Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Standard Errors for Nonparametric Regression

Ba M. Chu, David T. Jacho-Chávez & Oliver B. Linton

To cite this article: Ba M. Chu, David T. Jacho-Chávez & Oliver B. Linton (2020) Standard Errors for Nonparametric Regression, Econometric Reviews, 39:7, 674-690, DOI: 10.1080/07474938.2020.1772563

To link to this article: https://doi.org/10.1080/07474938.2020.1772563

ECONOMETRIC REVIEWS
2020, VOL. 39, NO. 7, 674–690
https://doi.org/10.1080/07474938.2020.1772563

Taylor & Francis
Taylor & Francis Group

Check for updates

Standard Errors for Nonparametric Regression

Ba M. Chu[a], David T. Jacho-Chávez[b], and Oliver B. Linton[c]

[a]Department of Economics, Carleton University, Ottawa, ON, Canada; [b]Department of Economics, Emory University, Atlanta, GA, USA; [c]Department of Economics, University of Cambridge, Cambridge, United Kingdom

ABSTRACT
This paper proposes five pointwise consistent and asymptotic normal estimators of the asymptotic variance function of the Nadaraya-Watson kernel estimator for nonparametric regression. The proposed estimators are constructed based on the first-stage nonparametric residuals, and their asymptotic properties are established under the assumption that the same bandwidth sequences are used throughout, which mimics what researchers do in practice while making derivations more complicated instead. A limited Monte Carlo experiment demonstrates that the proposed estimators possess smaller pointwise variability in small samples than the pair and wild bootstrap estimators which are commonly used in practice.

KEYWORDS
Nonparametric Regression; Nonparametric Standard Errors; Bootstrap

MATHEMATICS SUBJECT CLASSIFICATION (2000)
62G08; 62G20

1. Introduction

Peter C. B. Phillips has made so many important research contributions in different areas and has developed the career of many fine econometricians that it is a great honour to write a paper for his 70th birthday. Peter made many contributions to nonparametric estimation methodology, to standard error construction, to improving inference in parametric, semiparametric and nonparametric settings with stationary and nonstationary regressors, see, e.g., Bandi and Phillips (2003), Jeffrey et al. (2004), Phillips (2005, 2007, 2009), Sul et al. (2005), Phillips et al. (2006), Phillips et al. (2007), Xu and Phillips (2008), Sun et al. (2008), Wang and Phillips (2009a,b, 2011), Phillips and Su (2011), and Li et al. (2016). Our paper considers the issue of constructing standard errors for nonparametric regression.

Consider the nonparametric regression

$$Y_i = m_0(X_i) + \epsilon_i, \tag{1.1}$$

where X_i is a vector of d continuous regressors, $E[\epsilon_i|X_i = x] = 0,$ and $E[\epsilon_i^2|X_i = x] = \sigma^2(x)$. We consider the setting where a sequence of independent and identically distributed (i.i.d.) observations, $\{Y_i, X_i^\top\}_{i=1}^n$, is used to construct the Nadaraya-Watson (NW) estimator $\hat{m}(x)$ of $m_0(x)$. Let $v(x) = ||K||_2^2\sigma^2(x)/f_X(x)$ denote the first-order asymptotic variance of, $\sqrt{nh_n^d}\hat{m}(x)$, where h_n represents a common bandwidth parameter, $f_X(x)$ is the joint density function of X evaluated at x, and $||K||_2^2 = \int K^2(u)du$ for some given kernel function, $K(\cdot)$. This paper derives the pointwise asymptotic normality of five different estimators of $v(x)$, namely $\hat{v}_{n,l}(x)$ for $l = 1, 2, \dots, 5,$ which

CONTACT Oliver B. Linton obl20@cam.ac.uk Department of Economics, University of Cambridge, Cambridge CB3 9DD, United Kingdom.

Supplemental data for this article can be accessed on the publisher's website.

are based on the nonparametric residuals, $\hat{\epsilon}_i = Y_i - \hat{m}(X_i)$, and utilize the same bandwidth, h_n, and kernel function, $K(\cdot)$, used to construct $\hat{m}(x)$.

Estimation of $v(x)$ is important because it is needed to construct pointwise asymptotic confidence intervals for the NW estimates of $m_0(x)$. For example, the np package of Hayfield and Racine (2008) in the R statistical language calculates pointwise 95% confidence intervals as $\hat{m}_0(x) \pm 1.965\sqrt{\hat{v}_{n,2}(x)/nh_n^d}$ where $\hat{v}_{n,2}(\cdot)$ is the second estimator under consideration here and defined in (2.2) below. With the notable exceptions of Hall (1992), Fan and Yao (1998), Yu and Jones (2004) and Xu and Phillips (2011), to the best of our knowledge, there has been no further research on establishing the theoretical properties of standard errors and confidence intervals beyond their consistency, see, e.g., Härdle (1990) and references therein. Their properties such as pointwise biases and variances are important because they allow practitioners to choose between competing estimators in a given application. In a limited Monte Carlo experiment, we find that all the proposed estimators display smaller variances than the estimators of $v(\cdot)$ implied by the pair and wild bootstrap which are indiscriminately used in practice, see, e.g., Huynh and Jacho-Chávez (2009a,b). Similarly, it is shown below that an estimator (namely $\hat{v}_{n,4}(\cdot)$) based on the *internalization* idea of Mack and Müller (1989) and Linton and Jacho-Chávez (2010) has uniformly smaller variance that the remaining four proposed estimators when using a second-order Gaussian kernel and $\epsilon_i/\sigma(X_i)$ in (1.1), are i.i.d. Gaussian and independent of X_i. As shown below, the derived bias functions are different across estimators and depend on the higher derivatives of $m_0(\cdot)$ and $f_X(\cdot)$. Similarly, $\hat{v}_{n,l}(x)\hat{f}_n(x)/||K||_2^2$ for $l = 1, 2, \ldots, 5$ then become consistent and asymptotic normally distributed estimators of $\sigma^2(x)$ which are for example suitable competitors for Fan and Yao's (1998) estimator. Finally, our assumptions are mild in the sense that bandwidths obtained from commonly bandwidth-selection procedures such as cross-validation or Silverman's (1986) rule-of-thumb are consistent with our asymptotic theory.

We consider only the i.i.d. case for expositional reasons. A curious property of nonparametric regression is that the asymptotic distribution is the same in the i.i.d. world as in the weakly dependent time series world, see, e.g., Robinson (1983), which means that many results for the i.i.d. case carry over directly to the weakly dependent case, see also Fan and Yao (1998).

1.1. Literature Review

In the classical parametric world, there is much work on estimating asymptotic covariance matrices of parameter estimates. In the likelihood framework, Efron and Hinkley (1978) showed that the observed information was a better estimator of the variance of the MLE than the expected information in terms of asymptotic variance. In Econometrics much focus has been on robustness of standard errors and test statistics constructed from them. White (1980) developed standard errors for linear regression that are robust to heteroskedasticity, which was followed by Newey and West (1987) that developed consistent standard errors in the presence of autocorrelation of unknown form as well as heteroskedasticity. Andrews (1991) establishes the Mean Squared Error (MSE) of a class of long run variance estimators and uses this to derive an optimal estimator within his kernel class. Kauermann and Carroll (2001) investigate the efficiency of the sandwich estimator under the i.i.d. model. Chesher and Jewitt (1987) investigate the bias of White's (1980) estimator and some proposed modifications.

The paper is structured as follows: Section 2 introduces and discusses the proposed five estimators. Section 3 presents the main theoretical results of the paper, while the numerical performance of the proposed estimators in small samples are presented in Section 4. All proofs and auxiliary materials are collected in the Appendix and the online supplemental material of this paper.

2. Estimators of $v(x)$

We define our estimators of $v(x)$ here, and most of them are already well known. We only consider estimators that are robust to heteroskedasticity, which excludes differencing-type estimators of Hall and Yatchew (2005) and many others, see, e.g., Hastie and Tibshirani (1990). We also follow the conclusion of Fan and Yao (1998) in that we only consider estimators of the conditional variance that are based on the squared residuals; this guarantees that our estimators are all positive.

Let $K(\cdot)$ be some kernel function satisfying Assumption 3.1 below and let $K_h(\cdot) = h_n^{-d}K(\cdot/h_n)$ for any positive sequence of bandwidths h_n. Let:

$$\hat{v}_{n,1}(x) = nh_n^d \sum_{i=1}^{n} w_{n,i}^2(x)\hat{\epsilon}_i^2, \tag{2.1}$$

$$\hat{v}_{n,2}(x) = ||K||_2^2 \frac{\hat{\sigma}_a^2(x)}{\hat{f}_n(x)}, \tag{2.2}$$

$$\hat{v}_{n,3} = \hat{\sigma}_a^2(x) nh_n^d \sum_{i=1}^{n} w_{n,i}^2(x), \tag{2.3}$$

$$\hat{v}_{n,4} = nh_n^d \sum_{i=1}^{n} w_{n,i}^2(x)\hat{\sigma}_a^2(X_i), \tag{2.4}$$

where $w_{n,i}(x) = K_{h_n}(X_i - x)/\sum_{i=1}^{n} K_{h_n}(X_i - x)$ and $\hat{\epsilon}_i = Y_i - \hat{m}(X_i)$, while $\hat{\sigma}_a^2(x) = \sum_{i=1}^{n} w_{n,i}(x)\hat{\epsilon}_i^2$ and $\hat{\sigma}_a^2(X_i) = \sum_{j=1, j\neq i}^{n} w_{n,i,j}\hat{\epsilon}_j^2$ with $w_{n,i,j} = K_{h_n}(X_i - X_j)/\sum_{\ell=1, \ell\neq i}^{n} K_{h_n}(X_i - X_\ell)$, and finally, $\hat{f}_n(x) = \sum_{i=1}^{n} K_{h_n}(X_i - x)/n$ is the kernel estimate of $f_X(x)$.

