left(\frac{\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}}}{\partial x_1^{m_1} \cdots \partial x_d^{m_d}} \right)(x_1, \cdots, x_d) \\ &= 2^{\frac{j_* d}{2}} \cdot 2^{j_* |\boldsymbol{m}|} (\partial^{\boldsymbol{m}} \Phi)(2^{j_*} x_1 - k_1, \cdots, 2^{j_*} x_d - k_d).\end{aligned}$$

3. Two Lemmas

This section will provide two important lemmas, which are used to prove the main theorem in a later section. According to the following lemma, it is easy to see that our wavelet estimator $\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})$ is unbiased.

Lemma 1. *For the model (1) with* ***A1****,*

$$\mathbb{E}[\widehat{\alpha}_{j_*,\boldsymbol{k}}] = \alpha_{j_*,\boldsymbol{k}}.$$

Proof. By the definition of $\widehat{\alpha}_{j_*,\boldsymbol{k}}$ and the properties of random vectors $(\boldsymbol{X_i}, Y_i)$,

$$\begin{aligned}\mathbb{E}[\widehat{\alpha}_{j_*,\boldsymbol{k}}] &= \mathbb{E}\left[\frac{(-1)^{|\boldsymbol{m}|}}{n} \sum_{i=1}^{n} \frac{Y_i^2}{h(\boldsymbol{X_i})} (\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_i}) - (-1)^{|\boldsymbol{m}|} \int_{[a,b]^d} g^2(\boldsymbol{x})(\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x})d\boldsymbol{x} \right] \\ &= \mathbb{E}\left[(-1)^{|\boldsymbol{m}|} \frac{Y_1^2}{h(\boldsymbol{X_1})} (\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_1}) \right] - (-1)^{|\boldsymbol{m}|} \int_{[a,b]^d} g^2(\boldsymbol{x})(\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x})d\boldsymbol{x}.\end{aligned}$$

Then, it follows from (1) that

$$\begin{aligned}&\mathbb{E}[\widehat{\alpha}_{j_*,\boldsymbol{k}}] \\ &= \mathbb{E}\left[(-1)^{|\boldsymbol{m}|} \frac{r(\boldsymbol{X_1})}{h(\boldsymbol{X_1})} U_1^2 (\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_1}) \right] + 2\mathbb{E}\left[(-1)^{|\boldsymbol{m}|} \frac{f(\boldsymbol{X_1}) g(\boldsymbol{X_1})}{h(\boldsymbol{X_1})} U_1 (\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_1}) \right] \\ &+ \mathbb{E}\left[(-1)^{|\boldsymbol{m}|} \frac{g^2(\boldsymbol{X_1})}{h(\boldsymbol{X_1})} (\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_1}) \right] - (-1)^{|\boldsymbol{m}|} \int_{[a,b]^d} g^2(\boldsymbol{x})(\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x})d\boldsymbol{x}.\end{aligned}$$

Note that the conditions $\mathbb{E}[U_1] = 0$ and $Var[U_1] = 1$ imply $\mathbb{E}[U_1^2] = 1$. Furthermore, using the assumption of no correlation between U_i and X_i , one gets

$$\mathbb{E}\left[(-1)^{|\boldsymbol{m}|} \frac{r(\boldsymbol{X_1})}{h(\boldsymbol{X_1})} U_1^2 (\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_1}) \right] = \mathbb{E}\left[(-1)^{|\boldsymbol{m}|} \frac{r(\boldsymbol{X_1})}{h(\boldsymbol{X_1})} (\partial^{\boldsymbol{m}} \Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_1}) \right]$$

and

$$\mathbb{E}\left[(-1)^{|\boldsymbol{m}|}\frac{f(\boldsymbol{X_1})g(\boldsymbol{X_1})}{h(\boldsymbol{X_1})}U_1(\partial^{\boldsymbol{m}}\Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_1})\right]=0.$$

In addition, because the density function of the random vector $\boldsymbol{X}$ is $h(\boldsymbol{x})$, the following equation can be obtained easily:

$$\mathbb{E}\left[(-1)^{|\boldsymbol{m}|}\frac{g^2(\boldsymbol{X_1})}{h(\boldsymbol{X_1})}(\partial^{\boldsymbol{m}}\Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_1})\right]=(-1)^{|\boldsymbol{m}|}\int_{[a,b]^d}g^2(\boldsymbol{x})(\partial^{\boldsymbol{m}}\Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x})d\boldsymbol{x}.$$

According to the above results, one has

$$\mathbb{E}[\widehat{\alpha}_{j_*,\boldsymbol{k}}]=\mathbb{E}\left[(-1)^{|\boldsymbol{m}|}\frac{r(\boldsymbol{X_1})}{h(\boldsymbol{X_1})}(\partial^{\boldsymbol{m}}\Phi_{j_*,\boldsymbol{k}})(\boldsymbol{X_1}).\right]$$

