Nonparametric Estimation of Variance Function for Functional Data

Heng Lian

Division of Mathematical Sciences,
School of Physical and Mathematical Sciences,
Nanyang Technological University, Singapore 637371.

Abstract: This article investigates nonparametric estimation of variance functions for functional data when the mean function is unknown. We obtain asymptotic results for the kernel estimator based on squared residuals. Similar to the finite dimensional case, our asymptotic result shows the smoothness of the unknown mean function has an effect on the rate of convergence. Our simulation studies demonstrate that estimator based on residuals performs much better than that based on conditional second moment of the responses.

Key words and phrases: Functional data, Kernel regression, Rates of convergence, Variance estimation.

1 Introduction

Recently, there has been increased interest in the statistical modelling of functional data. In many experiments, functional data appear as the basic unit of observations. As a natural extension of the multivariate data analysis, functional data analysis provides valuable insights into these problems. Compared with the discrete multivariate analysis, functional analysis takes into account the smooth-
ness of the high dimensional covariates, and often suggests new approaches to the problems that have not been discovered before. Even for nonfunctional data, the functional approach can often offer new perspectives on the old problem.

The literature contains an impressive range of functional analysis tools for various problems including exploratory functional principal component analysis, canonical correlation analysis, classification and regression. Two major approaches exist. The more traditional approach, carefully documented in the monograph Ramsay & Silverman (2005), typically starts by representing functional data by an expansion with respect to a certain basis, and subsequent inferences are carried out on the coefficients. The most commonly utilized basis include B-spline basis for nonperiodic data and Fourier basis for periodic data. Another line of work by the French school Ferraty & Vieu (2002), taking a non-parametric point of view, extends the traditional nonparametric techniques, most notably the kernel estimate, to the functional case. Some theoretical results are also obtained as a generalization of the convergence properties of the classical kernel estimate.

The functional nonparametric regression model, introduced in Ferraty & Vieu (2002), is defined as

\[ Y_i = m(X_i) + \sqrt{v(X_i)}\epsilon_i, \]  

where we emphasized the heterogeneity of the regression model which is the focus of this article. We assume that \( \epsilon_i \)'s are random variables with \( E(\epsilon_i|X_i) = 0 \) and \( Var(\epsilon_i|X_i) = v(X_i) \). The covariates \( X_i \) are assumed to belong to some semi-metric vectorial space \( \mathcal{H} \) determined by the semi-metric \( d(\ldots) \). Unlike many previous nonparametric functional regression studies Ferraty & Vieu (2004); Masry (2005); Lian (2007) which focused on estimating the mean function \( m \), here we
are interested in estimating $v$ when $m$ is unknown, and thus the mean function only plays the role of a nuisance parameter.

Variance function estimation has received much attention since the 1980's when it was required for confidence interval construction for the mean function, and Muller & Stadtmuller (1987) discussed some utility of it in obtaining more efficient estimators of the mean function. There are two main approaches to variance function estimation. In Muller & Stadtmuller (1987, 1993), the variance function was estimated directly from local contrasts of the responses. More recently, Brown & Levine (2007); Wang et al. (2008) has obtained minimax convergence rates based on local difference and Cai et al. (2009) further extended this to multivariate regression. These asymptotic theory were developed based on fixed covariates on a grid and it is not straightforward to extend to the case with random covariates. For our functional data analysis, it is not clear how to define a grid on the semi-metric space $H$. A different direction was taken in Hall & Carroll (1989), where the variance function was estimated by a weighted smoothing of squared residuals after a fit for the mean function was obtained. This approach was also considered in Fan & Yao (1998) using local polynomial regression. Finally, we mention the adaptive estimation of variance function in Cai & Wang (2008) by thresholding of wavelet coefficients.

In the following sections, we adapt the idea of variance estimation in non-parametric regression based on squared residuals to the functional setting. In Section 2, we review the functional nonparametric regression model in a semi-metric functional vectorial space. Then we introduce functional nonparametric variance estimation in this general setting and describe the asymptotic results for our kernel-type estimator. We also discuss the effect of unknown mean function
on the variance estimator and relate it to the finite-dimensional case. In Section 3, we carry out a simulation study to demonstrate that the residual-based estimator is more efficient than the estimator based on nonparametric regression on the squared responses. Finally, we illustrate the approach on the popular spectrometric data for predicting the fat content. The technical proofs for our asymptotic results are deferred to the appendix.