In addition, we also consider the residual-based wild bootstrap scheme: $\tilde{\epsilon}_i^* = v_i\tilde{\epsilon}_i$, where $\tilde{\epsilon}_i = \hat{\epsilon}_i - \frac{1}{n}\hat{\epsilon}_i$ and v_i, $i = 1, \ldots, n$, are i.i.d.(0,1) such that $E[v_i^4] \in (1, \infty)$. We construct a bootstrap sample, $(\tilde{Y}_i^*, X_i^\top)^\top, i = 1, \ldots, n$, where $\tilde{Y}_i^* = \hat{m}(X_i) + \tilde{\epsilon}_i^*$, then nonparametrically regress $\tilde{Y}_i^*, i = 1, \ldots, n$, on X_i, $i = 1, \ldots, n$ and obtain a sequence of wild bootstrap residuals, $\hat{\epsilon}_i^* = \tilde{Y}_i^* - \hat{m}^*(X_i)$, where $\hat{m}^*(x) = \sum_{i=1}^{n} w_{n,i}(x)\tilde{Y}_i^*$. The residual-based wild bootstrap estimate of $v(x)$ is then given by

$$\hat{v}_{n,5}^*(x) = nh_n^d \sum_{i=1}^{n} w_{n,i}^2(x)\hat{\epsilon}_i^{*2}. \tag{2.5}$$

Härdle (1990, p. 100) proposes $\hat{v}_{n,2}(x)$, which is often used in practice. Estimators (2.1), (2.4) and (2.5) are closest in spirit to robust standard errors in linear regression, see, i.e., Eicker (1967) and White (1980), while Estimator (2.2) uses the specific structure of the variance of the limiting distribution.

Estimator (2.5) uses re-centred residuals because unlike ordinary least squares, nonparametric estimation does not impose the mean of the residuals to be zero. Similarly, unlike Estimators (2.1), (2.4) and (2.5), Estimators (2.2) and (2.3) separate out the estimator of the variance from the weights. Estimator (2.4) combines the leave-one-out principle with the internalization idea of Mack and Müller (1989) and Linton and Jacho-Chávez (2010). It is also shown below to have the smallest pointwise variance when using a second-order Gaussian kernel and $\epsilon_i/\sigma(X_i)$ in (1.1), are i.i.d. Gaussian and independent of X_i. Estimator (2.4) also displays the best pointwise performance among all estimators in our Monte Carlo experiments. Our numerical experimentation shows that Estimator (2.4) also provides the most accurate nominal coverage probability for non-linear designs in large samples.

Notice that one can replace $\hat{m}(\cdot)$ in the estimators above by the local polynomial estimator see Fan and Gijbels; 1996 which undoubtedly will yield the same pointwise asymptotic variance but different bias functions. In fact, this will still be the case if we were to replace it with other smoothers such as series, wavelets or splines for example. For this reason, we only consider the NW estimator here because of its unchallenged popularity in applied research.

3. Main Results

The following notation is used for the remaining part of the paper: C_0 is a generic constant that may vary from one context to another; $||f||_\infty = \sup_x |f|$; $\mathbb{I}(w \in W) = 1$ if $w \in W$ and zero otherwise; $\sum_{|\beta|=k}$ is a summation taken over all possible integer vectors, $\beta = (\beta_1, ..., \beta_d)^\top$, such that $|\beta| = \sum_{i=1}^d \beta_i = k$, where d is the dimension of β; $\beta! = \beta_1! \times \cdots \times \beta_d!$; $D^\beta f = \frac{\partial^{|\beta|}}{\partial^{\beta_1} \cdots \partial^{\beta_d}} f$; $u^\beta = u_1^{\beta_1} \times \cdots \times u_d^{\beta_d}$. The following assumptions are needed for the main results:

Assumption 3.1. K(u) is a twice continuously differentiable symmetric kernel function which is zero outside a bounded set, all the partial derivatives are bounded and Lipschitzian, $||K||_\infty < C(K)$, where C(K) is the upper bound specific to $K(\cdot)$; $\int K(u)du = 1$; and $\int ||u||^2 K(u)du < \infty$, where $||u|| = \sqrt{u_1^2 + \cdots + u_d^2}$.

Assumption 3.2. The errors ϵ_i, $i = 1, ..., n$, satisfy $E[\exp(\ell|\epsilon_1|)|X_1 = x] \le C$ almost surely for some constant $C > 0$ and $\ell > 0$ small enough. Also, $\sup_x E[|\epsilon_1|^2 | X_1 = x] < \infty$.

Assumption 3.3. The probability density function, f_X, of X_i is bounded and twice differentiable; $\sup_{|\beta|=1} ||D^\beta f_X||_\infty < \infty$ and $\sup_{|\beta|=2} |D^\beta f_X(u) - D^\beta f_X(v)| < C_0||u - v||$.

Assumption 3.4. $\sigma^2(x)$ is a twice continuously differentiable function and $\sup_{|\beta|=2} |D^\beta \sigma^2(u) - D^\beta \sigma^2(v)| < C_0||u - v||$.

Assumption 3.5. $m_0(x)$ is a bounded and twice differentiable function such that $\sup_{|\beta|=1} ||D^\beta m_0||_\infty < \infty$ and $\sup_{|\beta|=2} |D^\beta m_0(u) - D^\beta m_0(v)| < C_0||u - v||$.

Assumptions 3.1 is rather typical for kernel smoothing, while Assumptions 3.3–3.4 are also standard smoothness and boundedness assumptions in the literature. Assumption 3.2 is needed here to apply some results from Empirical Process theory used in the proofs and it amounts to having sub-exponential tails of ϵ conditional on X, see, e.g., Mammen et al. (2012) and Escanciano et al. (2014, 2016).

Assumption 3.6. $E[||X_1||^{\rho+1}] < \infty$ and $E[|Y_1|^p] < \infty$ for some $\rho > 0$ and $p > 2$.

Assumption 3.7. The bandwidth h_n verifies

1. $n^{1-\alpha} h_n^{\frac{p+2}{p-2}d} \uparrow \infty$ for some $\alpha \in (0, 1)$; $\frac{n^{1-2/p} h_n^d}{\log(n)} \uparrow \infty$; $n^{1-2\alpha} h_n^d \uparrow \infty$; $n h_n^{d+8} \downarrow 0$; and $\frac{h_n^d}{n^{4\rho - 1}} \to \text{const}$.
2. $\frac{1}{\xi_n \delta_n^{2-\alpha_m - 2\alpha_\eta}} \uparrow \infty$ and $\frac{\sqrt{n h_n^d}}{\xi_n \delta_n^{1-\alpha_m - \alpha_\eta} \log(n)} \uparrow \infty$ for some $\alpha_\eta \in (0, (2 - \alpha_m)/2)$, where $\delta_n = \max\left(\sqrt{\frac{n^\alpha}{n h_n^d}}, h_n, \frac{1}{n^\rho}\right)$, $\alpha_m \in (0, 2)$, and ξ_n depends on the choice of h_n

Assumption 3.6 requires the existence of higher moments and is rather standard. The first part of Assumption 3.7 imposes a minimum and a maximum rate of convergence of the bandwidth h_n see, e.g., Ai; 1997, Assumption 14, pp. 941. In particular, if we let $\alpha = 2/p$, $p > 4$, and $\rho \ge \frac{1}{4}$ then the first set of conditions can be simplified to $\min\left\{n^{1-\frac{4}{p}}h_n^d/\log(n), n^{1-\frac{2}{p}}h_n^{\frac{p+2}{p-2}d}/\log(n)\right\} \uparrow \infty$ and $nh_n^{d+8} \downarrow 0$. On the other hand the second part of this assumption is satisfied, for example, by $h_n \propto n^{-\frac{1}{d+4}}$. In particular, arguing along the line of Mammen et al. (2012), suppose that all the k-order partial derivatives of $\hat{m}(x) - E[\hat{m}(x)]$ are bounded by $C_0 n^{\xi^*}$ for some $k > d/2$ and $\xi^* > 0$, one can then choose $\xi_n = n^{\xi^* d/2}$ and $\alpha_m = d/k$. Therefore, commonly-used bandwidth selection procedures such as cross-validation or Silverman's (1986) rule-of-thumb produce bandwidths with asymptotic convergence rates that are compatible with the assumptions above.