Then, it is easy to see from **A1** that

$$\begin{aligned}\mathbb{E}[\widehat{\alpha}_{j_*,\boldsymbol{k}}]&=(-1)^{|\boldsymbol{m}|}\int_{[a,b]^d}r(\boldsymbol{x})(\partial^{\boldsymbol{m}}\Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x})d\boldsymbol{x}\\&=\int_{[a,b]^d}(\partial^{\boldsymbol{m}}r)(\boldsymbol{x})\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x}=\alpha_{j_*,\boldsymbol{k}}.\end{aligned}$$

□

For nonparametric estimation, wavelet estimators can be viewed as generalized kernel estimators. For any $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^d$, we introduce the kernel $K(\boldsymbol{u},\boldsymbol{v})$ by $K(\boldsymbol{u},\boldsymbol{v})=\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\Phi(\boldsymbol{u}-\boldsymbol{k})\Phi(\boldsymbol{v}-\boldsymbol{k})$. Now, we give some important properties of this kernel function, which will be used in the later discussion. Furthermore, we define

$$K^{(\boldsymbol{m})}(\boldsymbol{u},\boldsymbol{v}):=\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\Phi(\boldsymbol{u}-\boldsymbol{k})(\partial_{\boldsymbol{v}}^{\boldsymbol{m}}\Phi)(\boldsymbol{v}-\boldsymbol{k}),$$

where $K^{(\boldsymbol{m})}(\boldsymbol{u},\boldsymbol{v}):=(\partial_{\boldsymbol{v}}^{\boldsymbol{m}}K)(\boldsymbol{u},\boldsymbol{v})$ denotes the $\boldsymbol{m}$th partial derivative of $K(\boldsymbol{u},\boldsymbol{v})$ with respect to $\boldsymbol{v}$.

Let the scaling function Φ be λ-regular [20,22,23], i.e., $\Phi\in C^{\lambda}$ and $|D^{\boldsymbol{\alpha}}\Phi(\boldsymbol{x})|\le c_l(1+\|\boldsymbol{x}\|)^{-l}$ for any integer $l\ge 1$, $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_d)\in\mathbb{N}^d$ with $|\boldsymbol{\alpha}|=\sum_{i=1}^{d}\alpha_i\le\lambda$ and $\boldsymbol{x}\in\mathbb{R}^d$. Then, there exists a positive constant C_d such that

$$|K^{(\boldsymbol{m})}(\boldsymbol{u},\boldsymbol{v})|\le\frac{C_d}{(1+\|\boldsymbol{u}-\boldsymbol{v}\|)^{d+1}}. \tag{5}$$

Meanwhile, one can obtain that

$$|K^{(\boldsymbol{m})}(\boldsymbol{u},\boldsymbol{y})-K^{(\boldsymbol{m})}(\boldsymbol{v},\boldsymbol{y})|\lesssim\|\boldsymbol{u}-\boldsymbol{v}\|. \tag{6}$$

For more properties and details of kernel functions, one can see [24,25].

Lemma 2. *For the model (1) with conditions* **A3** *and* **A4***, the wavelet estimator* $\widehat{(\partial^{\boldsymbol{m}}r)}(\boldsymbol{x})$ *is defined by (3) and* $2^{j_*}\lesssim(\frac{n}{\ln n})^{1/d}$*, there exists a constant* $\kappa>0$ *such that*

$$\mathbb{P}\left[\left|\widehat{(\partial^{\boldsymbol{m}}r)}(\boldsymbol{x})-\mathbb{E}[\widehat{(\partial^{\boldsymbol{m}}r)}(\boldsymbol{x})]\right|\ge\kappa\eta_n\right]\lesssim n^{-z(\kappa)},$$

where $z(\kappa)=\frac{\kappa^2}{2(1+\kappa/3)}$ *and* $\eta_n\asymp 2^{j_*(\frac{d}{2}+|\boldsymbol{m}|)}\sqrt{\frac{\ln n}{n}}$.

Proof. According to the definition of $\widehat{(\partial^m r)}(x)$ by (3),

$$
\begin{aligned}
&\left|\widehat{(\partial^m r)}(x) - \mathbb{E}[\widehat{(\partial^m r)}(x)]\right| \\
&= \left|\sum_{k \in \Lambda_{j_*}} (\widehat{\alpha}_{j_*,k} - \mathbb{E}[\widehat{\alpha}_{j_*,k}]) \Phi_{j_*,k}(x)\right| \\
&= \frac{1}{n}\left|\sum_{k \in \Lambda_{j_*}} \left(\sum_{i=1}^{n} \frac{Y_i^2}{h(X_i)} (\partial^m \Phi_{j_*,k})(X_i) - \mathbb{E}\left[\sum_{i=1}^{n} \frac{Y_i^2}{h(X_i)} (\partial^m \Phi_{j_*,k})(X_i)\right]\right) \Phi_{j_*,k}(x)\right| \\
&= \frac{1}{n}\left|\sum_{i=1}^{n}\left(\frac{Y_i^2}{h(X_i)} K_{j_*}^{(m)}(x, X_i) - \mathbb{E}\left[\frac{Y_i^2}{h(X_i)} K_{j_*}^{(m)}(x, X_i)\right]\right)\right| = \frac{1}{n}\left|\sum_{i=1}^{n} B_i\right|,
\end{aligned}
$$