2 Nonparametric Functional Variance Estimation

In the functional nonparametric regression model (1) presented originally in Ferraty & Vieu (2002), the mean function is estimated by a kernel-type estimator

\[ \hat{m}(x) = \frac{\sum_{i=1}^{n} K(d_m(x, X_i)/h_m)Y_i}{\sum_{i=1}^{n} K(d_m(x, X_i)/h_m)}, \]

where \( Y_i \) is the real-valued responses and \( h_m \) is the bandwidth used for estimating the mean function. Note that we use \( d_m \) to denote the semi-metric for mean function estimation as we will use a different semi-metric for variance function estimation. Denote \( R(X, Y) = (Y - m(X))^2 \). Since under model (1), we have \( E(R(X, Y)|X) = v(X) \), a natural kernel-type estimator for \( v(x) \) (when the mean function is known) is

\[ \hat{v}(x) = \frac{\sum_{i=1}^{n} K(d_v(x, X_i)/h_v)R_i}{\sum_{i=1}^{n} K(d_v(x, X_i)/h_v)}, \quad (2) \]

where \( R_i = (Y_i - m(X_i))^2 \) and \( h_v \) is the chosen bandwidth of the kernel. Note that the semi-metric \( d_v \) used for estimating the variance function is in general different from the semi-metric \( d_m \) used in estimating the mean function. Using
different semi-metrics is important in some cases as demonstrated in our experiment with spectrometric data later. Although we could use different kernels for the mean and variance functions, we choose to use the same kernel here mainly for notational simplicity.

In practice, the mean function $m(\cdot)$ is typically unknown and a natural approach is to replace $m$ by the nonparametric estimator $\hat{m}$. Equivalently, we replace $R_i$ by $\hat{R}_i = (Y_i - \hat{m}(X_i))^2$ in (2).

Although only independent data are considered in our simulations and real data application, for our asymptotic analysis, we will present our results in a more general context by considering a strongly mixing sequence $\{(X_i, Y_i), i = 1, \ldots, n\}$. Our asymptotic result is stated for a fixed $x \in H$.

Following the notations in Ferraty & Vieu (2006), we have

\[
\begin{align*}
\Delta_m^i &= \frac{K(d_m(x, X_i)/h_m)}{EK(d_m(x, X_i)/h_m)} \\
\Delta_v^i &= \frac{K(d_v(x, X_i)/h_v)}{EK(d_v(x, X_i)/h_v)} \\
\end{align*}
\]

\[
\begin{align*}
r_1^m &= \sum_{i=1}^n \Delta_m^i / n \\
r_1^v &= \sum_{i=1}^n \Delta_v^i / n \\
r_2^m &= \sum_{i=1}^n Y_i \Delta_m^i / n \\
r_2^v &= \sum_{i=1}^n (Y_i - \hat{m}(X_i))^2 \Delta_v^i / n \\
\end{align*}
\]

so that $\hat{m}(x) = r_2^m / r_1^m$ and $\hat{v}(x) = r_2^v / r_1^v$. For notational simplicity, in the rest of the article, we denote $m_i = m(X_i), \hat{m}_i = \hat{m}(X_i), v_i = v(X_i), \hat{v}_i = \hat{v}(X_i)$. We also set $w_{ij} = K(d_m(X_i, X_j)/h_m) / \sum_k K(d_m(X_i, X_k)/h_m)$ so that $\hat{m}_i = \sum_j w_{ij} Y_j$. 5
Similar to Ferraty & Vieu (2004, 2006), the rate of convergence of \( \hat{v}(x) \) will critically depend on the quantities \( s^m_n \) and \( s^v_n \) defined by