The following set of theorems provides the main results of the paper. Let $\mu_{j,l}(K) = \int u^j K^l(u)du$ with $\mu_l(K) = \mu_{0,l}(K)$, and let $m_3(x) = E[\eta_i\epsilon_i/\sigma(X_i)|X_i = x]$, and $m_4(x) = E[\eta_i^2|X_i = x]$, where $\eta_i = \epsilon_i^2/\sigma^2(X_i) - 1$. Define the bias constants:

$$\beta_1(x) = \frac{1}{2f_X^2(x)}\sum_{|\alpha|=1}\sum_{|\beta|=1}\left(b_{\alpha,\beta}(x)\mu_{\alpha+\beta,2}(K) + \frac{1}{\alpha!\beta!}\sigma^2(x)\cdot D^{\alpha+\beta}f_X(x)\left\{\mu_{\alpha+\beta,2}(K) - 2||K||_2^2\mu_{\alpha+\beta,1}(K)\right\}\right),$$

$$\beta_2(x) = \frac{||K||_2^2}{2f_X^2(x)}\sum_{|\alpha|=1}\sum_{|\beta|=1}\mu_{\alpha+\beta,1}(K)\left(b_{\alpha,\beta}(x) - \frac{1}{\alpha!\beta!}\sigma^2(x)\cdot D^{\alpha+\beta}f_X(x)\right),$$

$$\beta_3(x) = \frac{||K||_2^2}{2f_X^2(x)}\sum_{|\alpha|=1}\sum_{|\beta|=1}\left(\mu_{\alpha+\beta,1}(K)b_{\alpha,\beta}(x) + \sigma^2(x)\frac{1}{\alpha!\beta!}D^{\alpha+\beta}f_X(x)\left\{\mu_{\alpha+\beta,2}(K)du - 2||K||_2^2\mu_{\alpha+\beta,1}(K)\right\}\right),$$

$$\begin{aligned}\beta_4(x) = \frac{1}{2f_X^2(x)}\sum_{|\alpha|=1}\sum_{|\beta|=1}&\left(\frac{1}{\alpha!\beta!}D^{\alpha+\beta}\{\sigma^2(x)\cdot f_X(x)\}\mu_{\alpha+\beta,2}(K)\right.\\ &+\frac{1}{\alpha!\beta!}f_X(x)\cdot D^{\alpha+\beta}\sigma^2(x)\int(v-u)^{\alpha+\beta}K^2(u)K(u-v)dudv\\ &\left.-2||K||_2^2\mu_{\alpha+\beta,1}(K)\frac{1}{\alpha!\beta!}\sigma^2(x)\cdot D^{\alpha+\beta}f_X(x)\right),\end{aligned}$$

where $b_{\alpha,\beta}(x) = \frac{1}{\alpha!\beta!}\left\{f_X(x)D^{\alpha+\beta}\sigma^2(x) + 2D^\alpha f_X(x)D^\beta\sigma^2(x)\right\}$, and define the limiting variances:

$$s_1(x) = \mu_4(K)\frac{\sigma^4(x)}{f_X^3(x)}(m_4(x)+1) + \left(4\mu_2^3(K) - 4\mu_2(K)\mu_3(K)\right)\frac{\sigma^4(x)}{f_X^3(x)}.$$

$$s_2(x) = \mu_2^3(K)\frac{\sigma^4(x)}{f_X^3(x)}(m_4(x)+1).$$

$$s_3(x) = \mu_2^3(K)\frac{\sigma^4(x)}{f_X^3(x)}m_4(x) + \left(\mu_4(K) + 4\mu_2^3(K) - 4\mu_2(K)\mu_3(K)\right)\frac{\sigma^4(x)}{f_X^3(x)}.$$

$$s_4(x) = \left(\mu_4(K) + 4\mu_2^3(K) - 4\mu_2(K)\mu_3(K)\right)\frac{\sigma^4(x)}{f_X^3(x)} + \mu_2(K^2 * K)m_4(x)]\frac{\sigma^4(x)}{f_X^3(x)}.$$

$$s_5(x) = \mu_4(K)\frac{\sigma^4(x)(m_4(x)+1)}{f_X^3(x)}\kappa_v,\ \text{where } \kappa_v = E\left[v_i^4\right] - 1.$$

Theorem 1. *Suppose that Assumptions 3.1-3.5 and 3.6-3.7 hold. Then,* $\sqrt{nh_n^d}(\hat{v}_{n,1}(x) - v(x) - h_n^2\beta_1(x)) \xrightarrow{d} N(0, s_1(x))$.

Theorem 2. *Suppose that Assumptions 3.1-3.5 and 3.6-3.7 hold. Then,* $\sqrt{nh_n^d}(\hat{v}_{n,2} - v(x) - h_n^2\beta_2(x)) \xrightarrow{d} N(0, s_2(x))$.

Theorem 3. *Let Assumptions 3.1-3.5 and 3.6-3.7 hold. Then,* $\sqrt{nh_n^d}(\hat{v}_{n,3}- v(x)- h_n^2\beta_3(x))\xrightarrow{d}N(0,s_3(x))$.

Theorem 4. *Suppose that Assumptions 3.1-3.5 and 3.6-3.7 hold. Then,* $\sqrt{nh_n^d}(\hat{v}_{n,4}(x)- v(x)- h_n^2\beta_4(x))\xrightarrow{d}N(0,s_4(x))$

Theorem 5. *Define* $\hat{v}_{n,5}(x) = nh_n^d\sum_{i=1}^n w_{n,i}^2(x)\tilde{\epsilon}_i^2$, *and suppose that Assumptions 3.1-3.5 and 3.6-3.7 hold. Then,* $\sqrt{nh_n^d}\{\hat{v}_{n,5}^*(x) - \hat{v}_{n,5}(x)\}\xrightarrow{d_{P*}}N(0,s_5(x))$, *where* $\xrightarrow{d_{P*}}$ *represents 'convergence in bootstrap distribution.'*

In general, there is no ranking in terms of the MSE of the proposed standard error estimators, and not even a ranking based on bias, which is usual in nonparametric estimation problems, see, e.g., Fan and Gijbels (1996). The bias terms all depend on the derivatives of the marginal density and the heteroskedastic function as well as kernel constants, but not on the curvature of the regression function m_0, as in Fan and Yao (1998). Note that we may also consider the estimators with the re-centred residuals $\tilde{\epsilon}_i = \hat{\epsilon}_i - n^{-1}\sum_{i=1}^n \hat{\epsilon}_i$ – This has exactly the same asymptotic bias and variance (as the corresponding case with raw residuals $\hat{\epsilon}_i$). Suppose that $d=1$ and $f_X(\cdot)$ is uniform between 0 and 1. Then the derived biases simplify to $\beta_1(x) = 0.5\times\mu_{2,2}(K) D^2\sigma^2(x), \beta_2(x) = \beta_3(x) = 0.5\times||K||_2^2\mu_{2,1}(K)D^2\sigma^2(x), \beta_4(x) = 0.5\times\left\{\mu_{2,2}(K)+\int(v-u)^2K^2(u)K(u-v)dudv\right\}D^2\sigma^2(x)$, and the following table summarizes the proportionality constants for various kernel functions:

This table shows that estimator (2.1) provides the smallest bias for all cases except when using the Epanechnikov kernel where (2.4) displays the smallest bias. However, in this uniform design, the bias of estimator (2.4) is larger than that of estimators (2.2) and (2.3) for the Uniform, Biweight, and Normal kernels. Therefore, there is no a clear winner in terms of bias.

The variance terms are also not generally rankable. However, suppose that $\epsilon_i/\sigma(X_i)$ are i.i.d. Gaussian and independent of X_i. Then, the following table shows that the estimator $\hat{v}_{n,4}$ defined by (2.4) has the smallest asymptotic pointwise variance across all popular non-uniform kernels. Our Monte Carlo exercise below confirms this.

We expect that our results apply exactly as stated in the case where $(Y_i, X_i^\top)$ is a stationary and weakly dependent (alpha mixing) time series process, see, i.e., Robinson (1983). We expect that in some cases our results continue to hold for the case where $(Y_i, X_i^\top)$ is a locally stationary process as in Vogt (2012). In that case, the limiting variance of the NW estimator is proportional to $(\int_0^1 f_u(x)du)^{-2}\int_0^1\sigma^2(x,u)f_u(x)du$, where $f_u(\cdot)$ is the density of the covariate process at location $u\in[0,1]$. Even in that case, the estimators $\hat{v}_{n,j}(x), j = 1,\ldots,4$ are consistent and continue to satisfy the Central Limit Theorems (CLT) given above where the limiting variances are suitably modified. In the case where X_i are globally nonstationary, see, i.e., Wang and Phillips (2009a,b, 2011), the situation becomes radically different because the CLT has a random limiting variance. Nevertheless, one may be able to make comparisons across different implied confidence intervals.