where $B_i := \frac{Y_i^2}{h(X_i)} K_{j_*}^{(m)}(x, X_i) - \mathbb{E}\left[\frac{Y_i^2}{h(X_i)} K_{j_*}^{(m)}(x, X_i)\right]$ and

$$
\begin{aligned}
K_{j_*}^{(m)}(x, X_i) &:= \sum_{k \in \Lambda_{j_*}} \Phi_{j_*,k}(x) (\partial^m \Phi_{j_*,k})(X_i) \\
&= \sum_{k \in \Lambda_{j_*}} 2^{\frac{j_* d}{2}} \Phi(2^{j_*} x_1 - k_1, \cdots, 2^{j_*} x_d - k_d) \cdot \\
&\qquad 2^{j_*(\frac{d}{2} + |m|)} (\partial^m \Phi)(2^{j_*} X_{i_1} - k_1, \cdots, 2^{j_*} X_{i_d} - k_d) \\
&= 2^{j_*(d+|m|)} K^{(m)}(2^{j_*} x, 2^{j_*} X_i).
\end{aligned}
$$

Then we can obtain that

$$
\mathbb{P}\left[\left|\widehat{(\partial^m r)}(x) - \mathbb{E}[\widehat{(\partial^m r)}(x)]\right| \geq \kappa \eta_n\right] = \mathbb{P}\left[\frac{1}{n}\left|\sum_{i=1}^{n} B_i\right| \geq \kappa \eta_n\right]. \tag{7}
$$

By the definition of B_i, $\mathbb{E}[B_i] = 0$. Meanwhile, note that $|K_{j_*}^{(m)}(x, X_i)| \lesssim 2^{j_*(d+|m|)}$ by (5). Now, it follows from **A3** and **A4** that

$$
|B_i| \lesssim \left|\frac{Y_i^2}{h(X_i)} K_{j_*}^{(m)}(x, X_i)\right| + \left|\mathbb{E}\left[\frac{Y_i^2}{h(X_i)} K_{j_*}^{(m)}(x, X_i)\right]\right| \lesssim 2^{j_*(d+|m|)}.
$$

Using the property of the kernel function in (5),

$$
\begin{aligned}
\int_{[a,b]^d} \left(K_{j_*}^{(m)}(u, v)\right)^2 dv &= 2^{2j_*(d+|m|)} \int_{[a,b]^d} \left(K^{(m)}(2^{j_*} u, 2^{j_*} v)\right)^2 dv \\
&= 2^{2j_*(d+|m|)} 2^{-j_* d} \int_{[a,b]^d} \left(K^{(m)}(2^{j_*} u, 2^{j_*} v)\right)^2 d(2^{j_*} v) \\
&\lesssim 2^{j_*(d+2|m|)}.
\end{aligned}
$$

Then, by **A3** and **A4**, one gets

$$
\begin{aligned}
\mathbb{E}[B_i^2] = Var\left[\frac{Y_i^2}{h(X_i)} K_{j_*}^{(m)}(x, X_i)\right] &\leq \mathbb{E}\left[\frac{Y_i^4}{h^2(X_i)} \left(K_{j_*}^{(m)}(x, X_i)\right)^2\right] \\
\lesssim \mathbb{E}\left[\left(K_{j_*}^{(m)}(x, X_i)\right)^2\right] &\lesssim \int_{[a,b]^d} \left(K_{j_*}^{(m)}(x, v)\right)^2 dv \lesssim 2^{j_*(d+2|m|)}.
\end{aligned}
$$

According to the Bernstein's inequality [26] and the above results, one can obtain that

$$\mathbb{P}\left[\frac{1}{n}\left|\sum_{i=1}^{n} B_i\right| \geq \kappa\eta_n\right] \lesssim \exp\left\{-\frac{n\kappa^2\eta_n^2}{2\left(2^{j_*(d+2|\boldsymbol{m}|)} + \frac{\kappa\eta_n 2^{j_*(d+|\boldsymbol{m}|)}}{3}\right)}\right\}. \quad (8)$$

