# LOCAL $M$-ESTIMATION FOR CONDITIONAL VARIANCE FUNCTION WITH DEPENDENT DATA

YUNYAN WANG AND MINGTIAN TANG

ABSTRACT. In this paper, a local $M$-estimation for the conditional variance function in heteroscedastic regression models under stationary $\alpha$-mixing dependent samples is developed. The local $M$-estimator is based on the local linear smoothing technique and the $M$-estimation technique, and it is shown to be not only asymptotically equivalent to the local linear estimator but also robust. The weak consistency as well as the asymptotic normality of the local $M$-estimator for the conditional variance function are obtained under mild conditions.

**1. Introduction.** Let $\{(Y_i, X_i)\}$ be a two-dimensional strictly stationary process with the same marginal distribution as $(Y, X)$. Let $m(x) = E(Y|X = x)$ and $\sigma^2(x) = \mathrm{Var}\,(Y|X = x) \neq 0$ be the conditional mean function and conditional variance function, respectively. Consider the nonparametric heteroscedastic regression model:

$$(1.1) \qquad Y_i = m(X_i) + \sigma(X_i)\varepsilon_i,$$

where $E(\varepsilon_i|X_i) = 0$ and $\mathrm{Var}\,(\varepsilon_i|X_i) = 1$.

As is known, there is much literature concerning the estimation of the conditional mean function $m(x)$, for example, [**8, 19, 41**] and the references therein. Here, we are interested in the nonparametric robust estimation of the conditional variance function $\sigma^2(x)$ under stationary $\alpha$-mixing dependent samples.

Conditional variance estimation plays an important role in many applications. It is crucial in inference for the parameters in the conditional mean function. For example, Holst et al. [**30**] used local polynomial

regression for evaluation of the concentration of atmospheric atomic mercury measured with the LIght Detection And Ranging (LIDAR) technique; because of severe heteroscedasticity, the variance function must be estimated to obtain a satisfactory bandwidth for the derivative of the mean function. In a time series context, Hansen [26] developed parameter estimation of the conditional mean function by incorporating nonparametric conditional variance estimates.

The variance function is of interest in its own right. Local variability of the data is often described by the conditional variance function in a statistical model. For example, measuring volatility or risk in finance (Andersen and Lund [2]), identifying homoscedastic transforms in regression (Carroll and Ruppert [9]) and monitoring the signal-to-noise ratios in quality control of experimental design (Box [7]).

Moreover, the conditional variance function is crucial for statistical analysis. It can be used to construct confidence intervals. In finance, conditional variance is also called volatility, which refers to the characteristic of the non-constant variation and is directly related to option pricing and risk measure quantification.

Since the conditional variance function $\sigma^2(x)$ can model the variability of statistical data and its applications in financial risk management, its estimation has become increasingly popular. In nonparametric theory, various methods of estimation of the volatility function have appeared. For example, Ruppert et al. [38] studied the conditional variance function in a nonparametric heteroscedastic regression model by linear smoothing of squared residuals. Fan and Yao [20] proposed a local linear estimator for the conditional variance of model (1.1). Vilar-Fernández and Francisco-Fernández [42] studied two nonparametric estimators of the conditional variance previously studied in Härdle and Tsybakov [29] and Fan and Yao [20] in different contexts, and showed that the leading term of the asymptotic variance of these estimators is different from that obtained in Härdle and Tsybakov [29] and Fan and Yao [20]. Chen et al. [12] applied variance reduction techniques to the estimator of Fan and Yao [20] to improve the inference for the conditional variance. Chan et al. [10] applied empirical likelihood methods to construct confidence intervals for the conditional variance based on the estimator of Fan and Yao [20] and the reduced variance modification of Chen et al. [12]. Xu and Phillips [44] developed a reweighted local constant estimator of the conditional variance

function based on an empirical likelihood modification of conventional local level nonparametric regression applied to squared mean regression residuals. Kulik and Wichelhaus [**32**] studied the asymptotic properties of a nonparametric kernel estimator of conditional variance function in nonparametric regression models with long memory errors and predictors, and obtained the nonparametric estimator of the error density. Moreover, this paper showed that the Nadaraya-Watson estimation of the conditional mean had influence on the estimation of the conditional variance and the error density.

However, the nonparametric methods mentioned above focus mainly on the independent identically distributed case. There are many statistical analysis problems about high-dimensional, dependent and incomplete data in many areas, such as modern science, technology and socioeconomic. With the development of computer technology, statistics is facing unprecedented challenges and opportunities, in which the statistical analysis of dependent data is a universal highlight problem. For example, the signal data which are received and transmitted by the radar network and information network are dependent. Thus, the study of dependent random variables has attracted the attention of statistical workers. The book by Györfi et al. [**24**] studied the nonparametric curve estimation from time series. Roussas [**37**] considered the nonparametric estimation for regression function under four kinds of mixing conditions, i.e. $\psi$-mixing, $\phi$ (or uniform)-mixing, $\rho$-mixing and $\alpha$ (or strong)-mixing, developed the strongly consistency of the estimator, and obtained the rates of convergence. Yoshihara [**45**] considered weakly dependent stochastic sequences and studied the curve estimation based on weakly dependent data. Masry and Fan [**34**] studied the local polynomial estimation for conditional mean functions and their derivatives in regression functions under strongly mixing and $\rho$-mixing conditions. Liang [**33**] established the asymptotic normality of nonparametric kernel estimator of the regression function for the left-truncation model when the data exhibit some kind of dependence (it is assumed that the lifetime observations with multivariate covariates form a stationary $\alpha$-mixing sequence), and obtained a uniform weak convergence rate for the product-limit estimator of the lifetime and truncated distributions under dependence, which is interesting independently.

Moreover, the nonparametric methods mentioned above are based

on local polynomial regression smoothers, and they are not robust. Unnoticed or difficult to detect unforeseen circumstances often happened in the process of accessing data, and all these unforeseen circumstances will make the data inevitably contain more or less abnormal values; hence, these abnormal observations are observed very often in economic time series and finance as well as in many other applied fields. Since individual outliers will make larger changes to the results of statistical inference and lead to unreasonable or wrong conclusions, such statistical methods are not "robust" and cannot adapt to complex changes in the actual situation.

In order to fine tune the abnormal observations, there is a growing literature on robust methods. Among the authors with the most interesting contributions in the area is Graciela Boente. For example, Boente et al. [5] introduced a family of robust estimates for the parametric and nonparametric components under a generalized partially linear model and established the consistency and asymptotic normality of the proposed estimators. Boente et al. [6] considered the local $M$-estimators of the scale function based on consecutive differences of the responses and developed the asymptotic behavior of the local $M$-estimators for general weight functions. Bianco et al. [4] presented two families of nonparametric and robust estimators of regression function in the presence of anomalous data, and studied the asymptotic behavior of the robust estimators.

One popular robust technique is the so-called $M$-estimators. $M$-estimation has been studied by many authors in parametric, semiparametric and nonparametric settings, for example, Delecroix et al. [15], Davis et al. [14], Hall and Jones [25] and the references therein. Furthermore, some modified $M$-estimators were proposed, such as the local $M$-estimator, which is a combination of the local linear smoothing technique and the $M$-estimation technique. The local $M$-estimators inherit nice properties from not only $M$-estimators but also from local linear estimators. For example, Fan and Jiang [19] developed local linear $M$-estimator with variable bandwidth for regression function and Jiang and Mack [31] considered a robust estimator of the regression function by using local polynomial regression techniques.