Our analysis has been based on large sample approximations. In some contexts it is appropriate to condition on the regressors as being ancillary. Write $\hat{\epsilon}_i^2 = \epsilon_i^2 + [\hat{m}(X_i) - m(X_i)]^2 - 2\epsilon_i[\hat{m}(X_i) - m(X_i)]$, and obtain $E\left[\hat{\epsilon}_i^2|\mathcal{X}^n\right] = \sigma^2(X_i) + E[(\hat{m}(X_i) - m(X_i))^2|\mathcal{X}^n] - 2E[\epsilon_i(\hat{m}(X_i) - m(X_i))|\mathcal{X}^n]$. In general, we have $E\left[\epsilon_i(\hat{m}(X_i) - m(X_i))|\mathcal{X}^n\right] = O_P(n^{-1}h_n^{-d})$, and so this term is of smaller order than the smoothing bias terms. However, this term is considered undesirable in cross-validation contexts and is analogous to the term explored in Chesher and Jewitt (1987) for the linear regression HAC (Heteroskedasticity Autocorrelation Consistent) estimator. One can easily remove this term by replacing $\hat{m}(X_i)$ by the leave-one-out version, but as we pointed out, this will not affect our asymptotic approximations.

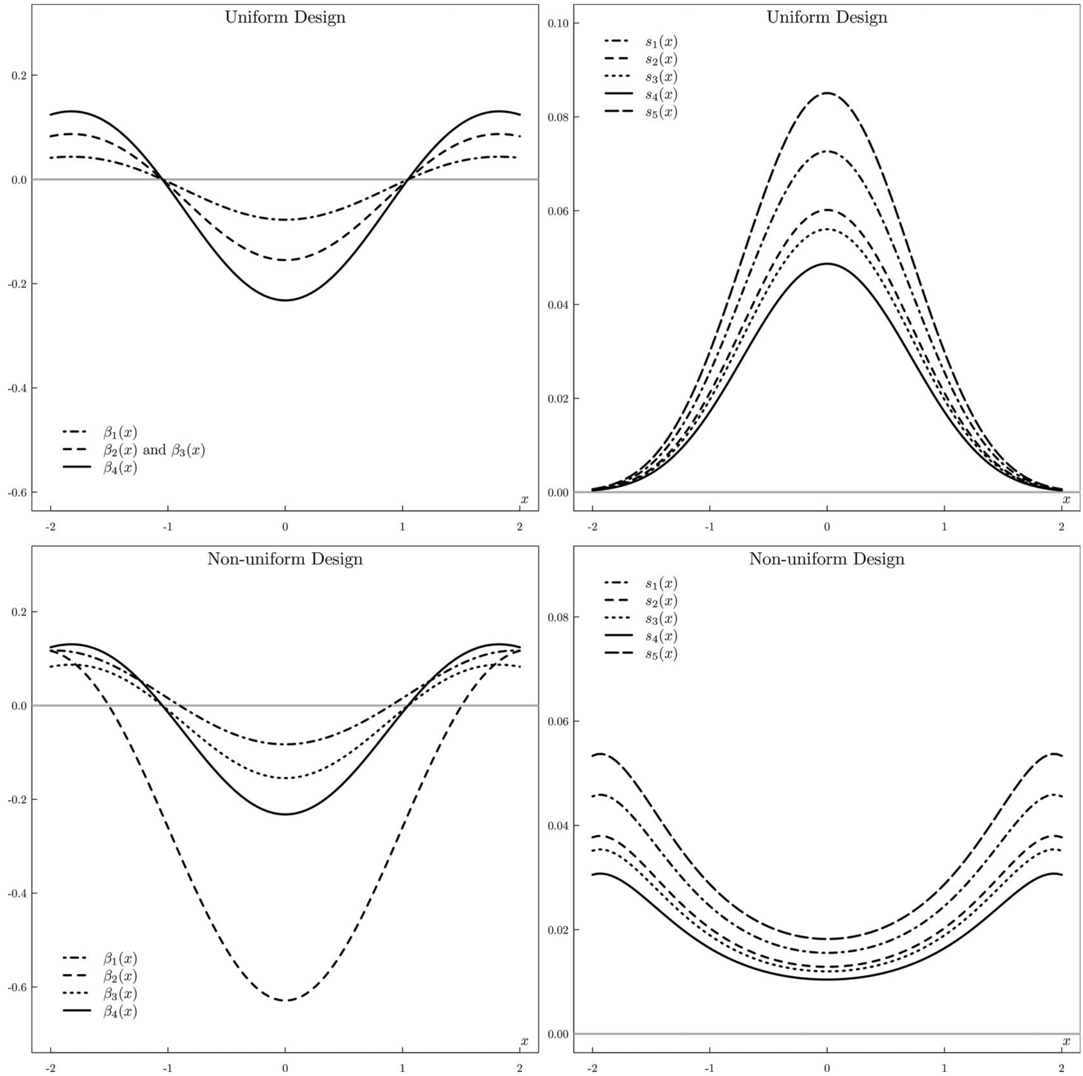

Figure 1. Monte Carlo Designs.
Note: Designs' biases and variances structures as derived in Theorems 1–5.

4. Monte Carlo

We proceed to generate 1,000 Monte Carlo samples of sizes $n \in \{100, 200, 400, 800\}$ from (1.1), using a slightly modified version of the design proposed in Linton and Jacho-Chávez (2010) and also used in Ho et al. (2014). Particularly, we set $d = 1$, $m_0(X_i) = X_i \cos(2\pi X_i), \epsilon_i = \sigma(X_i)\varepsilon_i, \sigma(X_i) = (1 + \cos(X_i))/[2\sin(2) + 4]$ for $i = 1, ..., n$, where ε_i are randomly drawn from a $N(0,\ 1)$ and statistically independent of X_i. As in Linton and Jacho-Chávez (2010), two designs are considered for X_i, i.e. an uniform design (X_i are drawn from an Uniform distribution between –2 and 2), and a truncated design (X_i are drawn from a truncated normal distribution between –2 and 2, and variance equals to 4).

In each replication, estimators (2.1) (Estimator 1), (2.2) (Estimator 2), (2.3) (Estimator 3), (2.4) (Estimator 4), (2.5) (Estimator 5), as well as the pair-wise (Estimator 6) and wild bootstrap (Estimator 7) estimators see Henderson and Parmeter; 2015, Sections 5.10.1 and 5.10.3 are implemented using a second-order Gaussian kernel everywhere and setting h_n to be equal to the least

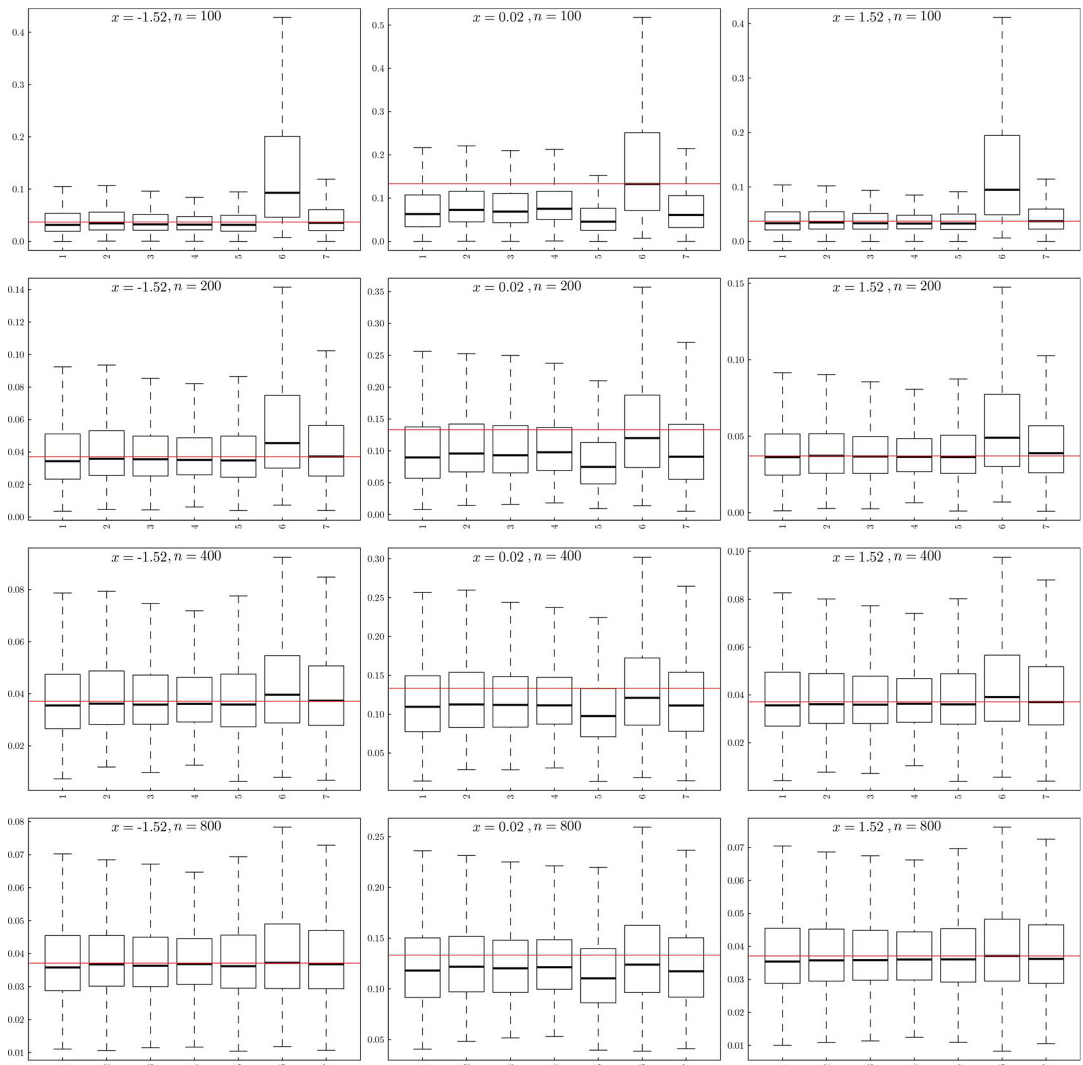

Figure 2. Monte Carlo Results - Uniform Design.
Note: Box plots based on 1,000 Monte Carlo replications of the proposed five estimators, i.e. Estimator 1 (2.1), Estimator 2 (2.2), Estimator 3 (2.3), Estimator 4 (2.4), and Estimator 5 (2.5), as well as of the pairwise (Estimator 6) and wild (Estimator 7) bootstrap. In each replication, all estimators use the same least squares cross-validated bandwidth, a second order gaussian kernel and estimators 5, 6 and 7 are based on 199 bootstrap replications. The red horizontal line represents the true value of v.