\[
\begin{align*}
    s^m_n &= \max\{s^m_{n,1}, s^m_{n,2}, s^m_{n,3}, s^m_{n,4}\} \\
    s^v_n &= \max\{s^v_{n,1}, s^v_{n,2}, s^v_{n,3}, s^v_{n,4}\} \\
    s^m_{n,1} &= \sum_{i=1}^{n} \sum_{j=1}^{n} |\text{Cov}(\Delta^m_i, \Delta^m_j)| \\ 
    s^m_{n,2} &= \sum_{i=1}^{n} \sum_{j=1}^{n} \text{E}[\text{Cov}(\Delta^m_i \epsilon_i, \Delta^m_j \epsilon_j | X^n_1)] \\ 
    s^m_{n,3} &= \sum_{i=1}^{n} \sum_{j=1}^{n} |\text{Cov}(\Delta^m_i m_i, \Delta^m_j m_j)| \\ 
    s^m_{n,4} &= n \text{E} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta^v_i w_i \epsilon_i \epsilon_j \\ 
    s^v_{n,1} &= \sum_{i=1}^{n} \sum_{j=1}^{n} |\text{Cov}(\Delta^v_i, \Delta^v_j)| \\ 
    s^v_{n,2} &= \sum_{i=1}^{n} \sum_{j=1}^{n} \text{E}[\text{Cov}(\Delta^v_i \epsilon_i, \Delta^v_j \epsilon_j | X^n_1)] \\ 
    s^v_{n,3} &= \sum_{i=1}^{n} \sum_{j=1}^{n} |\text{Cov}(\Delta^v_i v_i, \Delta^v_j v_j)| \\ 
    s^v_{n,4} &= \sum_{i,j,k,l=1} \text{E}[\text{Cov}(\Delta^v_i \epsilon_i \epsilon_j, \Delta^v_k \epsilon_k \epsilon_l | X^n_1)]
\end{align*}
\]

where in some of the expressions above, the covariances are conditioned on observed covariates \( X^n_1 = \{X_1, \ldots, X_n\} \).

We follow Ferraty & Vieu (2006) and impose the following condition on the kernel function

\[
    K \text{ is supported on } [0, 1], \text{ bounded and bounded away from zero on } [0, 1].
\]

As the case for mean function estimation, we need the following regularity
conditions

\[ |m(x_1) - m(x_2)| \leq C d_m(x_1, x_2)^\alpha, \quad |v(x_1) - v(x_2)| \leq C d_v(x_1, x_2)^\beta, \quad \alpha > 0, \beta > 0. \] (12)

In Ferraty & Vieu (2004, 2006), moment conditions are directly assumed on the response \( Y \). We figure that it is more natural to impose the moment condition on the error

\[ \exists p \geq 4, E|\epsilon|^p < \infty. \] (13)

For uniform convergence over a compact neighborhood \( C \) of \( H \) containing \( x \) for the mean function, which is needed in the proof below, we assume that \( C \) can be written as, for any \( l > 0 \),

\[ C = \sum_{k=1}^r B(t_k, l), \text{ with } \tau l^a = C \text{ for some } a > 0, C > 0. \] (14)

This condition is exactly the same as that in Ferraty & Vieu (2008), and interested readers can find some related discussions there.

Now we are ready to state our main result:

**Theorem 1** Under the conditions (11)-(14), for a fixed \( x \in H \), we have

\[ |\hat{v}(x) - v(x)| = O \left( h_m^{2\alpha} + s_n^m \log n \frac{\log n}{n^2} + h_v^\beta + \sqrt{s_v^m \log n} n \right) \text{ in probability.} \]

**Remark 1** In Ferraty & Vieu (2004, 2006), the asymptotic results are stated as almost complete convergence, which is stronger than convergence in probability. The difficulty of proving stronger convergence for our variance estimator comes from the appearance of \( U \)-type-statistics in the expressions in the proof, thus we settle with weaker type of convergence here.
Remark 2 In Ferraty & Vieu (2006), it was discussed in details how $s_n^m$ depends on the following two quantities: 

$$\phi_m(h) := P(d_m(x,X) \leq h) \text{ and } \psi_m(h) = P(d_m(x-X_1) \leq h, d_m(x,X_2) \leq h)$$

for strongly mixing data sequences. Those results can be adapted for our purposes. For example, for the independent and identically distributed data, as shown in the appendix, we have 

$$s_n^m = O(n/\phi_m(h_m))$$

and 

$$s_n^v = O(n/\phi_v(h_v))$$

with 

$$\phi_v(h) = P(d_v(x,X) \leq h).$$

Thus in the i.i.d. case we have the following direct consequence.