The conditions $\eta_n \asymp (\frac{\ln n}{n})^{\frac{1}{2}} \cdot 2^{j_*(\frac{d}{2}+|\boldsymbol{m}|)}$ and $2^{j_*} \lesssim (\frac{n}{\ln n})^{1/d}$ imply that $\eta_n \lesssim 2^{j_*|\boldsymbol{m}|}$. Meanwhile, one can easily obtain that

$$-\frac{n\kappa^2\eta_n^2}{2\left(2^{j_*(d+2|\boldsymbol{m}|)} + \frac{\kappa\eta_n 2^{j_*(d+|\boldsymbol{m}|)}}{3}\right)} = -\frac{\kappa^2}{2\left(1 + \frac{\kappa\eta_n}{3\cdot 2^{j_*|\boldsymbol{m}|}}\right)} \cdot \frac{n\eta_n^2}{2^{j_*(d+2|\boldsymbol{m}|)}} \lesssim -z(\kappa)\cdot \ln n,$$

with $z(\kappa) = \frac{\kappa^2}{2(1+\kappa/3)}$. Then, this result with (7) and (8) implies that

$$\mathbb{P}\left[\left|\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) - \mathbb{E}[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})]\right| \geq \kappa\eta_n\right] \lesssim n^{-z(\kappa)}.$$

□

4. Main Theorem

In this section, we will state the convergence rates of the wavelet estimator under different estimation error and mild conditions.

Theorem 1. *For the problem (1), the wavelet estimator $\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})$ is defined by (3), and the following results under different conditions are obtained.*

(i) *Let the model (1) satisfy the assumptions **A1–A4**,*

$$\sup_{\boldsymbol{x}\in[a,b]^d} |\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) - \mathbb{E}[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})]| = O_{a.s.}\left(\left(\frac{\ln n}{n}\right)^{\frac{1}{2}} 2^{j_*(\frac{d}{2}+|\boldsymbol{m}|)}\right). \quad (9)$$

(ii) *Assume that the model (1) satisfies the assumptions **A1–A4**, and the partial derivatives functions $(\partial^{\boldsymbol{m}} r)(\boldsymbol{x})$ belong to Hölder space $H^s(\mathbb{R}^d)(s > 0)$, one gets*

$$\sup_{\boldsymbol{x}\in[a,b]^d} |\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) - (\partial^{\boldsymbol{m}} r)(\boldsymbol{x})| = O_{a.s.}\left(\left(\frac{\ln n}{n}\right)^{\frac{1}{2}} 2^{j_*(\frac{d}{2}+|\boldsymbol{m}|)} + 2^{-j_* s}\right). \quad (10)$$

(iii) *Suppose that in the model (1) with conditions **A3** and **A4**, the random vector $\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}$ is independent. Then, one has*

$$Var\left[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})\right] \lesssim \frac{2^{j_*(d+2|\boldsymbol{m}|)}}{n}. \quad (11)$$

Remark 1. *If one takes $2^{j_*} \asymp (\frac{n}{\ln n})^{1/(d+2|\boldsymbol{m}|+2s)}$, then the result of (ii) reduces to*

$$\sup_{\boldsymbol{x}\in[a,b]^d} |\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) - (\partial^{\boldsymbol{m}} r)(\boldsymbol{x})| = O_{a.s.}\left(\left(\frac{\ln n}{n}\right)^{\frac{s}{d+2|\boldsymbol{m}|+2s}}\right).$$

Then, when $\boldsymbol{m} = \boldsymbol{0}$, this strong convergence rate of the wavelet estimator is the same as the optimal uniform almost sure convergence rate of nonparametric function problems [27].

Remark 2. *According to Lemma 1, it is easy to know that the wavelet estimator $\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})$ is an unbiased estimator. Hence, the estimation error of this wavelet estimator in the deviation sense is* *given by (i). In addition, the result (iii) considers the estimation error of the wavelet estimator in the variance sense.*

Proof. Proof of (i). Note that $[a,b]^d$ is a compact set, then it can be covered by a finite number L_n of cubes I_ℓ. Meanwhile, one defines that the centre of I_ℓ is $\boldsymbol{x}_\ell := (x_{\ell_1}, x_{\ell_2}, \cdots, x_{\ell_d})$, and the radius length is $l_n := \frac{c}{L_n^{1/d}}$ with a positive constant c. The parametric L_n will be taken in the following discussions. Using the triangle inequality, one gets

$$\sup_{\boldsymbol{x}\in[a,b]^d} |\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) - \mathbb{E}[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})]| \le Q_1 + Q_2 + Q_3, \tag{12}$$

where

$$Q_1 := \max_{1\le\ell\le L_n} \sup_{\boldsymbol{x}\in I_\ell} |\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) - \widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell)|,$$
$$Q_2 := \max_{1\le\ell\le L_n} \sup_{\boldsymbol{x}\in I_\ell} |\mathbb{E}[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell)] - \mathbb{E}[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})]|,$$
$$Q_3 := \max_{1\le\ell\le L_n} |\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell) - \mathbb{E}[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell)]|.$$