squares cross-validated bandwidth, $\hat{h}_n$, for the Nadaraya-Watson estimator of m_0. For Estimators 5 and 7, we set v_i to follow a two-point distribution such that it takes the value $\left(1-\sqrt{5}\right)/2$ with probability $\left(1+\sqrt{5}\right)/\left(2\sqrt{5}\right)$ and $\left(1+\sqrt{5}\right)/2$ with probability $1-\left(1+\sqrt{5}\right)/\left(2\sqrt{5}\right)$ in each of the 199 bootstrap replications. The original cross-validated bandwidth is used across estimators and in each bootstrap replication.

Figure 1 depicts the bias structures in Theorems 1–4, and variance functions derived in Theorems 1–5 for the designs under consideration. This figure shows the different implied bias and variance structures under consideration in this Monte Carlo. As predicted by the derived asymptotic properties, Estimator 4 has the smallest first-order variance function uniformly over the support of X for the uniform and non-uniform designs. However, for the uniform design, Estimator 4 has the largest pointwise bias function, but otherwise no other estimator uniformly

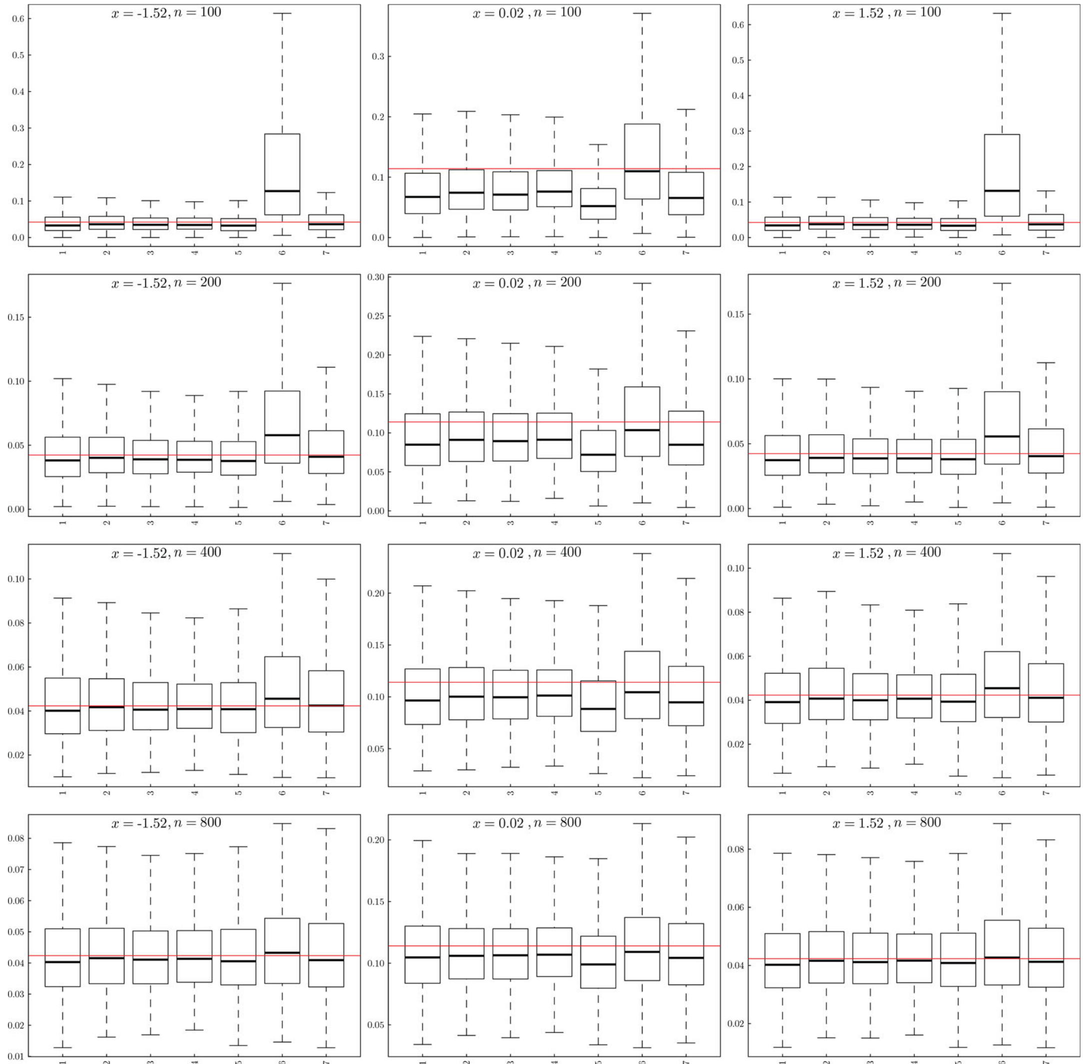

Figure 3. Monte Carlo Results - Non-uniform Design.
Note: Box plots based on 1,000 Monte Carlo replications of the proposed five estimators, i.e. Estimator 1 (2.1), Estimator 2 (2.2), Estimator 3 (2.3), Estimator 4 (2.4), and Estimator 5 (2.5), as well as of the pairwise (Estimator 6) and wild (Estimator 7) bootstrap. In each replication, all estimators use the same least squares cross-validated bandwidth, a second order gaussian kernel and estimators 5, 6 and 7 are based on 199 bootstrap replications. The red horizontal line represents the true value of v.

has better bias than the rest in the non-uniform design case. Estimator 1 has a smaller bias in the interior of the support of X, while Estimator 3's bias is smaller in the tails.

The results are reported in Figures 2 and 3. They display box plots based on the 1,000 Monte Carlo replications of all 7 estimators at three different points in the interior of the support of X, namely $x \in \{-1.52, 0.02, 1.52\}$ for all 4 sample sizes. Firstly, as depicted in Figure 1, all estimators have larger downward biases in the interior than near the boundaries of the support of X. Interestingly across all estimators, sample sizes, designs and estimation points, the pair and wild bootstrap estimators (6 and 7 respectively) show the largest Monte Carlo variability and inter-quartile range, although they have better small-sample median bias performance for most sample sizes. As discussed in the text, Estimator 4 has the best performance in terms of spread among all estimators.

Table 1. Pointwise Bias Rankings for Uniform Designs.

	Uniform	Triangular	Biweight	Epanechnikov	Triweight	Normal
$\beta_1(\cdot) \propto$	8.3333×10^{-2}	3.3333×10^{-2}	3.2468×10^{-2}	4.2857×10^{-2}	2.7195×10^{-2}	7.0525×10^{-2}
$\beta_2(\cdot) \propto$	8.3333×10^{-2}	5.5556×10^{-2}	5.1020×10^{-2}	6×10^{-2}	4.5325×10^{-2}	14.1050×10^{-2}
$\beta_4(\cdot) \propto$	25×10^{-2}	5.0794×10^{-2}	13.0320×10^{-2}	1.7143×10^{-2}	3.3106×10^{-2}	21.1570×10^{-2}

Note: $\beta_3(\cdot) = \beta_2(\cdot)$ in the uniform design and therefore omitted. Kernels functions can be found in (Yatchew; 2003, Fig. 3.2, p. 33)

Table 2. Pointwise Variance Rankings for Non-Uniform Designs.

	Uniform	Triangular	Biweight	Epanechnikov	Triweight	Normal
$s_1(\cdot) \propto$	0.3750	1.0519	1.2405	0.7097	1.8797	8.1349×10^{-2}
$s_2(\cdot) \propto$	0.3750	0.8889	1.0933	0.6480	1.6291	6.7345×10^{-2}
$s_3(\cdot) \propto$	0.3750	0.8444	1.0438	0.6274	1.5452	6.2752×10^{-2}
$s_4(\cdot) \propto$	0.3750	0.7086	0.9334	0.4897	1.3280	5.4514×10^{-2}
$s_5(\cdot) \propto$	0.3750	1.2000	1.3883	0.7714	2.1309	9.5240×10^{-2}

Note: Kernels functions can be found in (Yatchew; 2003, Fig. 3.2, p. 33)

Table 3. Simulated 95% Nominal Level Coverage Probability.