**Corollary 1** Under the conditions (11)-(14), assuming in addition the data \{(X_i,Y_i), i = 1,\ldots,n\} are i.i.d. and the bandwidths are chosen such that $h_m \to 0, h_v \to 0, n\phi_m(h_m) \to \infty, n\phi_v(h_v) \to \infty$, we have

$$|\hat{v}(x) - v(x)| = O \left( h_m^{2\alpha} + \frac{\log n}{n\phi_m(h_m)} + h_v^\beta + \sqrt{\frac{\log n}{n\phi_v(h_v)}} \right)$$

in probability.

Remark 3 From the corollary, we can observe some interesting effect of unknown mean for variance function estimation. For simplicity and specificity, assume that $X$ is of fractal order $d$ with respect to both $d_m$ and $d_v$, i.e. 

$$\phi_m(h) \sim \phi_v(h) \sim h^d.$$ 

It was shown in Ferraty & Vieu (2006) Lemma 13.6 that if $\mathcal{H}$ is a separable Hilbert space with semi-metric defined by the projection onto the first $d$ elements of an orthonormal basis, then $\phi(h) \sim h^d$. This is also true for $d$-dimensional regression (i.e., $\mathcal{H} = R^d$). With $h_m \sim (\log n/n)^{1/(2\alpha+d)}, h_v \sim (\log n/n)^{1/(2\beta+d)},$ we obtain the rate of convergence $\max\{ (\log n/n)^{2\alpha/(2\alpha+d)}, (\log n/n)^{\beta/(2\beta+d)} \}.$

If $2\alpha/(2\alpha+d) \geq \beta/(2\beta+d)$, the rate becomes $\log n/n)^{\beta/(2\beta+d)}$. This rate is the same as the rate obtained when the mean function $m(.)$ is known. Thus we observe that when the mean function is smooth enough, it has no effect on variance function estimation, while its effect cannot be ignored for less smooth mean functions. In particular, it can be easily verified that $2\alpha/(2\alpha+d) \geq \beta/(2\beta+d)$
is true as soon as $\alpha \geq d/2$. This results is the same as what was observed in Hall & Carroll (1989) for one-dimensional regression where the author observed that the mean has no effect on variance function estimation as long as $\alpha \geq 1/2$ (the last sentence in section 2.2 of Hall & Carroll (1989)).

Remark 4 The simple relationship $v(x) = E(Y^2|X = x) - (E(Y|X = x))^2$ motivates the direct estimator based on estimating conditional expectation of squared responses and setting $\hat{v}(x) = \hat{s}(x) - \hat{m}^2(x)$ where $\hat{s}(x)$ is the nonparametric kernel-type estimate of $E(Y^2|X = x)$. This estimator is briefly mentioned in Ferraty et al. (2007). It can be shown that this estimator has the same convergence rate as above. However, in one-dimensional case, Fan & Yao (1998) pointed out the direct method can create a very large bias. The intuitive explanation provided for the large bias is that the direct estimator is obtained when replacing $\hat{R}_i = (Y_i - \hat{m}(X_i))^2$ in the residual-based method by $(Y_i - \hat{m}(x))^2$. This explanation also applies to our functional context. In our simulation study to be presented next, it is clear that the performance of the direct method is much worse than the residual based method.

3 Experiments

3.1 Simulation Study

We now consider in this section the finite sample performance of our variance estimator and also compare the results with the direct squared responses based method. We use three examples with different mean and variance functions to illustrate their performances. For each example, 100 simulations are performed
with \( n = 200 \) data points generated in each simulation. In all three examples, \( X_i \) is a random function supported on \([-1,1]\).