For Q_1. By the definitions of $\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})$ and $\widehat{\alpha}_{j_*,\boldsymbol{k}}$ in (3) and (4), for any $\boldsymbol{x} \in [a,b]^d$, one can easily obtain

$$\begin{aligned}
&|\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) - \widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell)| \\
&= \left|\sum_{\boldsymbol{k}\in\Lambda_{j_*}} \widehat{\alpha}_{j_*,\boldsymbol{k}}(\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}) - \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}_\ell))\right| \\
&\le \frac{1}{n}\left|\sum_{i=1}^{n} \frac{Y_i^2}{h(\boldsymbol{X}_i)}\left[K_{j_*}^{(\boldsymbol{m})}(\boldsymbol{x},\boldsymbol{X}_i) - K_{j_*}^{(\boldsymbol{m})}(\boldsymbol{x}_\ell,\boldsymbol{X}_i)\right]\right| \\
&\quad + \sum_{\boldsymbol{k}\in\Lambda_{j_*}}\left|\int_{[a,b]^d} g^2(\boldsymbol{x})(\partial^{\boldsymbol{m}}\Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x})d\boldsymbol{x}\cdot(\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}) - \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}_\ell))\right| \\
&=: Q_{11} + Q_{12}.
\end{aligned} \tag{13}$$

Using **A3**, **A4** and (6),

$$Q_{11} \lesssim \frac{1}{n}\left|\sum_{i=1}^{n}\left[K_{j_*}^{(\boldsymbol{m})}(\boldsymbol{x},\boldsymbol{X}_i) - K_{j_*}^{(\boldsymbol{m})}(\boldsymbol{x}_\ell,\boldsymbol{X}_i)\right]\right| \lesssim 2^{j_*(d+1+|\boldsymbol{m}|)}\|\boldsymbol{x} - \boldsymbol{x}_\ell\|. \tag{14}$$

In addition, by the boundedness assumption of function $g(\boldsymbol{x})$ in **A2**, the following inequalities are true:

$$\begin{aligned}
&\left|\int_{[a,b]^d} g^2(\boldsymbol{x})(\partial^{\boldsymbol{m}}\Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x})d\boldsymbol{x}\right| \\
&\lesssim \left|\int_{[a,b]^d} (\partial^{\boldsymbol{m}}\Phi_{j_*,\boldsymbol{k}})(\boldsymbol{x})d\boldsymbol{x}\right| \\
&= 2^{j_*(\frac{d}{2}+|\boldsymbol{m}|)}\left|\int_{[a,b]^d}(\partial^{\boldsymbol{m}}\Phi)(2^{j_*}x_1 - k_1,\cdots,2^{j_*}x_d - k_d)d\boldsymbol{x}\right| \\
&\le 2^{j_*(-\frac{d}{2}+|\boldsymbol{m}|)}\int_{[a,b]^d}\left|(\partial^{\boldsymbol{m}}\Phi)(2^{j_*}x_1 - k_1,\cdots,2^{j_*}x_d - k_d)\right|d(2^{j_*}\boldsymbol{x} - \boldsymbol{k}) \\
&\lesssim 2^{j_*(-\frac{d}{2}+|\boldsymbol{m}|)}.
\end{aligned}$$

Furthermore, it follows from the property of wavelet scaling function Φ that

$$\begin{aligned} Q_{12} &\lesssim 2^{j_*(-\frac{d}{2}+|\boldsymbol{m}|)} \sum_{\boldsymbol{k}\in\Lambda_{j_*}} \left|\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}) - \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}_\ell)\right| \\ &= 2^{j_*(-\frac{d}{2}+|\boldsymbol{m}|)} \cdot 2^{\frac{j_* d}{2}} \sum_{\boldsymbol{k}\in\Lambda_{j_*}} \left|\Phi(2^{j_*}\boldsymbol{x}-\boldsymbol{k}) - \Phi(2^{j_*}\boldsymbol{x}_\ell - \boldsymbol{k})\right| \\ &= 2^{j_*|\boldsymbol{m}|} \sum_{\boldsymbol{k}\in\Lambda_{j_*}} \left|\triangledown\Phi(2^{j_*}\boldsymbol{x}-\boldsymbol{k}) \cdot (2^{j_*}x_1 - 2^{j_*}x_{\ell_1}, \cdots, 2^{j_*}x_d - 2^{j_*}x_{\ell_d})^T\right| \\ &\lesssim 2^{j_*(1+|\boldsymbol{m}|)} \sum_{\boldsymbol{k}\in\Lambda_{j_*}} \|\boldsymbol{x}-\boldsymbol{x}_\ell\| \lesssim 2^{j_*(d+1+|\boldsymbol{m}|)}\|\boldsymbol{x}-\boldsymbol{x}_\ell\|. \end{aligned} \tag{15}$$