The paper is structured as follows. Section 2 introduces the local $M$-estimator of the conditional variance function in heteroscedastic regression models under stationary $\alpha$-mixing dependent samples. The

consistency and asymptotic normality of the new estimator are developed under some mild conditions in Section 3. Some technical lemmas and all mathematical proofs of our main results are presented in Section 4. Section 5 is the conclusion.

**2. Local $M$-estimator and assumptions.** Our main interest in this work is in estimating the conditional variance function $\sigma^2(\cdot)$. The local linear estimator of $\sigma^2(x)$ proposed by Fan and Yao [20] is defined as the solution to the following problem: find $\alpha_0$ and $\alpha_1$ to minimize the following weighted sum

$$(2.1) \qquad \sum_{i=1}^{n} \left([Y_i - \widehat{m}(X_i)]^2 - \alpha_0 - \alpha_1(X_i - x)\right)^2 K\left(\frac{X_i - x}{h}\right),$$

where $K(\cdot)$ is kernel function and $h$ is a sequence of positive numbers tending to zero, called the *smoothing parameter* or *bandwidth*, and $\widehat{m}(x)$ is the local linear estimator of the conditional mean function $m(x)$, that is,

$$\widehat{m}(x) = \frac{\sum\limits_{i=1}^{n} \omega_{1i} W_{\bar{h}}(X_i - x) Y_i}{\sum\limits_{i=1}^{n} \omega_{1i} W_{\bar{h}}(X_i - x)},$$

where $W(\cdot)$ is a kernel function and $\bar{h}$ is a bandwidth,

$$\omega_{1i} = \overline{S}_{n,2} - (X_i - x)\overline{S}_{n,1}$$

with

$$\overline{S}_{n,j} = \sum_{i=1}^{n} (X_i - x)^j W_{\bar{h}}(X_i - x), \quad j = 1, 2$$

and $W_{\bar{h}}(\cdot) = W(\cdot/\bar{h})/\bar{h}$.

In fact, the local linear estimator which is obtained from (2.1) is based on the least-squares principle and is not robust. Therefore, we propose to choose $a$ and $b$ to minimize

$$(2.2) \qquad \sum_{i=1}^{n} \rho\left([Y_i - \widehat{m}(X_i)]^2 - a - b(X_i - x)\right) K\left(\frac{X_i - x}{h}\right)$$

or to satisfy the local estimation equations:

(2.3)
$$\sum_{i=1}^{n} \psi\left([Y_i - \widehat{m}(X_i)]^2 - a - b(X_i - x)\right) K\left(\frac{X_i - x}{h}\right) \left(\begin{array}{c} 1 \\ X_i - x/h \end{array}\right)$$
$$= \left(\begin{array}{c} 0 \\ 0 \end{array}\right),$$

where $\rho(\cdot)$ is a given outlier-resistant loss function and $\psi(\cdot)$ is the derivative of $\rho(\cdot)$.

Let $\dot{\sigma}^2(x)$ be the first order derivative of the conditional variance $\sigma^2(x)$, and the $M$-type estimators of $\sigma^2(x)$ and $\dot{\sigma}^2(x)$ are denoted as $\widehat{\sigma}^2(x) = \widehat{a}$ and $\widehat{\dot{\sigma}}^2(x) = \widehat{b}$, which are the solutions of (2.3).

Let $p(\cdot)$ be the marginal density function of $X$, and let $x_0$ be a given point in $[-1, 1]$. The conditions that are necessary in deriving the asymptotic properties of the local $M$-estimators are as follows.

**Assumption 1.**

(i) The marginal density function $p(x)$ of the process $X$ is continuous at $x_0$, and $p(x_0) > 0$.
(ii) The second derivative $\ddot{m}(x) = d^2[m(x)]/dx^2$ and $\ddot{\sigma}^2(x) = d^2[\sigma^2(x)]/dx^2$ are continuous at the point $x_0$.

**Assumption 2.** Assume that the kernel functions $W(\cdot)$ and $K(\cdot)$ are continuous probability density functions each with bounded support on $[-1, 1]$.

**Assumption 3.** Assume that $n \to \infty$, $\bar{h} \to 0$ and $h \to 0$ such that $n\bar{h} \to \infty$ and $nh \to \infty$.

**Assumption 4.** $E[\psi(u_i)|\, X_i = x] = o(1)$ with $u_i = [Y_i - \widehat{m}(X_i)]^2 - \sigma^2(X_i)$.

**Assumption 5.** The function $\psi'(\cdot)$ satisfies

$$E[\sup_{|z|\leq\delta} |\psi'(u_i + z) - \psi'(u_i)||\, X_i = x] = o(1)$$

and

$$E[\sup_{|z|\leq\delta} |\psi(u_i + z) - \psi(u_i) - \psi'(u_i)z||\, X_i = x] = o(\delta)$$

as $\delta \to 0$ uniformly in $x$ in a neighborhood of $x_0$;

**Assumption 6.** The function $\psi(\cdot)$ is continuous and has a derivative $\psi'(\cdot)$ almost everywhere. Furthermore, it is assumed that functions

$$E[\psi'(u_i)| X_i = x] > 0, \qquad E[\psi^2(u_i)\big| X_i = x] > 0,$$
$$E[\psi'^2(u_i)\Big| X_i = x] > 0,$$

and continuous at the point $x_0$, and there exists a constant $\gamma > 0$ such that

$$E[|\psi(u_i)|^{2+\gamma}\Big| X_i = x], \qquad E[|\psi'(u_i)|^{2+\gamma}\Big| X_i = x]$$

are bounded in a neighborhood of $x_0$.

**Assumption 7.** For any $i, j$,

$$E[\psi^2(u_i) + \psi^2(u_j)\big| X_i = x, X_j = y],$$
$$E[\psi'^2(u_i) + \psi'^2(u_j)\Big| X_i = x, X_j = y]$$

are bounded in the neighborhood of $x_0$.

**Remark 2.1.** The conditions in Assumptions 4–7 imposed on $\psi(\cdot)$ and $\psi'(\cdot)$ are mild and satisfied for many applications. Particularly, they are fulfilled for Huber's $\psi(\cdot)$ function. In this paper, unlike Härdle [**27, 28**], we do not need the monotonicity and boundedness of $\psi(\cdot)$ nor the convexity of the function $\rho(\cdot)$. In the paper, the function $\rho(\cdot)$ can choose $\rho(x) = x^2$; in other words, the well-known least square estimator (LSE) can be used in practice and satisfies all the technical conditions above. For more details about these conditions we refer to Fan and Jiang [**19**] or Cai and Ould-Said [**8**].

**Assumption 8.** Assume that $\{(X_j, Y_j)\}$ is a stationary $\alpha$-mixing process, and the mixing coefficient $\alpha(n)$ satisfies

$$\sum_{n \geq 1} n^a (\alpha(n))^{\gamma/(2+\gamma)} < \infty$$

for some $a > \gamma/(2 + \gamma)$, where $\gamma$ is given in Assumption 6.