For the first example, we set
\[
m(x) = 0, \quad v(x) = \int_{-1}^{1} |\cos x(t)| \, dt,
\]
and the \( X_i \)'s are generated as realizations of Brownian Motion starting at time \( t = -1 \) with random start point \( x(-1) \) distributed as uniform random variables on \([-1,1]\). For the second example, we have
\[
m(x) = \int_{-1}^{1} tx(t) \, dt, \quad v(x) = \int_{-1}^{1} |t|x^2(t) \, dt,
\]
and the \( X_i \)'s are generated the same way as in the first example. For the third example, we follow Ferraty et al. (2007) and set
\[
m(x) = \int_{-1}^{1} |x'(t)|(1 - \cos(\pi t)) \, dt, \quad v(x) = \int_{-1}^{1} |x'(t)|(1 + \cos(\pi t)) \, dt.
\]
The random curves in this example are simulated from
\[
X(t) = \sin(\omega t) + (a + 2\pi)t + b, \quad \omega \sim Unif(0,2\pi), a,b \sim Unif(0,1).
\]
The simulations are performed in R with the publicly available npfda package \( \text{http://www.lsp.ups-tlse.fr/staph/npfda/} \). The default quadratic kernel is used in the implementation. The bandwidths \( h_m \) and \( h_v \) are chosen using cross-validation. The choice of semi-metric is in general a difficult problem. In our current simulations, their choices are suggested by our knowledge of the true mean and variance functions. Thus for the first two examples, we use \( d_m(x_1, x_2) = d_v(x_1, x_2) = \int_{-1}^{1} (x_1(t) - x_2(t))^2 \, dt \) and we use \( d_m(x_1, x_2) = d_v(x_1, x_2) = \int_{-1}^{1} (x_1'(t) - x_2'(t))^2 \, dt \) for the third example. Our simulation also
Table 1: Simulation results (MSE) for comparing two variance function estimators.

| Estimators            | Example 1 | Example 2 | Example 3 |
|-----------------------|-----------|-----------|-----------|
| residual based method | 0.10      | 0.27      | 4.37      |
| direct method         | 0.10      | 0.38      | 19.24     |

 shows that these choices of semi-metrics are the best among semi-metrics based on different orders of derivatives (results not represented here). For evaluation of performance, we adopt the discrete mean squared error

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{v}(X_i) - v(X_i))^2.
\]

We report in Table 1 the median MSE for variance function estimators based on 100 simulations. It is easily seen from the table that and the residual based two-step method performs much better than the direct method in terms of MSE, except in the first example with constant mean function, which is as expected.

### 3.2 Illustration with Chemometric Data

We illustrate our approach on the real chemometric dataset, which contains 215 spectra of light absorbance for meat samples as functions of the wavelengths. Because of the denseness of wavelengths at which the measurements are made, the subjects are naturally treated as continuous curves. This dataset has been previously used in nonparametric regression studies where the covariate is the spectra curve and the response is the percentage of fat content in the piece of meat \cite{FerratyVieu2002, FerratyVieu2006, FerratyVieu2007}. We will estimate the variance function for this regression problem. Previous study suggested that for
mean function estimation, taking as the semi-metric the $L_2$ distance between the second derivatives of the spectra gives favorable result, thus this semi-metric is used for mean function estimation. As in previous studies, we train on the first 150 spectra and use the rest as validation. We examine the estimation accuracy of the variance function for semi-metrics defined as $L_2$ distance between the curves using different orders of derivatives, measured as mean squared error

$$MSE = \frac{1}{65} \sum_{i=151}^{215} (\hat{R}_i^2 - \hat{v}(X_i))$$

and find that using $L_2$ distance between 1st derivatives gives the best result. The estimated variance function value and squared residuals for the validation data are shown in Fig. 1, giving a MSE of 33.18. Heterogeneity of the problem are clearly seen from the figure.

4 Conclusion

In this article, we study the problem of nonparametrically estimating variance function in functional data analysis. We derived the asymptotic property for the squared residuals based estimator and its superiority to the direct squared responses based method is demonstrated through simulations. Our asymptotic result shows an interesting interaction between the smoothness of the mean function and that of the variance function. Finally, we show there exists clear heterogeneity in the regression problem for the chemometric data as an illustration.