	Uniform Design				Non-uniform Design			
	$n=100$	$n=200$	$n=400$	$n=800$	$n=100$	$n=200$	$n=400$	$n=800$
$x=-1.52$								
[1]	0.766	0.865	0.913	0.922	0.799	0.870	0.898	0.923
[2]	0.797	0.891	0.917	0.933	0.820	0.888	0.912	0.928
[3]	0.788	0.891	0.920	0.928	0.817	0.886	0.910	0.926
[4]	0.810	0.898	0.924	0.934	0.837	0.895	0.913	0.931
[5]	0.711	0.838	0.891	0.918	0.743	0.846	0.884	0.916
[6]	0.907	0.916	0.925	0.936	0.911	0.900	0.906	0.927
[7]	0.748	0.866	0.907	0.925	0.789	0.875	0.897	0.917
$x=0.02$								
[1]	0.767	0.850	0.885	0.927	0.790	0.870	0.900	0.930
[2]	0.798	0.865	0.897	0.932	0.820	0.883	0.910	0.933
[3]	0.789	0.860	0.895	0.930	0.819	0.878	0.906	0.934
[4]	0.812	0.878	0.901	0.931	0.833	0.887	0.914	0.935
[5]	0.714	0.810	0.872	0.919	0.747	0.841	0.883	0.920
[6]	0.906	0.893	0.907	0.932	0.883	0.912	0.922	0.930
[7]	0.765	0.854	0.889	0.927	0.786	0.868	0.905	0.931
$x=1.52$								
[1]	0.714	0.816	0.885	0.907	0.743	0.835	0.898	0.913
[2]	0.753	0.843	0.895	0.905	0.776	0.859	0.903	0.923
[3]	0.739	0.836	0.890	0.906	0.762	0.851	0.903	0.923
[4]	0.768	0.843	0.890	0.908	0.785	0.858	0.909	0.925
[5]	0.669	0.778	0.872	0.896	0.684	0.804	0.890	0.904
[6]	0.868	0.868	0.909	0.912	0.853	0.882	0.913	0.921
[7]	0.703	0.818	0.890	0.904	0.736	0.840	0.905	0.908

Note: This table reports the frequency $m_0(x)$ was found inside the interval $\hat{m}(x) \pm 1.965 \times \sqrt{\hat{v}_{n,l}(x)/n\hat{h}_n}$ in 1,000 replications for $l = 1, ..., 7$ and $\hat{h}_n$ represents the least squares cross-validated bandwidth for $\hat{m}(x)$. We use the notation $\hat{m}(x)$ for Estimator l where $[l]$.

Similarly, in each Monte Carlo replication and for each estimator we recorded whether $m(x)$ was inside the pointwise 95% confidence interval $\hat{m}(x) \pm 1.965 \times \sqrt{\hat{v}_{n,l}(x)/n\hat{h}_n}$ for $l = 1, ..., 7, x \in \{-1.52, 0.02, 1.52\}$, using all 4 sample sizes. Table 3 displays the percentage of times this happened in the 1,000 Monte Carlo replications. As expected, the simulated 95% nominal level coverage probability improves with sample size for all estimators. However, Estimators 4 and 6 (pairwise bootstrap) provide the more accurate nominal coverage at all estimation points and sample sizes, with Estimator 4 again dominating in terms of performance in the non-uniform design for larger samples.

5. Conclusions

We find some small-sample differences in performance between the various proposed estimators of the first-order asymptotic variance for nonparametric regression just as in linear regression. On the basis of our work, the proposed estimator (2.4), i.e., $\hat{v}_{n,4} = nh_n^d \sum_{i=1}^n w_{n,i}^2(x)\hat{\sigma}_a^2(X_i)$ performs the best in moderate sized samples. This estimator shares some features with Bandi and Phillips's (2003, eq. 4.3, pp. 248) volatility estimator, which also involves a kind of double smoothing. Finally, it should be noted that all proposed estimators will remain consistent and asymptotic normally distributed in the presence of weak dependence, see, e.g., Robinson (1983).

A. Proof of Main Theorems

Proof of Theorem 1. Let $\hat{g}_n(x) = (nh_n^d)^{-1}\sum_{i=1}^n K^2((X_i - x)/h_n)$ and $\hat{f}_n(x) = (nh_n^d)^{-1}\sum_{i=1}^n K((X_i - x)/h_n)$. We have

$$\hat{v}_{n,1}(x) - v(x) = nh_n^d \sum_{i=1}^n w_{n,i}^2(x)(\hat{\epsilon}_i^2 - \epsilon_i^2) + nh_n^d \sum_{i=1}^n w_{n,i}^2(x)(\epsilon_i^2 - \sigma(X_i)^2) + nh_n^d \sum_{i=1}^n w_{n,i}^2(x)\left\{a\sigma^2(X_i) - \sigma^2(x)\right\} + \sigma^2(x)\left(\frac{\hat{g}_n(x)}{\hat{f}_n^2(x)} - \frac{||K||_2^2 f_X(x)}{f_X^2(x)}\right). \tag{A.1}$$

Notice that

$$\begin{aligned}\frac{\hat{g}_n(x)}{\hat{f}_n^2(x)} - \frac{||K||_2^2 f_X(x)}{f_X^2(x)} &= \frac{1}{f_X^2(x)}\left(\hat{g}_n(x) - f_X(x)||K||_2^2 - 2||K||_2^2\left(\hat{f}_n(x) - f_X(x)\right)\right) \\ &+ \left(\frac{1}{E\left[\hat{f}_n(x)\right]^2} - \frac{1}{f_X^2(x)}\right)\left(\hat{g}_n(x) - f_X(x)||K||_2^2 - ||K||_2^2 \frac{1}{f_X(x)}\left(\frac{1}{E\left[\hat{f}_n(x)\right]} - \frac{1}{f_X(x)}\right)\left(\hat{f}_n(x) - f_X(x)\right)\right) \\ &+ \left(\frac{1}{\hat{f}_n^2(x)} - \frac{1}{E\left[\hat{f}_n(x)\right]^2}\right)\left(\hat{g}_n(x) - f_X(x)||K||_2^2 - ||K||_2^2\left(1 + \frac{\hat{f}_n(x)}{f_X(x)}\right)\left(\hat{f}_n(x) - f_X(x)\right)\right) \\ &+ \frac{1}{E\left[\hat{f}_n(x)\right]^2}||K||_2^2 \frac{E\left[\hat{f}_n(x)\right] - \hat{f}_n(x)}{f_X(x)}\left(\hat{f}_n(x) - f_X(x)\right).\end{aligned} \tag{A.2}$$

By the asymptotic normality of kernel density estimators e.g., Rosenblatt; 1985, p. 192 and their uniform consistency e.g., Prakasa Rao; 1983, p. 185, the third and fourth terms on the right-hand side of (A.2) is of order $o_p\left(\frac{1}{\sqrt{nh_n^d}}\right)$. Therefore, we have that

$$\begin{aligned}\frac{\hat{g}_n(x)}{\hat{f}_n^2(x)} - \frac{||K||_2^2 f_X(x)}{f_X^2(x)} &= \frac{1}{f_X^2(x)}\left(\hat{g}_n(x) - f_X(x)||K||_2^2 - 2||K||_2^2\left(\hat{f}_n(x) - f_X(x)\right)\right) \\ &+ \left(\frac{1}{E\left[\hat{f}_n(x)\right]^2} - \frac{1}{f_X^2(x)}\right)\left(\hat{g}_n(x) - f_X(x)||K||_2^2 - ||K||_2^2 \frac{1}{f_X(x)}\left(\frac{1}{E\left[\hat{f}_n(x)\right]} - \frac{1}{f_X(x)}\right)\left(\hat{f}_n(x) - f_X(x)\right)\right) \\ &+ o_p\left(\frac{1}{\sqrt{nh_n^d}}\right).\end{aligned}$$

Moreover, since $E[\hat{f}_n(x)] = f_X(x) + o(h_n^2)$, $\hat{f}_n(x) - f_X(x) = O_p\left(\frac{1}{\sqrt{nh_n^d}}\right) + O(h_n^2)$, and $\hat{g}_n(x) - f_X(x)||K||_2^2 = O_p\left(\frac{1}{\sqrt{nh_n^d}}\right) + O(h_n^2)$, it then follows that

$$\frac{\hat{g}_n(x)}{\hat{f}_n^2(x)} - \frac{||K||_2^2 f_X(x)}{f_X^2(x)} = \frac{1}{f_X^2(x)}\left(\hat{g}_n(x) - f_X(x)||K||_2^2 - 2||K||_2^2\left(\hat{f}_n(x) - f_X(x)\right)\right) + o_p\left(\frac{1}{\sqrt{nh_n^d}}\right) + O_p\left(\frac{h_n^2}{\sqrt{nh_n^d}}\right) + O\left(h_n^4\right). \tag{A.3}$$