**Assumption 9.** Assume that there exists a sequence of positive integers $q_n$ such that $q_n \to \infty$, $q_n = o((nh)^{1/2})$ and $(n/h)^{1/2}\alpha(q_n) \to 0$ as $n \to \infty$.

**Assumption 10.** There exists $\tau > 2 + \gamma$, where $\gamma$ is given in Assumption 6, such that $E\{|\psi(u_i)|^\tau | X_i = x\}$ is bounded for all $x$ in a neighborhood of $x_0$, and $\alpha(n) = O(n^{-\theta})$, where $\theta \geq (2 + \gamma)\tau / \{2(\tau - 2 - \gamma)\}$.

**Assumption 11.** $n^{-\gamma/4} h^{(2+\gamma)/\tau - 1 - \gamma/4} = O(1)$, where $\gamma$ is given in Assumption 6 and $\tau$ is given in Assumption 10.

**3. Main results.** Throughout the whole paper, let

$$K_l = \int u^l K(u)\, du, \qquad\qquad J_l = \int u^l K^2(u)\, du, \quad \text{for } l \geq 0,$$

$$U = \begin{pmatrix} K_0 & K_1 \\ K_1 & K_2 \end{pmatrix}, \qquad\qquad V = \begin{pmatrix} J_0 & J_1 \\ J_1 & J_2 \end{pmatrix}, \quad A = \begin{pmatrix} K_2 \\ K_3 \end{pmatrix},$$

$$G_1(x) = E[\psi'(u_i)| X_i = x],$$
$$G_2(x) = E[\psi^2(u_i)| X_i = x],$$
$$G_3(x) = E[\psi'^2(u_i)| X_i = x].$$

Our main results are as follows:

**Theorem 3.1.** *Under Assumptions 1–8, there exist solutions $\widehat{\sigma}^2(x_0)$ and $\widehat{\dot{\sigma}}^2(x_0)$ to equations (2.3) such that*

$$\begin{pmatrix} \widehat{\sigma}^2(x_0) - \sigma^2(x_0) \\ h[\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0)] \end{pmatrix} \xrightarrow{P} 0, \quad n \to \infty.$$

**Theorem 3.2.** *Under Assumptions 1–11, for the solutions given in Theorem 3.1, we have*

$$\sqrt{nh}\left[\begin{pmatrix} \widehat{\sigma}^2(x_0) - \sigma^2(x_0) \\ h[\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0)] \end{pmatrix} - \frac{h^2 \ddot{\sigma}^2(x_0)}{2} U^{-1} A\right] \xrightarrow{D} N(0, \Sigma),$$

*where*

$$\Sigma = \frac{G_2(x_0)}{G_1^2(x_0) p(x_0)} U^{-1} V U^{-1}.$$

**4. Lemmas and proofs.** In order to prove Theorems 3.1 and 3.2, we need to establish the following lemmas.

**Lemma 4.1.** *Under Assumptions* 1–8, *we have*

$$\sum_{i=1}^{n} \psi'(u_i)K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^l = nh^{l+1}G_1(x_0)p(x_0)K_l(1 + o_p(1)),$$

*and*

$$\sum_{i=1}^{n} \psi'(u_i)R(X_i)K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^l$$
$$= \frac{nh^{l+3}}{2}G_1(x_0)\ddot{\sigma}^2(x_0)p(x_0)K_{l+2}(1 + o_p(1)),$$

*where* $R(X_i) = \sigma^2(X_i) - \sigma^2(x_0) - \dot{\sigma}^2(x_0)(X_i - x_0)$.

*Proof of Lemma* 4.1. Since the second part of the lemma can be proved by the same arguments as the first one, we only prove the first part. Let

$$Z_{n,i} = \psi'(u_i)K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^l.$$

By continuation at the point $x_0$ of $K(\cdot)$, $G_1(\cdot)$ and $p(\cdot)$, we obtain

$$E(Z_{n,1}) = \int G_1(x)K\left(\frac{x - x_0}{h}\right)(x - x_0)^l p(x)\, dx$$
$$= \int G_1(x_0 + yh)K(y)(yh)^l p(x_0 + yh)h\, dy$$
$$= h^{l+1}G_1(x_0)p(x_0)\int K(y)y^l\, dy(1 + o(1))$$
$$= h^{l+1}G_1(x_0)p(x_0)K_l(1 + o(1)).$$

Therefore, we have

$$E\left(\sum_{i=1}^{n} \psi'(u_i)K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^l\right) = nh^{l+1}G_1(x_0)p(x_0)(x_0)K_l(1 + o(1)).$$

Note that

$$\sum_{i=1}^{n} Z_{n,i} = E\left(\sum_{i=1}^{n} Z_{n,i}\right) + O_p\left(\sqrt{\text{Var}\left(\sum_{i=1}^{n} Z_{n,i}\right)}\right),$$

and

$$\text{Var}\left(\sum_{i=1}^{n} Z_{n,i}\right) = nEZ_{n,1}^2 + 2\sum_{j=2}^{n}(n-j+1)\text{Cov}\left(Z_{n,1}, Z_{n,j}\right).$$

By continuation at the point $x_0$ of $K(\cdot)$, $G_3(\cdot)$ and $p(\cdot)$, we obtain

$$
\begin{aligned}
EZ_{n,1}^2 &= \int G_3(x)K^2\left(\frac{x-x_0}{h}\right)(x-x_0)^{2l}p(x)\,dx \\
&= \int G_3(x_0+yh)K^2\left(y\right)(yh)^{2l}p(x_0+yh)h\,dy \\
&= h^{2l+1}G_3(x_0)p(x_0)\int K^2\left(y\right)y^{2l}dy(1+o(1)) = O(h^{2l+1}).
\end{aligned}
$$

Let $d_n$ be a sequence of positive integers satisfying $d_n \to \infty$ and $hd_n \to 0$. Then we have

$$\sum_{j=2}^{n}|\text{Cov}\left(Z_{n,1}, Z_{n,j}\right)| = \sum_{j=2}^{d_n}|\text{Cov}\left(Z_{n,1}, Z_{n,j}\right)| + \sum_{j=d_n+1}^{n}|\text{Cov}\left(Z_{n,1}, Z_{n,j}\right)|.$$

By Assumption 7 and the bounded support of $K(\cdot)$, we have

$$
\begin{aligned}
|EZ_{n,i}Z_{n,j}| \leq E\,|Z_{n,i}Z_{n,j}| &= E\Big|E[\psi'(u_i)\psi'(u_j)\,|\,X_i, X_j] \\
&\quad \cdot K\left(\frac{X_i-x_0}{h}\right)(X_i-x_0)^l K\left(\frac{X_j-x_0}{h}\right)(X_j-x_0)^l\Big| \\
&\leq C_1 E\Big|K\left(\frac{X_i-x_0}{h}\right)(X_i-x_0)^l K\left(\frac{X_j-x_0}{h}\right)(X_j-x_0)^l\Big| \\
0 &\leq C_2 h^{2l+2},
\end{aligned}
$$

where $C_1$ and $C_2$ are constants. Therefore, we have

$$\sum_{j=2}^{d_n}|\text{Cov}\left(Z_{n,1}, Z_{n,j}\right)| \leq C_2 h^{2l+2}\sum_{j=2}^{d_n}1 = o(nh^{2l+1}).$$