Acknowledgements

This research is funded by Singapore MOE Tier 1.
Figure 1: Estimated variance function vs. squared residuals on validation data.
Appendix

First, we make the remark that under condition (12), we can assume \( m(x) \leq M_m \) and \( v(x) \leq M_v \), that is the mean and variance functions are bounded, without loss of generality. The reason is that we always consider only values of both functions inside a compact neighborhood of the fixed \( x \). As an illustration, in the definition of the estimator \( \hat{v}(x) \), \( \Delta_i^v > 0 \) only when \( d_v(x, X_i) \leq h_v \), so the sum over \( i \) is only for all \( X_i \)’s contained in a neighborhood of \( x \).

To make the presentation clear, we first state the asymptotics for mean function estimation in a Lemma. Note all asymptotic orders obtained below are in the sense of convergence in probability.

**Lemma 1** Under conditions (11)-(13), we have

\[
| r_1^m(x) - 1 | = O\left( \sqrt{s_m^m \log n / n^2} \right) \tag{15}
\]

\[
| r_2^m(x) - m(x) | = O\left( h_m^m + \sqrt{s_m^m \log n / n^2} \right) \tag{16}
\]

\[
| E( r_2^m / r_1^m ) - m(x) | = O\left( h_m^m + \sqrt{s_m^m \log n / n^2} \right) \tag{17}
\]

\[
| r_2^m / r_1^m - E( r_2^m / r_1^m ) | = O\left( h_m^m + \sqrt{s_m^m \log n / n^2} \right). \tag{18}
\]

If in addition, condition (14) is satisfied, the above convergence is uniform over a compact neighborhood of \( x \) in \( \mathcal{H} \).

**Proof:** The proofs of (15) and (16) are similar to that contained in Ferraty & Vieu (2004, 2006). On one hand, the proof is simplified by the observation that we only require convergence in probability. On the other hand, the fact that we impose conditions directly on the errors instead of the responses make the proof slightly more complicated. Equations (17) and (18) are direct consequences of the first
two equations and the last statement of the lemma follows from Ferraty & Vieu (2008). We only show (16) below.

The bias $|E r^m_2(x) - m(x)| = O(h^\alpha_m)$ is shown exactly as in Ferraty & Vieu (2004). For variance calculation, we have

$$Var(r^m_2 - E(r^m_2|X^n_1)) = E[Var(r^m_2 - E(r^m_2|X^n_1)|X^n_1)]$$
$$= E[Var(\frac{1}{n} \sum_i \Delta^m_i v_i \epsilon_i|X^n_1)]$$
$$\leq \frac{M^2}{n^2} s^m_{n,2} = O(\frac{1}{n^2} s^m_{n,2}).$$

Similarly, $Var(E(r^m_2|X^n_1) - E r^m_2) = O(s^m_{n,3}/n^2)$ using equation (5). Since $Var(r^m_2) = Var(r^m_2 - E(r^m_2|X^n_1)) + Var(E(r^m_2|X^n_1) - E r^m_2) = O(s^m_n/n^2)$, (16) follows from the Markov inequality. □

**Proof of Theorem 1:**

Using the decomposition

$$\hat{R}_i = (Y_i - \hat{m}_i)^2 = v_i + 2\sqrt{v_i}(m_i - \hat{m}_i)\epsilon_i + (m_i - \hat{m}_i)^2 + v_i(\epsilon_i^2 - 1) =: A_i + B_i + C_i + D_i,$$

and similar to the proof of Lemma 1 we have

$$|r^u_2 - 1| = O(\sqrt{v} s^u_{n,1} \log n/n^2), \quad (19)$$

and we only need to show that

$$r^u_2 = \sum_i \Delta^u_i(A_i + B_i + C_i + D_i)/n = O(h^\beta_u + \sqrt{s^u_n \log n/n^2} + h^{2\alpha}_m + s^m_n \log n/n^2).$$

Using conditions (11)-(13), we have

$$\sum_i \Delta^u_i(A_i + D_i)/n - v(x) = O(h^\beta_u + \sqrt{(s^u_{n,2} + s^u_{n,3}) \log n/n^2}), \quad (20)$$

15
following the same steps as the proof of (16). Also,

\[
\sum_i \Delta_i^v C_i / n = \sum_i (m_i - \hat{m}_i)^2 \Delta_i^v / n \\
\leq \sup_i (m_i - \hat{m}_i)^2 r_i^v = O(h_m^2 n + s_n^m \log n / n^2), \tag{21}
\]

where the supremum over \(i\) obeys the same rate as for a fixed \(x\) because we can take only \(i\) such that \(d_v(x, X_i) \leq h_v\), which is contained in any fixed compact neighborhood of \(x\) and note the final statement of Lemma 1.