In the same manner, write

$$nh_n^d\sum_{i=1}^n w_{n,i}^2(x)\left\{\sigma^2(X_i) - \sigma^2(x)\right\} = \frac{1}{f_X^2(x)}\frac{h_n^d}{n}\sum_{i=1}^n K_h^2(X_i - x)\left(\sigma^2(X_i) - \sigma^2(x)\right) + \left(\frac{1}{\hat{f}_n^2(x)} - \frac{1}{E\left[\hat{f}_n(x)\right]^2}\right)\frac{h_n^d}{n}\left(\sum_{i=1}^n K_h^2(X_i - x)\left(\sigma^2(X_i) - \sigma^2(x)\right)\right). \tag{A.4}$$

Since $\frac{1}{nh_n^d}\sum_{i=1}^n K^2\left(\frac{X_i - x}{h_n}\right)(\sigma^2(X_i) - \sigma(x)) = O_p(h_n^2)$ in view of the weak law of large numbers and (A.9) below, the second term of (A.4) is of order $O_p(h_n^4)$. Therefore,

$$nh_n^d\sum_{i=1}^n w_{n,i}^2(x)\left\{\sigma^2(X_i) - \sigma^2(x)\right\} = \frac{1}{f_X^2(x)}\frac{h_n^d}{n}\sum_{i=1}^n K_h^2(X_i - x)\left(\sigma^2(X_i) - \sigma^2(x)\right) + O_p\left(h_n^4\right). \tag{A.5}$$

In view of (A.1), (A.3), and (A.5), after rearranging all the stochastic terms to the right-hand side, one immediately obtains that

$$\begin{aligned}
&\hat{v}_{n,1} - v(x) - f_X^{-2}(x)\frac{h_n^d}{n}\sum_{i=1}^n E\left[K_h^2(X_i - x)\{\sigma^2(X_i) - \sigma^2(x)\}\right] \\
&-\frac{\sigma^2(x)}{f_X^2(x)}\left(E\left[\hat{g}_n(x)\right] - f_X(x)||K||_2^2 - 2||K||_2^2\{E\left[\hat{f}_n(x)\right] - f_X(x)\}\right) = nh_n^d\sum_{i=1}^n w_{n,i}^2(x)\{\hat{\epsilon}_i^2 - \epsilon_i^2\} \\
&+\frac{1}{f_X^2(x)}\frac{h_n^d}{n}\sum_{i=1}^n\left(K_h^2(X_i - x)\{\sigma^2(X_i) - \sigma^2(x)\} - E\left[K_h^2(X_i - x)\{\sigma^2(X_i) - \sigma^2(x)\}\right]\right) \\
&+\frac{1}{f_X^2(x)}\frac{h_n^d}{n}\sum_{i=1}^n K_h^2(X_i - x)\{\epsilon_i^2 - \sigma^2(X_i)\} + \frac{\sigma^2(x)}{f_X^2(x)}\left(\hat{g}_n(x) - E\left[\hat{g}_n(x)\right] - 2||K||_2^2\{\hat{f}_n(x) - E\left[\hat{f}_n(x)\right]\}\right) \\
&+ o_p\left(\frac{1}{\sqrt{nh_n^d}}\right) + O_p\left(\frac{h_n^2}{\sqrt{nh_n^d}}\right) + O_p\left(h_n^4\right).
\end{aligned} \tag{A.6}$$

First, we examine the non-stochastic terms on the left-hand side of (A.6): Let

$$\mathcal{A}_{n,1} = \frac{h_n^d}{n}\sum_{i=1}^n E\left[K_h^2(X_i - x)\{\sigma^2(X_i) - \sigma^2(x)\}\right].$$

By a Taylor's expansion and Assumption 3.4,

$$\sigma^2(X_i) - \sigma^2(x) = \sum_{|\beta|=1}\frac{1}{\beta!}D^\beta\sigma^2(x)(X_i - x)^\beta + \sum_{|\beta|=1}\left(\frac{1}{\beta!}\int_0^1\left\{D^\beta\sigma^2(x + v(X_i - x)) - D^\beta\sigma^2(x)\right\}dv\right)(X_i - x)^\beta, \tag{A.7}$$

we can show that

$$\mathcal{A}_{n,1} = h_n\sum_{|\beta|=1}\frac{1}{\beta!}D^\beta\sigma^2(x)\int u^\beta K^2(u)f_X(x + h_n u)du + h_n\sum_{|\beta|=1}\frac{1}{\beta!}\int u^\beta K^2(u)f_X(x + h_n u)\int_0^1\left\{D^\beta\sigma^2(x + h_n vu) - D^\beta\sigma^2(x)\right\}dvdu = \mathcal{A}_{n,1,a} + \mathcal{A}_{n,1,b}.$$

Notice that

$$f_X(x+h_n u) = f_X(x) + \sum_{|\gamma|=1} \frac{1}{\gamma!} D^\gamma f_X(x) h_n u^\gamma + \sum_{|\gamma|=1} \left(\frac{1}{\gamma!} \int_0^1 \{ D^\gamma f_X(x+h_n v u) - D^\gamma f_X(x) \} dv \right) h_n u^\gamma. \tag{A.8}$$

In view of Assumptions 3.1 and 3.3, $\sum_{|\beta|=1} \int u^\beta K^2(u) = 0$ because $K^2(u)$ is also symmetric. One can then obtain that

$$\mathcal{A}_{n,1,a} = h_n^2 \sum_{|\beta|=1} \sum_{|\gamma|=1} \frac{1}{\beta!\gamma!} D^\beta \sigma^2(x) D^\gamma f_X(x) \int u^{\beta+\gamma} K^2(u) du + o(h_n^2).$$

Moreover, by Assumption 3.4 and a Taylor's expansion,

$$D^\beta \sigma^2(x+h_n v u) - D^\beta \sigma^2(x) = \sum_{|\alpha|=1} \frac{1}{\alpha!} D^{\alpha+\beta} \sigma^2(x) h_n v u^\alpha + \sum_{|\alpha|=1} \left(\frac{1}{\alpha!} \int_0^1 \{ D^{\alpha+\beta} \sigma^2(x+h_n v \xi u) - D^{\alpha+\beta} \sigma^2(x) \} d\xi \right) h_n v u^\alpha.$$

Therefore, in view of (A.8), we obtain

$$\mathcal{A}_{n,1,b} = \frac{1}{2} h_n^2 f_X(x) \sum_{|\alpha|=1} \sum_{|\beta|=1} \frac{1}{\alpha!\beta!} D^{\alpha+\beta} \sigma^2(x) \int u^{\alpha+\beta} K^2(u) du + o(h_n^2).$$

It follows that

$$\mathcal{A}_{n,1} = \frac{1}{2} h_n^2 \sum_{|\alpha|=1} \sum_{|\beta|=1} \frac{1}{\alpha!\beta!} \int u^{\alpha+\beta} K^2(u) du \{ f_X(x) D^{\alpha+\beta} \sigma^2(x) + 2 D^\alpha f_X(x) D^\beta \sigma^2(x) \}. \tag{A.9}$$

Let

$$\mathcal{A}_{n,2} = E[\hat{g}_n(x)] - f_X(x) ||K||_2^2.$$

Therefore, in view of (A.8),

$$\mathcal{A}_{n,2} = h_n \int K^2(u) \left\{ \sum_{|\beta|=1} \frac{1}{\beta!} D^\beta f_X(x) u^\beta + \sum_{|\beta|=1} \left(\frac{1}{\beta!} \int_0^1 \{ D^\beta f_X(x+h_n v u) - D^\beta f_X(x) \} dv \right) u^\beta \right\} du.$$

Moreover, in view of Assumption 3.3 and a Taylor's expansion, one can show that

$$D^\beta f_X(x+h_n v u) - D^\beta f_X(x) = \sum_{|\alpha|=1} \frac{1}{\alpha!} D^{\alpha+\beta} f_X(x) h_n v u^\alpha + \sum_{|\alpha|=1} \left(\frac{1}{\alpha!} \int_0^1 \{ D^{\alpha+\beta} f_X(x+\xi h_n v u) - D^{\alpha+\beta} f_X(x) \} d\xi \right) h_n v u^\alpha. \tag{A.10}$$

Therefore, in view of Assumptions 3.1 and 3.3, we have

$$\mathcal{A}_{n,2} = \frac{h_n^2}{2} \sum_{|\alpha|=1} \sum_{|\beta|=1} \frac{1}{\alpha!\beta!} D^{\alpha+\beta} f_X(x) \int u^{\alpha+\beta} K^2(u) du + o(h_n^2). \tag{A.11}$$

It can also be shown from (A.10) that

$$\mathcal{A}_{n,3} = E\left[\hat{f}_n(x)\right] - f_X(x) = \frac{h_n^2}{2} \sum_{|\alpha|=1} \sum_{|\beta|=1} \frac{1}{\alpha!\beta!} D^{\alpha+\beta} f_X(x) \int u^{\alpha+\beta} K(u) du + o(h_n^2). \tag{A.12}$$

Next, we examine the stochastic terms on the right-hand side (RHS) of (A.6).

$$\mathcal{B}_{n,1} = n h_n^d \sum_{i=1}^n w_{n,i}^2(x) \{ \hat{\epsilon}_i^2 - \epsilon_i^2 \} = 2 n h_n^d \sum_{i=1}^n w_{n,i}^2(x) (m_0(X_i) - \hat{m}(X_i)) \epsilon_i + n h_n^d \sum_{i=1}^n w_{n,i}^2(x) [\hat{m}(X_i) - m_0(X_i)]^2 = 2\mathcal{B}_{n,1,a} + \mathcal{B}_{n,1,b}.$$