By using Davydov's inequality, we have

$$|\text{Cov}\left(Z_{n,1}, Z_{n,j}\right)| \leq C_3[\alpha(j-1)]^{\gamma/(2+\gamma)}\left(E|Z_{n,1}|^{2+\gamma}\right)^{2/(2+\gamma)},$$

and, by Assumption 6, we have

$$E|Z_{n,i}|^{2+\gamma} = E\left|E\left[\psi'(u_i)|\,X_i\right]K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^l\right|^{2+\gamma}$$

$$\leq C_4 E\left|K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^l\right|^{2+\gamma} \leq C_5 h^{(2+\gamma)l+1},$$

where $C_3$, $C_4$ and $C_5$ are constants. Therefore, by using Assumption 8 and choosing $d_n$ such that $d_n^a h^{\gamma/(2+\gamma)} = O(1)$, we have

$$\sum_{j=d_n+1}^{n} |\text{Cov}\,(Z_{n,1}, Z_{n,j})| \leq C_6 \sum_{j=d_n+1}^{n} [\alpha(j-1)]^{\gamma/(2+\gamma)} \left(h^{(2+\gamma)l+1}\right)^{2/(2+\gamma)}$$

$$= C_6 h^{2l+2/(2+\gamma)} \sum_{k=d_n}^{n} [\alpha(k)]^{\gamma/(2+\gamma)}$$

$$\leq C_6 d_n^{-a} h^{2l+2/(2+\gamma)} \sum_{k=d_n}^{n} k^a [\alpha(k)]^{\gamma/(2+\gamma)} = o(nh^{2l+1}),$$

where $C_6$ is a constant. In summary, we have

$$\text{Var}\left(\sum_{i=1}^{n} Z_{n,i}\right) = O(nh^{2l+1}).$$

Therefore,

$$\sum_{i=1}^{n} \psi'(u_i)K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^l = nh^{l+1}G_1(x_0)p(x_0)K_l(1 + o_p(1)).$$

This completes the proof of Lemma 4.1. $\qquad\square$

**Lemma 4.2.** *Under Assumptions 1–4 and 6–11, we have*

$$\frac{1}{\sqrt{nh}}\begin{pmatrix} \sum_{i=1}^{n} \psi(u_i)K\left(\frac{X_i-x_0}{h}\right) \\ \sum_{i=1}^{n} \psi(u_i)K\left(\frac{X_i-x_0}{h}\right)\frac{X_i-x_0}{h} \end{pmatrix} \xrightarrow{D} N(0, \Sigma_1),$$

*where* $\Sigma_1 = G_2(x_0)p(x_0)V$.

*Proof of Lemma* 4.2. Let

$$W_n = \sum_{i=1}^{n} W_{n,i} = \sum_{i=1}^{n} \psi(u_i) K\left(\frac{X_i - x_0}{h}\right) \left(\begin{array}{c} 1 \\ \frac{X_i - x_0}{h} \end{array}\right),$$

then by Assumption 4, we have $EW_n = o(1)$, and

$$\text{Var } W_n = \text{Var}\left(\sum_{i=1}^{n} W_{n,i}\right) = nEW_{n,1}^2 + 2\sum_{j=2}^{n}(n-j+1)\text{Cov}(W_{n,1}, W_{n,j}).$$

Using the same lines of arguments as in Lemma 4.1, we have

$$\text{Var } W_n = nhG_2(x)p(x_0)V(1 + o(1)). \qquad \square$$

Next, we will prove the asymptotic normality of $1/\sqrt{nh}W_n$, and this can be shown by using similar methods as in Cai and Ould-Said [**8,** Theorem 2]. This completes Lemma 4.2.

*Proof of Theorem* 3.1. We prove the consistency of the local $M$-estimators of $\sigma^2(x)$ and $\dot{\sigma}^2(x)$. Let

$$r = (a, hb)^T, \quad r_0 = (\sigma^2(x_0), h\dot{\sigma}^2(x_0))^T,$$

$$r_i = (r - r_0)^T \left(\begin{array}{c} 1 \\ \frac{X_i - x_0}{h} \end{array}\right),$$

and

$$L_n(r) = \sum_{i=1}^{n} \rho\left((Y_i - \widehat{m}(X_i))^2 - a - b(X_i - x_0)\right) K\left(\frac{X_i - x_0}{h}\right).$$

Then we have

$$r_i = (r - r_0)^T \left(\begin{array}{c} 1 \\ \frac{X_i - x_0}{h} \end{array}\right)$$

$$= (a - \sigma^2(x_0), hb - h\dot{\sigma}^2(x_0)) \left(\begin{array}{c} 1 \\ \frac{X_i - x_0}{h} \end{array}\right)$$

$$= a - \sigma^2(x_0) + (hb - h\dot{\sigma}^2(x_0))\frac{X_i - x_0}{h}$$

$$= a - \sigma^2(x_0) + (b - \dot{\sigma}^2(x_0))(X_i - x_0)$$

$$= a + b(X_i - x_0) - \sigma^2(x_0) - \dot{\sigma}^2(x_0)(X_i - x_0)$$

$$= a + b(X_i - x_0) + R(X_i) - \sigma^2(X_i)$$

$$= a + b(X_i - x_0) + R(X_i) - [(Y_i - \widehat{m}(X_i))^2 - u_i].$$

Let $S_\delta$ be the circle centered at $r_0$ with radius $\delta$. We will show that, for any sufficiently small $\delta$,

(4.1) $$\lim_{n \to \infty} P \left\{ \inf_{r \in S_\delta} L_n(r) > L_n(r_0) \right\} = 1.$$

In fact, for $r \in S_\delta$, we have

$$L_n(r) - L_n(r_0)$$

$$= \sum_{i=1}^{n} \rho \left( (Y_i - \widehat{m}(X_i))^2 - a - b(X_i - x_0) \right) K \left( \frac{X_i - x_0}{h} \right)$$

$$- \sum_{i=1}^{n} \rho \left( (Y_i - \widehat{m}(X_i))^2 - \sigma^2(x_0) - \dot{\sigma}^2(x_0)(X_i - x_0) \right) K \left( \frac{X_i - x_0}{h} \right)$$

$$= \sum_{i=1}^{n} K \left( \frac{X_i - x_0}{h} \right) [\rho(u_i + R(X_i) - r_i) - \rho(u_i + R(X_i))]$$

$$= \sum_{i=1}^{n} K \left( \frac{X_i - x_0}{h} \right) \int_{u_i + R(X_i)}^{u_i + R(X_i) - r_i} \psi(t) \, dt$$

$$= \sum_{i=1}^{n} K \left( \frac{X_i - x_0}{h} \right) \int_{u_i + R(X_i)}^{u_i + R(X_i) - r_i} \psi(u_i) \, dt$$

$$+ \sum_{i=1}^{n} K \left( \frac{X_i - x_0}{h} \right) \int_{u_i + R(X_i)}^{u_i + R(X_i) - r_i} \psi'(u_i)(t - u_i) \, dt$$

$$+ \sum_{i=1}^{n} K \left( \frac{X_i - x_0}{h} \right) \int_{u_i + R(X_i)}^{u_i + R(X_i) - r_i} [\psi(t) - \psi(u_i) - \psi'(u_i)(t - u_i)] \, dt$$

$$:= L_{n1} + L_{n2} + L_{n3}.$$

Next, we will show that

(4.2) $$L_{n1} = o_p(nh\delta),$$

(4.3) $$L_{n2} = \frac{nh}{2}(r - r_0)^T G_1(x_0) p(x_0) U(1 + o_p(1))(r - r_0) + O_p(nh^3\delta),$$