Finally, the term \(\sum_i \Delta_i^v B_i / n\) is dealt with in Lemma 2. The theorem is proved combining the following lemma with (19), (20) and (21).

**Lemma 2** In the context of Theorem 1, we have \(\sum_i \Delta_i^v B_i / n = O(\sqrt{(s_{n,2}^v + s_{n,4}^v) \log n / n^2} + s_{n,4}^v / n^2)\).

**Proof:** Writing

\[
\sum_i \Delta_i^v B_i / n = \frac{2}{n} \sum_i \Delta_i^v \sqrt{v_i} (m_i - E\hat{m}_i) \epsilon_i + \frac{2}{n} \sum_i \Delta_i^v \sqrt{v_i} (E\hat{m}_i - \hat{m}_i) \epsilon_i \\
= \frac{2}{n} \sum_i \Delta_i^v \sqrt{v_i} (m_i - E\hat{m}_i) \epsilon_i + \frac{2}{n} \sum_i \Delta_i^v \sqrt{v_i} \epsilon_i (E(\sum_j w_{ij} m_j) - \sum_j (w_{ij} m_j)) \\
- \frac{2}{n} \sum_i \Delta_i^v \sqrt{v_i} \epsilon_i (\sum_j (w_{ij} \sqrt{v_j} \epsilon_j)) =: F + G + H. \tag{22}
\]

We have \(E(F) = E(F|X_1^n) = 0\) and

\[
Var(F|X_1^n) = \frac{4}{n^2} \sum_{i,j} (m_i - E\hat{m}_i)(m_j - E\hat{m}_j) Cov(\Delta_i^v \sqrt{v_i} \epsilon_i, \Delta_j^v \sqrt{v_j} \epsilon_j | X_1^n) \\
= o(\frac{4M}{n^2} \sum_{i,j} |Cov(\Delta_i^v \epsilon_i, \Delta_j^v \epsilon_j | X_1^n)|).
\]

Thus \(Var(F) = E(Var(F|X_1^n)) = o(s_{n,2}^v / n^2)\) and \(F = o(\sqrt{s_{n,2}^v \log n / n^2})\).
Also, for the second term in (22), we have $E(G) = 0$ and

$$\text{Var}(G|X^n_1) = \frac{4}{n^2} \text{Var}(\sum_i \Delta^v_i \sqrt{v_i} \epsilon_i (\sum_j w_{ij} m_j - E \sum_j w_{ij} m_j)|X^n_1)$$

$$\leq \frac{4M^2_v M_v}{n^2} \sum_{i,j} |\text{Cov}(\Delta^v_i \epsilon_i, \Delta^v_j \epsilon_j|X^n_1)|$$

Thus $\text{Var}(G) = E(\text{Var}(G|X^n_1)) = O(s_{n,2}/n^2)$ and $G = O(\sqrt{s_{n,2} \log n/n^2})$. □

Finally, for the third term $H$,

$$E(H) = \frac{2}{n} E \sum_i \Delta^v_i \sqrt{v_i} \epsilon_i \sum_j w_{ij} \sqrt{v_j} \epsilon_j$$

$$= O(s_{n,4}/n^2)$$

and

$$\text{Var}(H|X^n_1) = \frac{4}{n^2} \text{Var}(\sum_i \Delta^v_i \sqrt{v_i} \epsilon_i \sum_j w_{ij} \sqrt{v_j} \epsilon_j|X^n_1)$$

$$= \frac{4M^2_v}{n^2} \sum_{i,j,k,l} |\text{Cov}(\Delta^v_i w_{ij} \epsilon_i \epsilon_j, \Delta^v_k w_{kl} \epsilon_k \epsilon_l|X^n_1)|.$$}

Thus $H = O(s_{n,4}/n^2 + \sqrt{s_{n,4} \log n/n})$.