Combining (13), (14) and (15), one can get

$$|\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) - \widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell)| \lesssim 2^{j_*(d+1+|\boldsymbol{m}|)}\|\boldsymbol{x}-\boldsymbol{x}_\ell\|.$$

Because the centre of I_ℓ is $\boldsymbol{x}_\ell$, $\|\boldsymbol{x}-\boldsymbol{x}_\ell\| \lesssim l_n$. Then, by the definition of l_n,

$$Q_1 \lesssim 2^{j_*(d+1+|\boldsymbol{m}|)} l_n \lesssim \frac{2^{j_*(d+1+|\boldsymbol{m}|)}}{{L_n}^{1/d}}.$$

Now, one takes

$$L_n \asymp \left(\frac{2^{j_*(d+2)}n}{\ln n}\right)^{\frac{d}{2}}.$$

Then, the following conclusion is true,

$$Q_1 \lesssim \left(\frac{\ln n}{n}\right)^{\frac{1}{2}} 2^{j_*(\frac{d}{2}+|\boldsymbol{m}|)}. \tag{16}$$

For Q_2. Using the above discussions of Q_1, one knows

$$Q_2 \le \max_{1\le\ell\le L_n} \sup_{\boldsymbol{x}\in I_\ell} \mathbb{E}[|\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell) - \widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})|] \lesssim \left(\frac{\ln n}{n}\right)^{\frac{1}{2}} 2^{j_*(\frac{d}{2}+|\boldsymbol{m}|)}. \tag{17}$$

For Q_3. Note that

$$\begin{aligned} \mathbb{P}[Q_3 \ge \kappa\eta_n] &= \mathbb{P}\left[\max_{1\le\ell\le L_n} |\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell) - \mathbb{E}[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell)]| \ge \kappa\eta_n\right] \\ &\le \sum_{\ell=1}^{L_n} \mathbb{P}\left[|\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell) - \mathbb{E}[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}_\ell)]| \ge \kappa\eta_n\right] \\ &\le L_n \sup_{\boldsymbol{x}\in[a,b]^d} \mathbb{P}\left[|\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x}) - \mathbb{E}[\widehat{(\partial^{\boldsymbol{m}} r)}(\boldsymbol{x})]| \ge \kappa\eta_n\right]. \end{aligned}$$

By Lemma 2 and $2^{j_*} \lesssim (\frac{n}{\ln n})^{1/d}$, one can choose a large enough constant κ such that

$$\sum_{i=1}^{n} \mathbb{P}[Q_3 \ge \kappa\eta_n] \lesssim \sum_{i=1}^{n} L_n n^{-z(\kappa)} < \infty.$$

Furthermore, this result with the Borel–Cantelli lemma implies

$$Q_3 = O_{a.s.}\left(\left(\frac{\ln n}{n}\right)^{\frac{1}{2}} \cdot 2^{j_*(\frac{d}{2}+|\boldsymbol{m}|)}\right). \tag{18}$$

Finally, together with (12), (16), (17), and (18), one gets

$$\sup_{x\in[a,b]^d}|\widehat{(\partial^m r)}(x)-\mathbb{E}[\widehat{(\partial^m r)}(x)]|=O_{a.s.}\left(\left(\frac{\ln n}{n}\right)^{\frac{1}{2}}2^{j_*(\frac{d}{2}+|m|)}\right). \tag{19}$$

Proof of (ii). Using Lemma 1 and the property (2) of wavelets,

$$\left|\widehat{(\partial^m r)}(x)-(\partial^m r)(x)\right|\leq\left|\widehat{(\partial^m r)}(x)-\mathbb{E}[\widehat{(\partial^m r)}(x)]\right|+\left|\sum_{j=j_*}^{\infty}\sum_{u=1}^{2^d-1}\sum_{k\in\Lambda_j}\beta_{j,k,u}\Psi_{j,k,u}(x)\right|.$$

Hence,

$$\sup_{x\in[a,b]^d}\left|\widehat{(\partial^m r)}(x)-(\partial^m r)(x)\right|\leq I_1+I_2, \tag{20}$$

where

$$I_1:=\sup_{x\in[a,b]^d}\left|\widehat{(\partial^m r)}(x)-\mathbb{E}[\widehat{(\partial^m r)}(x)]\right|,$$

$$I_2:=\sup_{x\in[a,b]^d}\left|\sum_{j=j_*}^{\infty}\sum_{u=1}^{2^d-1}\sum_{k\in\Lambda_j}\beta_{j,k,u}\Psi_{j,k,u}(x)\right|.$$

For I_1. According to the conclusion of (i),

$$I_1=O_{a.s.}\left(\left(\frac{\ln n}{n}\right)^{\frac{1}{2}}2^{j_*(\frac{d}{2}+|m|)}\right). \tag{21}$$