(4.4) $$L_{n3} = o_p(nh\delta^2).$$

For (4.2), we have

$$
\begin{aligned}
L_{n1} &= \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \int_{u_i + R(X_i)}^{u_i + R(X_i) - r_i} \psi(u_i)\, dt \\
&= \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \psi(u_i)(-r_i) \\
&= -(r - r_0)^T \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \psi(u_i) \begin{pmatrix} 1 \\ \frac{X_i - x_0}{h} \end{pmatrix} = -(r - r_0)^T W_n,
\end{aligned}
$$

where

$$
W_n = \begin{pmatrix} \sum\limits_{i=1}^{n} \psi(u_i) K\left(\frac{X_i - x_0}{h}\right) \\ \sum\limits_{i=1}^{n} \psi(u_i) K\left(\frac{X_i - x_0}{h}\right) \frac{X_i - x_0}{h} \end{pmatrix}.
$$

By Assumption 4 and Lemma 4.2, we have $EW_n = o(1)$, and

$$\operatorname{Var} W_n = nhG_2(x)p(x_0)V(1 + o(1)),$$

which implies that (4.2) holds.

For (4.3), we have

$$
\begin{aligned}
L_{n2} &= \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \int_{u_i + R(X_i)}^{u_i + R(X_i) - r_i} [\psi'(u_i)(t - u_i)]\, dt \\
&= \frac{1}{2} \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \psi'(u_i)(r_i^2 - 2R(X_i)r_i) \\
&= \frac{1}{2} \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \psi'(u_i)(r - r_0)^T \begin{pmatrix} 1 & \frac{X_i - x_0}{h} \\ \frac{X_i - x_0}{h} & \frac{(X_i - x_0)^2}{h^2} \end{pmatrix} (r - r_0) \\
&\quad - \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \psi'(u_i) R(X_i) r_i \\
&:= L_{n21} + L_{n22}.
\end{aligned}
$$

From Lemma 4.1 with $l = 0, l = 1$ and $l = 2$, respectively, we have

$$
\begin{aligned}
L_{n21} &= \frac{1}{2} \sum_{i=1}^{n} K\Big(\frac{X_i - x_0}{h}\Big) \psi'(u_i)(r - r_0)^T \begin{pmatrix} 1 & \frac{X_i - x_0}{h} \\ \frac{X_i - x_0}{h} & \frac{(X_i - x_0)^2}{h^2} \end{pmatrix} (r - r_0) \\
&= \frac{nh}{2}(r - r_0)^T G_1(x_0)p(x_0) \begin{pmatrix} K_0 & K_1 \\ K_1 & K_2 \end{pmatrix} (1 + o_p(1))(r - r_0) \\
&= \frac{nh}{2}(r - r_0)^T G_1(x_0)p(x_0)U(1 + o_p(1))(r - r_0)
\end{aligned}
$$

and

$$
\begin{aligned}
L_{n22} &= -\sum_{i=1}^{n} K\Big(\frac{X_i - x_0}{h}\Big) \psi'(u_i)R(X_i)r_i \\
&= -(r - r_0)^T \sum_{i=1}^{n} K\Big(\frac{X_i - x_0}{h}\Big) \psi'(u_i)R(X_i) \begin{pmatrix} 1 \\ \frac{X_i - x_0}{h} \end{pmatrix} \\
&= -\frac{nh^3}{2}(r - r_0)^T G_1(x_0)\ddot{\sigma}^2(x_0)p(x_0) \begin{pmatrix} K_2 \\ K_3 \end{pmatrix} (1 + o_p(1)) \\
&= O_p(nh^3\delta).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
L_{n2} &= L_{n21} + L_{n22} \\
&= \frac{nh}{2}(r - r_0)^T G_1(x_0)p(x_0)U(1 + o_p(1))(r - r_0) + O_p(nh^3\delta).
\end{aligned}
$$

For (4.4), we have

$$
\begin{aligned}
L_{n3} &= \sum_{i=1}^{n} K\Big(\frac{X_i - x_0}{h}\Big) \int_{u_i + R(X_i)}^{u_i + R(X_i) - r_i} [\psi(t) - \psi(u_i) - \psi'(u_i)(t - u_i)]\, dt \\
&= \sum_{i=1}^{n} K\Big(\frac{X_i - x_0}{h}\Big) \int_{R(X_i)}^{R(X_i) - r_i} [\psi(t + u_i) - \psi(u_i) - \psi'(u_i)t]\, dt \\
&= \sum_{i=1}^{n} K\Big(\frac{X_i - x_0}{h}\Big) [\psi(z_i + u_i) - \psi(u_i) - \psi'(u_i)z_i](-r_i)
\end{aligned}
$$

$$= -(r - r_0)^T \sum_{i=1}^{n}$$

$$\cdot K\left(\frac{X_i - x_0}{h}\right)[\psi(z_i + u_i) - \psi(u_i) - \psi'(u_i)z_i]\left(\begin{array}{c} 1 \\ \frac{X_i - x_0}{h} \end{array}\right),$$

where the second-to-last equality follows from the integral mean value theorem, and $z_i$ lies between $R(X_i)$ and $R(X_i) - r_i$, for $i = 1, 2, \ldots, n$. Since, for $|X_i - x_0| \le h$, we have

$$(4.5) \qquad \max_i |z_i| \le \max_i |R(X_i)| + \left|(r - r_0)^T\left(\begin{array}{c} 1 \\ \frac{X_i - x_0}{h} \end{array}\right)\right|$$

$$\le \max_i |R(X_i)| + 2\delta,$$

and, by Taylor's expansion,

$$(4.6) \qquad \max_i |R(X_i)| = \max_i \left|\sigma^2(X_i) - \sigma^2(x_0) - \dot\sigma^2(x_0)(X_i - x_0)\right|$$

$$= \max_i \left|\frac{1}{2}\ddot\sigma^2(\xi_i)(X_i - x_0)^2\right| \le O_p(h^2),$$

where $\xi_i$ lies between $X_i$ and $x_0$, for $i = 1, 2, \ldots, n$.

For any given $\eta > 0$, let $D_\eta = \{(\delta_1, \delta_2, \ldots, \delta_n)^T : |\delta_i| \le \eta, \forall i \le n\}$, by Assumption 5 and $|X_i - x_0| \le h$, we have

$$E\left[\sup_{D_\eta} \left|\sum_{i=1}^{n} [\psi(\delta_i + u_i) - \psi(u_i) - \psi'(u_i)\delta_i]K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^l\right|\right]$$

$$\le E\left[\sum_{i=1}^{n} \sup_{D_\eta} |\psi(\delta_i + u_i) - \psi(u_i) - \psi'(u_i)\delta_i|K\left(\frac{X_i - x_0}{h}\right)|X_i - x_0|^l\right]$$

$$\le a_\eta \delta E\left[\sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right)|X_i - x_0|^l\right] \le b_\eta \delta,$$

where $a_\eta$ and $b_\eta$ are two sequences of positive numbers, tending to zero as $\eta \to 0$. Therefore, by (4.5) and (4.6), we have

$$\sum_{i=1}^{n} [\psi(z_i + u_i) - \psi(u_i) - \psi'(u_i)z_i]K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^l = o_p(nh^{l+1}\delta),$$

which implies that (4.4) holds.