**Proof of Corollary 1**: We need to show that in the i.i.d. case, $s_n^m = O(n/\phi_m(h_m))$ and $s_n^v = O(n/\phi_v(h_v))$. We choose to calculate $s_{n,1}^m, s_{n,4}^m$ and $s_{n,4}^v$, the calculations are similar for the others.

In the i.i.d. case, we have

$$\sum_{i,j} |\text{Cov}(\Delta^m_i \epsilon_i, \Delta^m_j \epsilon_j|X^n_1)| = \sum_i (\Delta^m_i)^2.$$}

Thus $s_{n,1}^m = nE\Delta^2 = O(n/\phi_m(h_m))$ by Lemma 4.3 of Ferraty & Vieu (2006).
For $s_{n,4}^m$, we have
\[
E \sum_{i,j} \Delta_i^v w_{ij} \epsilon_i \epsilon_j = E \sum_i \Delta_i^v w_{ii} \epsilon_i^2 = O\left(\frac{1}{n \phi_m(h_m)} E \sum_i \Delta_i^v\right) = O\left(\frac{1}{\phi_m(h_m)}\right),
\]
where we used the fact $w_{ii} = K(0)/\sum_j K(h^{-1}_m d_m(X_i, X_j)) = O((n \phi_m(h_m))^{-1})$ obtained from (15) and Lemma 4.3 of Ferraty & Vieu (2006).

For $s_{n,4}^v$, we have
\[
\sum_{i,j,k,l} |Cov(\Delta_i^v w_{ij} \epsilon_i \epsilon_j, \Delta_k^v w_{kl} \epsilon_k \epsilon_l|X_i^n) = \sum_i (\Delta_i^v)^2 w_{ii}^2 + 2 \sum_{i \neq j} (\Delta_i^v)^2 w_{ij}^2 \leq O\left(\frac{1}{n \phi_m(h_m) \phi_v(h_v)}\right) + O\left(\frac{1}{\phi_m(h_m) \phi_v(h_v)}\right) = O\left(n/\phi_v(h_v)\right),
\]
since it is assumed that $n \phi_m(h_m) \to \infty$.

References

Brown, L. D. & Levine, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. *Annals of Statistics* 35, 2219–2232.

Cai, T. T., Levine, M. & Wang, L. (2009). Variance function estimation in multivariate nonparametric regression with fixed design. *Journal of Multivariate Analysis* 100, 126–136.
CAI, T. T. & WANG, L. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *Annals of Statistics* **36**, 2025–2054.

FAN, J. Q. & YAO, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645–660.

FERRATY, F., MAS, A. & VIEU, P. (2007). Nonparametric regression on functional data: Inference and practical aspects. *Australian & New Zealand Journal of Statistics* **49**, 267–286.

FERRATY, F. & VIEU, P. (2002). The functional nonparametric model and application to spectrometric data. *Computational Statistics* **17**, 545–564.

FERRATY, F. & VIEU, P. (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *Journal of nonparametric statistics* **16**, 111–125.

FERRATY, F. & VIEU, P. (2006). *Nonparametric Functional Data Analysis Theory and Practice*. Springer series in statistics,. New York, NY: Springer.

FERRATY, F. & VIEU, P. (2008). Erratum of: Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *Journal of nonparametric statistics* **20**, 187–189.

HALL, P. & CARROLL, R. J. (1989). Variance function estimation in regression - the effect of estimating the mean. *Journal of the Royal Statistical Society Series B-Methodological* **51**, 3–14.
Lian, H. (2007). Nonlinear functional models for functional responses in reproducing kernel hilbert spaces. *Canadian Journal of Statistics-Revue Canadienne De Statistique* **35**, 597–606.

Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and Their Applications* **115**, 155–177.

Muller, H. G. & Stadtmuller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Annals of Statistics* **15**, 610–625.

Muller, H. G. & Stadtmuller, U. (1993). On variance function estimation with quadratic forms. *Journal of Statistical Planning and Inference* **35**, 213–231.

Ramsay, J. O. & Silverman, B. W. (2005). *Functional data analysis*. Springer series in statistics. New York: Springer, 2nd ed.

Wang, L., Brown, L. D., Cai, T. T. & Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *Annals of Statistics* **36**, 646–664.