For I_2. Let a function $f(x)$ belong to Hölder space $H^s(\mathbb{R}^d)$, and let $\Psi_{j,k,u}$ be a wavelet function, then one can prove that $\left|\sum_{j=j_*}^{\infty}\sum_{u=1}^{2^d-1}\sum_{k\in\Lambda_j}\beta_{j,k,u}\Psi_{j,k,u}(x)\right|\lesssim 2^{-j_*s}$. More details and proofs of this above conclusion can be found in [28–30]. Furthermore, because the partial derivatives functions $(\partial^m r)(x)$ belong to Hölder space $H^s(\mathbb{R}^d)$, one can easily obtain that

$$I_2\lesssim 2^{-j_*s}.$$

Now, this conclusion with (20) and (21) shows that

$$\sup_{x\in[a,b]^d}|\widehat{(\partial^m r)}(x)-(\partial^m r)(x)|=O_{a.s.}\left(\left(\frac{\ln n}{n}\right)^{\frac{1}{2}}2^{j_*(\frac{d}{2}+|m|)}+2^{-j_*s}\right).$$

Proof of (iii). By the definition of $\widehat{(\partial^m r)}(x)$ and the properties of the variance function, one has

$$\begin{aligned}Var\left[\widehat{(\partial^m r)}(x)\right]&=Var\left[\sum_{k\in\Lambda_{j_*}}\left(\frac{(-1)^{|m|}}{n}\sum_{i=1}^{n}\frac{Y_i^2}{h(X_i)}(\partial^m\Phi_{j_*,k})(X_i)\right)\Phi_{j_*,k}(x)\right]\\&=Var\left[\frac{(-1)^{|m|}}{n}\sum_{i=1}^{n}\frac{Y_i^2}{h(X_i)}K_{j_*}^{(m)}(x,X_i)\right].\end{aligned}$$

Moreover, the assumptions of random vector X, **A3** and **A4** imply that

$$Var\left[\widehat{(\partial^m r)}(x)\right]\lesssim\frac{1}{n}Var\left[K_{j_*}^{(m)}(x,X_1)\right].$$

Using the property of the kernel function in (5) and condition **A3**,

$$Var\left[K_{j_*}^{(m)}(x, X_1)\right] \le \mathbb{E}\left[(K_{j_*}^{(m)}(x, X_1))^2\right] \lesssim \int_{[a,b]^d} (K_{j_*}^{(m)}(x, v))^2 dv \lesssim 2^{j_*(d+2|m|)}.$$

Hence,

$$Var\left[\widehat{(\partial^m r)}(x)\right] \lesssim \frac{2^{j_*(d+2|m|)}}{n}.$$

□

5. Conclusions

For nonparametric derivative estimation, classical research results pay more attention to the derivative estimation of one-dimensional functions. However, this paper studies the nonparametric estimation of partial derivatives of a multivariate function. Firstly, a wavelet estimator of the partial derivatives of the multivariate variance function in a heteroscedastic model is given. More importantly, this wavelet estimator is an unbiased estimator. Secondly, two important lemma are proved, which discuss the key properties of the wavelet estimator. Finally, the convergence rates over different estimation errors of the wavelet estimator are considered. According to the main theorem, it is easy to see that the strong convergence rate of the wavelet estimator is the same as the optimal uniform almost sure convergence rate of nonparametric function estimations.

Because the local analysis characteristics in the time and frequency domains of the wavelet, the wavelet estimator can choose an appropriate wavelet scaling parameter to get the optimal convergence rate. Hence, this paper considers partial derivatives estimation based on the wavelet method. The theoretical results of asymptotic property of the wavelet estimator are discussed in this paper. In addition, it is difficult to present the corresponding practical illustration of the wavelet estimator, which needs more investigations and some new skills. We will consider this in future work.

Author Contributions: Conceptualization, J.K. and H.Z.; writing—original draft preparation, J.K. and H.Z.; writing—review and editing, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is supported by the Guangxi Natural Science Foundation (No. 2023GXNSFAA026042), the National Natural Science Foundation of China (No. 12361016), the Center for Applied Mathematics of Guangxi (GUET), and the Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation.

Data Availability Statement: Data are contained within the article.

Acknowledgments: All authors would like to thank the reviewers for their important and constructive comments.

Conflicts of Interest: All authors state that there are no conflicts of interest.