Let $\lambda$ be the smallest eigenvalue of the positive definite matrix $U$. Then, for any $r \in S_\delta$, we have

$$
\begin{aligned}
L_n(r) - L_n(r_0) &= L_{n1} + L_{n2} + L_{n3} \\
&= \frac{nh}{2} G_1(x_0) p(x_0) (r - r_0)^T U (r - r_0)(1 + o_p(1)) \\
&\quad + O_p(nh^3\delta) \\
&\geq \frac{nh}{2} G_1(x_0) p(x_0) \lambda \delta^2 (1 + o_p(1)) + O_p(nh^3\delta).
\end{aligned}
$$

So, as $n \to \infty$, we have

$$
P\Big\{ \inf_{r \in S_\delta} L_n(r) - L_n(r_0) > \frac{nh}{2} G_1(x_0) p(x_0) \lambda \delta^2 > 0 \Big\} \longrightarrow 1,
$$

which implies that (4.1) holds. From (4.1), we know that $L_n(r)$ has a local minimum in the interior of $S_\delta$, so there exist solutions to equations (2.3). Let $(\widehat{\sigma}^2(x_0), h\widehat{\dot{\sigma}}^2(x_0))^T$ be the closest solutions to $r_0 = (\sigma^2(x_0), h\dot{\sigma}^2(x_0))^T$. Then

$$
\lim_{n \to \infty} P\Big\{ (\widehat{\sigma}^2(x_0) - \sigma^2(x_0))^2 + h^2(\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0))^2 \leq \delta^2 \Big\} = 1,
$$

which implies the consistency of the local *M*-estimators of $\sigma^2(x)$ and $\dot{\sigma}^2(x)$. This completes the proof. $\qquad\square$

*Proof of Theorem* 3.2. We prove the asymptotic normality of the local *M*-estimators of $\sigma^2(x)$ and $\dot{\sigma}^2(x)$. Let

$$
(4.7) \quad \widehat{\eta}_i = R(X_i) - (\widehat{\sigma}^2(x_0) - \sigma^2(x_0)) - (\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0))(X_i - x_0).
$$

Then, we have

$$
\begin{aligned}
[Y_i - \widehat{m}(X_i)]^2 &= \sigma^2(X_i) + u_i \\
&= u_i + \sigma^2(X_i) - \sigma^2(x_0) - \dot{\sigma}^2(x_0)(X_i - x_0) + \sigma^2(x_0) \\
&\quad + \dot{\sigma}^2(x_0)(X_i - x_0) \\
&= u_i + R(X_i) + \widehat{\sigma}^2(x_0) + \widehat{\dot{\sigma}}^2(x_0)(X_i - x_0) + \widehat{\eta}_i - R(X_i) \\
&= \widehat{\sigma}^2(x_0) + \widehat{\dot{\sigma}}^2(x_0)(X_i - x_0) + u_i + \widehat{\eta}_i.
\end{aligned}
$$

Therefore, by (2.3), we have

$$(4.8) \qquad \sum_{i=1}^{n} \psi(u_i + \widehat{\eta}_i) K\left(\frac{X_i - x_0}{h}\right) \left(\begin{array}{c} 1 \\ \frac{X_i - x}{h} \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \end{array}\right).$$

Let

$$T_{n1} = \sum_{i=1}^{n} \psi(u_i) K\left(\frac{X_i - x_0}{h}\right) \left(\begin{array}{c} 1 \\ \frac{X_i - x}{h} \end{array}\right) = W_n,$$

$$T_{n2} = \sum_{i=1}^{n} \psi'(u_i) \widehat{\eta}_i K\left(\frac{X_i - x_0}{h}\right) \left(\begin{array}{c} 1 \\ \frac{X_i - x}{h} \end{array}\right),$$

$$T_{n3} = \sum_{i=1}^{n} [\psi(u_i + \widehat{\eta}_i) - \psi(u_i) - \psi'(u_i)\widehat{\eta}_i] K\left(\frac{X_i - x_0}{h}\right) \left(\begin{array}{c} 1 \\ \frac{X_i - x}{h}. \end{array}\right)$$

Then, by (4.8), we have $T_{n1} + T_{n2} + T_{n3} = 0$. And, by (4.7), we have

$$T_{n2} = \sum_{i=1}^{n} \psi'(u_i) R(X_i) K\left(\frac{X_i - x_0}{h}\right) \left(\begin{array}{c} 1 \\ \frac{X_i - x}{h} \end{array}\right) - \sum_{i=1}^{n} \psi'(u_i) K\left(\frac{X_i - x_0}{h}\right)$$

$$\cdot \left(\begin{array}{c} (\widehat{\sigma}^2(x_0) - \sigma^2(x_0)) + (\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0))(X_i - x_0) \\ \frac{X_i - x}{h}[(\widehat{\sigma}^2(x_0) - \sigma^2(x_0)) + (\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0))(X_i - x_0)] \end{array}\right)$$

$$= \sum_{i=1}^{n} \psi'(u_i) R(X_i) K\left(\frac{X_i - x_0}{h}\right) \left(\begin{array}{c} 1 \\ \frac{X_i - x}{h} \end{array}\right)$$

$$- \sum_{i=1}^{n} \psi'(u_i) K\left(\frac{X_i - x_0}{h}\right) \left(\begin{array}{cc} 1 & \frac{X_i - x}{h} \\ \frac{X_i - x}{h} & \frac{(X_i - x)^2}{h^2} \end{array}\right)$$

$$\cdot \left(\begin{array}{c} \widehat{\sigma}^2(x_0) - \sigma^2(x_0) \\ h(\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0)) \end{array}\right)$$

$$= \frac{nh^3}{2} G_1(x_0) \ddot{\sigma}^2(x_0) p(x_0) \left(\begin{array}{c} K_2 \\ K_3 \end{array}\right)(1 + o_p(1))$$

$$- nh G_1(x_0) p(x_0) \left(\begin{array}{cc} K_0 & K_1 \\ K_1 & K_2 \end{array}\right)(1 + o_p(1))$$

$$\cdot \left(\begin{array}{c} \widehat{\sigma}^2(x_0) - \sigma^2(x_0) \\ h(\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0)) \end{array}\right)$$

$$= \frac{nh^3 G_1(x_0) \ddot{\sigma}^2(x_0) p(x_0)}{2} A(1 + o_p(1)) - nh G_1(x_0) p(x_0) U(1 + o_p(1))$$

$$\cdot \left( \begin{array}{c} \widehat{\sigma}^2(x_0) - \sigma^2(x_0) \\ h(\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0)) \end{array} \right)$$

$$:= T_{n21} + T_{n22},$$

where the third equality follows from Lemma 4.1.