References

1. Box, G. Signal-to-noise ratios, performance criteria, and transformations. *Technometrics* **1988**, *30*, 1–17. [CrossRef]
2. Smeds, K.; Wolters, F.; Rung, M. Estimation of signal-to-noise ratios in realistic sound scenarios. *J. Am. Acad. Audiol.* **2015**, *26*, 183–196. [CrossRef] [PubMed]
3. Yao, Q.W.; Tong, H. Quantifying the influence of initial values on nonlinear prediction. *J. R. Stat. Soc. Ser. B* **1994**, *56*, 701–725.
4. Härdle, W.; Tsybakov, A. Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econom.* **1997**, *81*, 223–242. [CrossRef]
5. Mak, T.K. Estimation of parameters in heteroscedastic linear models. *J. R. Stat. Soc. Ser. B* **1992**, *54*, 649–655. [CrossRef]
6. Shen, S.L.; Mei, C.L. Estimation of the variance function in heteroscedastic linear regression models. *Commun. Stat. Theory Methods* **2009**, *38*, 1098–1112. [CrossRef]

7. Su, Z.H.; Cook, R.D. Estimation of multivariate means with heteroscedastic errors using envelope models. *Stat. Sin.* **2013**, *23*, 213–230. [CrossRef]
8. Zhang, J.; Huang, Z.S. Efficient simultaneous partial envelope model in multivariate linear regression. *J. Stat. Comput. Simul.* **2022**, *92*, 1373–1400. [CrossRef]
9. Fan, J.Q.; Yao, Q.W. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **1998**, *85*, 645–660. [CrossRef]
10. Brown, D.L.; Levine, M. Variance estimation in nonparametric regression via the difference sequence method. *Ann. Stat.* **2007**, *35*, 2219–2232. [CrossRef]
11. Wang, L.; Brown, D.L.; Cai, T.T. Effect of mean on variance function estimation in nonparametric regression. *Ann. Stat.* **2008**, *36*, 646–664. [CrossRef]
12. Shen, Y.D.; Gao, C.; Witten, D.; Han, F. Optimal estimation of variance in nonparametric regression with random design. *Ann. Stat.* **2020**, *48*, 3589–3618. [CrossRef]
13. Zaoui A. Variance function estimation in regression model via aggregation procedures. *J. Nonparametric Stat.* **2023**, *35*, 397–436. [CrossRef]
14. Zhou, S.G.; Wolfe, D.A. On derivatives estimation in spline regression. *Stat. Sin.* **2000**, *10*, 93–108.
15. Cai, T.T. On adaptive wavelet estimation of a derivative and other related linear inverse problems. *J. Stat. Plan. Inference* **2002**, *108*, 329–349. [CrossRef]
16. Chaubey, Y.P.; Doosti, H.D.; Rao, B.L.S.P. Wavelet based estimation of the derivatives of a density for a negatively associated process. *J. Stat. Theory Pract.* **2008**, *2*, 453–463. [CrossRef]
17. Hosseinioun, N.; Doosti, H.; Nirum, H.A. Nonparametric estimation of the derivative of a density by the method of wavelet for mixing sequences. *Stat. Pap.* **2012**, *53*, 195–203. [CrossRef]
18. Kou, J.K.; Zhang, H. Wavelet estimation of the derivatives of variance function in heteroscedastic model. *AIMS Math.* **2023**, *8*, 14340–14361. [CrossRef]
19. Daubechies, I. *Ten Lectures on Wavelets*; SIAM: Philadelphia, PA, USA, 1992.
20. Härdle, W.; Kerkyacharian, G.; Picard, D.; Tsybakov, A. *Wavelets, Approximation and Statistical Application*; Springer: New York, NY, USA, 1998.
21. Meyer, Y. *Wavelet and Operators*; Hermann: Paris, France, 1990.
22. Walnut, D.F. *An Introduction to Wavelet Analysis*; Birkhäuser: Basel, Switzerland, 2001.
23. Triebel, H. *Theory of Function Spaces*; Birkhäuser: Basel, Switzerland, 2001.
24. Masry, E. Multivariate probability density estimation by wavelet methods: Strong consistency and rates for stationary time series. *Stoch. Process. Their Appl.* **1997**, *67*, 177–193. [CrossRef]
25. Allaoui, S.; Bouzebda, S.; Chesneau, C.; Liu, J.C. Uniform almost sure convergence and asymptotic distribution of the wavelet-based estimators of partial derivatives of multivariate density function under weak dependence. *J. Nonparametric Stat.* **2021**, *33*, 170–196. [CrossRef]
26. Kou, J.K.; Huang, Q.M.; Guo, H.J. Pointwise wavelet estimations for a regression model in local Hölder space. *Axioms* **2022**, *11*, 466. [CrossRef]
27. Giné, E.; Nickl, R. Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **2009**, *37*, 1605–1646. [CrossRef]
28. Devore, R.; Kerkyacharian, G.; Picard, D.; Temlyakov, V. Approximation methods for supervised learning. *Found. Comput. Math.* **2006**, *6*, 3–58. [CrossRef]
29. Liu, Y.M.; Wu, C. Point-wise estimation for anisotropic densities. *J. Multivar. Anal.* **2019**, *171*, 112–125. [CrossRef]
30. Chen, L.; Chesneau, C.; Kou, J.K.; Xu, J.L. Wavelet optimal estimation for a multivariate probability density function under weighted distribution. *Results Math.* **2023**, *78*, 66. [CrossRef]