Noting that, for $|X_i - x_0| \le h$, we have

$$\sup_i |\widehat{\eta}_i| = \sup_i \left| R(X_i) - (\widehat{\sigma}^2(x_0) - \sigma^2(x_0)) - (\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2 x_0))(X_i - x_0) \right|$$

$$\le \sup_i |R(X_i)| + \left| \widehat{\sigma}^2(x_0) - \sigma^2(x_0) \right| + h \left| \widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0) \right|$$

$$= O_p(h^2 + (\widehat{\sigma}^2(x_0) - \sigma^2(x_0)) + h(\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0))) = o_p(1),$$

where the last equality follows from the consistency of $(\widehat{\sigma}^2(x_0), h\widehat{\dot{\sigma}}^2(x_0))$. Then, by Assumption 5 and the same argument as that in Theorem 3.1, we have

$$T_{n3} = \sum_{i=1}^n [\psi(u_i + \widehat{\eta}_i) - \psi(u_i) - \psi'(u_i)\widehat{\eta}_i] K \left( \frac{X_i - x_0}{h} \right) \left( \begin{array}{c} 1 \\ \frac{X_i - x}{h} \end{array} \right)$$

$$= o_p(nh)[h^2 + (\widehat{\sigma}^2(x_0) - \sigma^2(x_0)) + h(\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0))] = o_p(T_{n22}).$$

Therefore, by $T_{n1} + T_{n2} + T_{n3} = 0$, we have

$$\left( \begin{array}{c} \widehat{\sigma}^2(x_0) - \sigma^2(x_0) \\ h(\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0)) \end{array} \right) = \frac{1}{nh} G_1^{-1}(x_0) p^{-1}(x_0) U^{-1}(1 + o_p(1)) W_n$$

$$+ \frac{h^2}{2} \ddot{\sigma}^2(x_0) U^{-1} A(1 + o_p(1)).$$

It follows that

$$\sqrt{nh} \left[ \left( \begin{array}{c} \widehat{\sigma}^2(x_0) - \sigma^2(x_0) \\ h(\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0)) \end{array} \right) - \frac{h^2 \ddot{\sigma}^2(x_0)}{2} U^{-1} A(1 + o_p(1)) \right]$$

$$= G_1^{-1}(x_0) p^{-1}(x_0) U^{-1}(1 + o_p(1)) \frac{1}{\sqrt{nh}} W_n.$$

So, by Lemma 4.2 and Slutsky's theorem, we have

$$\sqrt{nh} \left[ \left( \begin{array}{c} \widehat{\sigma}^2(x_0) - \sigma^2(x_0) \\ h(\widehat{\dot{\sigma}}^2(x_0) - \dot{\sigma}^2(x_0)) \end{array} \right) - \frac{h^2 \ddot{\sigma}^2(x_0)}{2} U^{-1} A \right]$$

$$\xrightarrow{D} G_1^{-1}(x_0)p^{-1}(x_0)U^{-1}N(0,\Sigma_1)$$
$$= N\left(0, \frac{G_2(x_0)}{G_1^2(x_0)p(x_0)}U^{-1}VU^{-1}\right) = N(0,\Sigma).$$

This completes the proof.                                              □

**5. Conclusions.** This paper concentrates exclusively on the uni-variate predictor; the robust approach considered here is available if $X$ is multivariate dimensional, and the method of proof of consistency and asymptotic normality is similar to the one-dimensional case, but more cumbersome. Chen and Zhang [**11**] considered the case of the two-dimensional predictor; they investigated the local linear $M$-estimation for regression function in a fixed-design model when the errors are from a strongly mixing random field. Moreover, our robust method is also available in the case of functional data. Functional data are more and more frequently involved in statistical problems, so many statistical methods in this special framework have been proposed during the last few years, particularly with the book by Ferraty and Vieu [**23**], which presented in an original way new nonparametric statistical methods for functional data analysis. In addition, Ferraty et al. [**22**] proposed a functional nonparametric model for time series prediction, and the paper used an infinite number of past values as predictors. Barrientos-Marin et al. [**3**] proposed a new local modelling approach for dealing with the nonparametric regression of a scalar response on a functional variable, and obtained an explicit expression of a kernel-type estimator which makes its computation easy and fast while keeping good predic-tive performance. Moreover, a functional approach in the same spirit as local linear ideas in nonparametric regression was introduced in the same paper. Delsol [**16**] developed some nonparametric methods for $\alpha$-mixing functional random variables.

To employ the local $M$-estimator in the paper, one needs to choose the kernel function and the bandwidth. It is well known, both empir-ically and theoretically, that the choice of kernel functions is not very important. Thus, a large bandwidth produces an oversmooth estimate, that is, the estimate can create large biases when a large bandwidth is used. While a small bandwidth produces undersmooth estimate, that is, there are not many local data points available to reduce the variance of the estimate when a small bandwidth is used. The usual important

techniques for selecting the bandwidth are cross-validation (CV) and plug-in bandwidth selectors. The cross-validation method uses ideas of Allen [1] and Stone [40], it is very useful for assessing the performance of an estimator via estimating its prediction error, but, it is computationally intensive. In order to facilitate the computation, Wahba [43] and Craven and Wahba [13] proposed the generalized cross-validation (GCV). The GCV method also has its disadvantages; one is its inherent variability, another is that it cannot be directly applied to select bandwidths for estimating derivative curves. The plug-in method avoids these problems as it is considered to be a conceptually simple technique, with theoretical justification and a good empirical performance method. The plug-in method was proposed by Ruppert et al. [38]; it is an asymptotic substitution method. More details on bandwidth selections can be found in Fan and Gijbels [18] and Fan and Yao [20].

The volatility function and the nonparametric local $M$-estimators introduced can be used in a variety of applications. For example, Poon and Granger [36] provided comprehensive coverage of the status of volatility forecasting. Patton [35] derived explicit analytical results for the distortions that may arise when some common loss functions are employed, provided necessary and sufficient conditions on the functional form of the loss function, and volatility forecasting was the focus as a specific case of the more general problem of latent variable forecasting in the same paper.

In this paper, a local $M$-estimation for the conditional variance function in heteroscedastic regression models under stationary $\alpha$-mixing dependent samples is developed. Long memory structure is a more general dependence structure than mixing; long memory process constitutes a broad class of models for stationary and nonstationary time series data in economics, finance, and other fields, so it has become a rapidly developing subject. Estéveza and Vieu [17] considered a general form of long dependence processes, studied asymptotic behavior of the nonparametric kernel estimators under this dependence structure and obtained the rates of weak consistency and of mean squared consistency. The roust method under stationary $\alpha$-mixing dependent samples developed in our paper can be used to establish roust estimator under long memory structure if one can get similar results with Lemma 4.1 and Lemma 4.2 in this paper under some mild conditions.

## REFERENCES

**1**. D.M. Allen, *The relationship between variable and data augmentation and a method of prediction*, Technometrics **16** (1974), 125–127.

**2**. T.G. Andersen and J. Lund, *Estimating continuous time stochastic volatility models of the short term interest rate*, J. Econometrics **77** (1997), 343–377.

**3**. J. Barrientos-Marin, F. Ferraty and P. Vieu, *Locally modelled regression and functional data*, J. Nonparam. Stat. **22** (2010), 617–632.

**4**. A.M. Bianco, G. Boente and S. Sombielle, *Robust estimation for nonparametric generalized regression*, Stat. Prob. Lett. **81** (2011), 1986–1994.

**5**. G. Boente, X. He and J. Zhou, *Robust estimates in generalized partially linear models*, The Annals Stat. **34** (2006), 2856–2878.

**6**. G. Boente, M. Ruiz and R.H. Zamar, *On a robust local estimator for the scale function in heteroscedastic nonparametric regression*, Stat. Prob. Lett. **80** (2010), 1185–1195.

**7**. G.E.P. Box, *Signal-to-noise ratios, performance criteria, and transformations*, Technometrics **30** (1988), 1–17.

**8**. Z. Cai and E. Ould-Said, *Local M-estimator for nonparametric time series*, Stat. Prob. Lett. **65** (2003), 433–449.

**9**. R.J. Carroll and D. Ruppert, *Transformations and weighting in regression*, Chapman & Hall, London, 1988.

**10**. N.H. Chan, L. Peng and D.B. Zhang, *Empirical-likelihood-based confidence intervals for conditional variance in heteroskedastic regression models*, Econo. Theor. **27** (2011), 154–177.

**11**. J. Chen and L.X. Zhang, *Local linear M-estimation for spatial processes in fixed-design models*, Metrika **71** (2010), 319–340.

**12**. L. Chen, M. Cheng and L. Peng, *Conditional variance estimation in heteroscedastic regression models*, J. Stat. Plan. Infer. **139** (2009), 236–245.

**13**. P. Craven and G. Wahba, *Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation*, Numer. Math. **31** (1979), 377–403.

**14**. R.A. Davis, K. Knight and J. Liuc, *M-estimation for autoregressions with infinite variance*, Stoch. Proc. Appl. **40** (1992), 145–180.

**15**. M. Delecroix, M. Hristache and V. Patilea, *On semiparametric M-estimation in single-index regression*, J. Stat. Plan. Infer. **136** (2006), 730–769.

**16**. L. Delsol, *Nonparametric methods for $\alpha$-mixing functional random variables. The Oxford handbook of functional data analysis*, Oxford University Press, Oxford, 2011.

**17**. G. Estéveza and P. Vieu, *Nonparametric estimation under long memory dependence*, J. Nonparam. Stat. **15** (2003), 535–551.

**18**. J. Fan and I. Gijbels, *Local polynomial modeling and its applications*, Chapman & Hall, London, 1996.

**19**. J. Fan and J. Jiang, *Variable bandwidth and one-step local M-estimator*, Sci. China **43** (2000), 65–81.

**20**. J. Fan and Q. Yao, *Efficient estimation of conditional variance functions in stochastic regression*, Biometrika **85** (1998), 645–660.

**21**. ――――, *Nonlinear time series*: *Nonparametric and parametric methods*, Springer-Verlag, New York, 2003.

**22**. F. Ferraty, A. Goia and P. Vieu, *Functional nonparametric model for time series*: *A fractal approach for dimension reduction*, Test **11** (2002), 317–344.

**23**. F. Ferraty and P. Vieu, *Nonparametric functional data analysis*: *Theory and practice*, Springer, New York, 2006.

**24**. L. Györfi, W. Härdle, P. Sarda and P. Vieu, *Nonparametric curve estimation from time series*, Lect. Notes Stat. **60** (1989), Springer-Verlag, Berlin.

**25**. P. Hall and M.C. Jones, *Adaptive M-estimation in nonparametric regression*, Ann. Stat. **18** (1990), 1712–1728.

**26**. B.E. Hansen, *Regression with nonstationary volatility*, Econometrica **63** (1995), 1113–1132.

**27**. W. Härdle, *Robust regression function estimation*, J. Multivar. Anal. **14** (1984), 169–180.

**28**. ――――, *Asymptotic maximal deviation of M-smoothers*, J. Multivar. Anal. **29** (1989), 163–179.

**29**. W. Härdle and A. Tsybakov, *Local polynomial estimators of the volatility function in nonparametric autoregression*, J. Econom. **81** (1997), 223–242.

**30**. U. Holst, O. Hössjer, C. Björklund, P. Ragnarson and H. Edner, *Locally weighted least squares kernel regression and statistical evaluation of LIDAR measurements*, Environmetrics **7** (1996), 401–416.

**31**. J.C. Jiang and Y.P. Mack, *Robust local polynomial regression for dependent data*, Stat. Sinica **11** (2001), 705–722.

**32**. R. Kulik and C. Wichelhaus, *Nonparametric conditional variance and error density estimation in regression models with dependent errors and predictors*, Electr. J. Stat. **5** (2011), 856–898.

**33**. H.Y. Liang, *Asymptotic normality for regression function estimate under truncation and $\alpha$-mixing conditions*, Comm. Stat. Theor. Meth. **40** (2011), 1999–2021.

**34**. E. Masry and J. Fan, *Local polynomial estimation of regression functions for mixing processes*, Scandinavian J. Stat. **24** (1997), 165–179.

**35**. A.J. Patton, *Volatility forecast comparison using imperfect volatility proxies*, J. Econometrics **160** (2011), 246–256.

**36**. S.-H. Poon and C.W.J. Granger, *Forecasting volatility in financial markets*: *A review*, J. Econ. Lit. **41** (2003), 478–539.

**37**. G.G. Roussas, *Nonparametric regression estimation under mixing conditions*, Stoch. Proc. Appl. **36** (1990), 107–116.

**38**. D. Ruppert, S.J. Sheather and M.P. Wand, *An effective bandwidth selector for local least squares regression*, J. Amer. Stat. Assoc. **90** (1995), 1257–1270.

**39**. D. Ruppert, M.P. Wand, U. Holst and O. Hössjer, *Local polynomial variance function estimation*, Technometrics **39** (1997), 262–273.

**40**. M. Stone, *Cross-validatory choice and assessment of statistical predictions* (*with discussion*), J. Roy. Stat. Soc. **36** (1974), 111–147.

**41**. D. Tjøstheim, *Nonlinear time series*: *A selective review*, Scandinvian J. Stat. **21** (1994), 97–130.

**42**. J.M. Vilar-Fernández and M. Francisco-Fernández, *Nonparametric estimation of the conditional variance function with correlated errors*, J. Nonparam. Stat. **18** (2006), 375–391.

**43**. G. Wahba, *A survey of some smoothing problems and the method of generalized cross-validation for solving them*, in *Applications of statistics*, P.R. Krishnaiah, ed., North-Holland, Amsterdam, 1977.

**44**. K.L. Xu and P.C.B. Phillips, *Tilted nonparametric estimation of volatility functions with empirical applications*, J. Business Econ. Stat. **29** (2011), 518–528.

**45**. K. Yoshihara, *Weakly dependent stochastic sequences and their applications*, Volume IV, *Curve estimation based on weakly dependent data*, Sanseido, Tokyo, 1994.

School of Science, Jiangxi University of Science and Technology, Ganzhou 341000, China
**Email address**: **yunyanwang0329@yahoo.com**

School of Science, Jiangxi University of Science and Technology, Ganzhou 341000, China
**Email address**: **mttang_csu@yahoo.com**