Notice that, by the uniform consistency of $\hat{f}_n(x)$,

$$\mathcal{B}_{n,1,a} = \left(f_X^{-2}(x) + o_{a.s.}(1)\right) \frac{1}{nh_n^d} \sum_{i=1}^{n} K^2\left(\frac{X_i - x}{h_n}\right)[m_0(X_i) - \hat{m}(X_i)]\epsilon_i.$$

We shall employ Lemmas 1 and 2 to prove that $\sqrt{nh_n^d}\mathcal{B}_{n,1,a} = o_p(1)$. Suppose that the function space $\mathcal{M}_n$ in Lemma 1 consists of functions with their domain in some compact set of $\mathbb{R}^d$; and all the k-order partial derivatives exist and are uniformly bounded by some multiple of n^{ξ^*} for some $k > d/2$ and $\xi^* > 0$. Then, Condition (1) holds with $\xi_n = \xi^*\alpha_m, \alpha_m = d/k$ [cf. van der Vaart and Wellner; 1996, sec. 2.7]. To verify that the NW estimator $\hat{m} \in \mathcal{M}_n$ with probability approaching 1, choose $h_n = n^{-\frac{1}{d+4}}$ for instance, by the same argument as in Lemma 2, it can be shown that all the k-order partial derivatives of $\hat{m}(x) - E[\hat{m}(x)]$ are stochastically bounded by $C_0 n^{\xi^*}$, where ξ^* is some positive constant. Moreover, Lemma 2 implies that the rate of convergence δ_n required by Lemma 1 can be set to $\max\left(\sqrt{\frac{n^\alpha}{nh_n^d}}, h_n, \frac{1}{n^\rho}\right)$. Setting $\eta_n = \delta^{\alpha_\eta}, \alpha_\eta \in (0, (2-\alpha_m)/2)$, in view of Assumptions 3.6 and 3.7(1)-3.7(2), one has $\sqrt{nh_n^d}\mathcal{B}_{n,1,a} = o_p(1)$. Next, notice that

$$\begin{aligned}
\mathcal{B}_{n,1,b} &\leq \max_{1\leq i\leq n} |\hat{m}(X_i) - m_0(X_i)|^2 nh_n^d \sum_{i=1}^{n} w_{n,i}^2(x) \\
&\leq \{f_X^{-1}(x) + o_{a.s.}(1)\} \int K^2(u)du \max_{1\leq i\leq n} |\hat{m}(X_i) - m_0(X_i)|^2
\end{aligned}$$

uniformly in x. Invoking Lemma 2, we have

$$\sqrt{nh_n^d}\mathcal{B}_{n,1,b} = o_p\left(\frac{1}{n^{1/2-\alpha}h_n^{d/2}}\right) + O\left(n^{1/2}h_n^{d/2+4}\right) + o_p\left(\frac{h_n^{d/2}}{n^{2\rho-1/2}}\right).$$

It then follows from Assumption 3.7(1) that $\sqrt{nh_n^d}\mathcal{B}_{n,1,b} = o_p(1)$. Thus,

$$\sqrt{nh_n^d}\mathcal{B}_{n,1} = o_p(1). \tag{A.13}$$

Define the second stochastic term on the RHS of (A.6) as

$$\mathcal{B}_{n,2} = \frac{h_n^d}{n}\sum_{i=1}^{n}\left(K_h^2(X_i - x)\{\sigma^2(X_i) - \sigma^2(x)\} - E\left[K_h^2(X_i - x)\{\sigma^2(X_i) - \sigma^2(x)\}\right]\right).$$

By the same argument employed to obtain $\mathcal{A}_{n,1}$ above, one can show that

$$\frac{h_n^{2d}}{n^2}\sum_{i=1}^{n} E\left[K_h^4(X_i - x)\left(\sigma^2(X_i) - \sigma^2(x)\right)^2\right] = O\left(\frac{1}{nh_n^{d-2}}\right)$$

and, in view of (A.9),

$$\frac{h_n^{2d}}{n^2}\sum_{i=1}^{n} E^2\left[K_h^2(X_i - x)\{\sigma^2(X_i) - \sigma^2(x)\}\right] = O\left(\frac{h_n^4}{n}\right).$$

It then follows that $nh_n^d E\left[\mathcal{B}_{n,2}^2\right] = O(h_n^2 + h_n^{d+4})$. As $E[\mathcal{B}_{n,2}] = 0$, it then follows that

$$\sqrt{nh_n^d}\mathcal{B}_{n,2} = o_p(1). \tag{A.14}$$

For remaining stochastic terms on the RHS of (A.6), let $T_{n,2}(x) = n^{-1}h_n^d\sum_{i=1}^n K_h^2(X_i - x)\{\epsilon_i^2 - \sigma^2(X_i)\}$; $T_{n,3}(x) = \hat{f}_n(x) - E[\hat{f}_n(x)]$; and $T_{n,4}(x) = \hat{g}_n(x) - E[\hat{g}_n(x)]$. In view of (A.9), (A.11), (A.12)-(A.14) together with Assumption 3.7(1), the main theorem then follows by applying Lemma 3.

Proof of Theorem 2: See the online supplemental material of this paper.
Proof of Theorem 3: See the online supplemental material of this paper.
Proof of Theorem 4: See the online supplemental material of this paper.
Proof of Theorem 5: See the online supplemental material of this paper.

B. Auxiliary Lemmas

Lemma 1. *Suppose that $\hat{m}(x)$ is some nonparametric estimate of $m_0(x)$ such that $||\hat{m} - m_0||_\infty = o_p(\delta_n), \delta_n \downarrow 0$. There exist a sequence of sets $\mathcal{M}_n$ such that $P(\hat{m} \in \mathcal{M}_n) \to 1$ as $n \to \infty$. Moreover,*

1. *For a constant,* $C_M > 0$, *and a function,* $m_n(x)$, *with* $||m_n - m_0||_\infty = o(\delta_n)$, *the sequence of sets* $\bar{\mathcal{M}}_n = \mathcal{M}_n \cap \{m: \ ||m - m_n||_\infty \le \delta_n\}$ *can be covered by at most* $C_M \exp(\lambda^{-\alpha_m}\xi_n)$ *balls with* $||\cdot||_\infty$*−radius* λ *for all* $\lambda \le \delta_n$ *and some sequence,* ξ_n, *such that* $\xi_n \delta_n^{-\alpha_m} \uparrow \infty$, *where* $\alpha_m \in (0, 2]$.
2. $\frac{\eta_n^2}{\xi_n \delta_n^{2-\alpha_m}} \uparrow \infty$; $\frac{\eta_n \sqrt{nh_n^d}}{\xi_n \delta_n^{1-\alpha_m} \log(n)} \uparrow \infty$; *and* $\frac{\delta_n h_n^{d/2}}{\eta_n n^{1/2}} (\min\left\{\frac{\eta_n^2}{\xi_n \delta_n^{2-\alpha_m}}, \frac{\eta_n \sqrt{nh_n^d}}{\xi_n \delta^{1-\alpha_m} \log(n)}\right\})^{\frac{1-c_*}{2-\alpha_m-2a}} \downarrow 0$ *for some* $a > 1$ *and* $c_* \in (0, 1)$, *where* η_n *is some decaying sequence of constants.*

Let Assumptions 3.1 and 3.2 hold. Then,

$$\frac{1}{\sqrt{nh_n^d}} \left| \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \{\hat{m}(X_i) - m_0(X_i)\}\epsilon_i \right| = O_p(\eta_n). \tag{B.1}$$

Proof of Lemma 1: See the online supplemental material of this paper.

Lemma 2. *Let* $\hat{m}(x)$ *be the NW estimate of* $m_0(x)$. *Suppose that Assumptions 3.1, 3.3, and 3.5 hold. Moreover,*

1. $E\left[||X_i||^{\rho+1}\right] < \infty$ *and* $E\left[|Y_i|^p\right] < \infty$ *for some* $\rho > 0$ *and* $p > 2$.
2. $n^{1-\alpha} h_n^{\frac{p+2}{p-2}d} \uparrow \infty$ *for some* $\alpha \in (0, 1)$; $\frac{n^{1-2/p} h_n^d}{\log(n)} \uparrow \infty$.

Then,
$\max_{1 \le i \le n} |\hat{m}(X_i) - m_0(X_i)| = o_p\left(\sqrt{\frac{n^\alpha}{nh_n^d}}\right) + O(h_n^2) + o_p\left(\frac{1}{n^\rho}\right)$.

Proof of Lemma 2. See the online supplemental material of this paper.

Lemma 3. Consider *the random vector*

$$T_n(x) = \begin{bmatrix} T_{n,0}(x) \\ T_{n,1}(x) \\ T_{n,2}(x) \\ T_{n,3}(x) \\ T_{n,4}(x) \end{bmatrix} = \begin{bmatrix} \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \varepsilon_i \\ \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \sigma^2(X_i)\eta_i \\ \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)^2 \sigma^2(X_i)\eta_i \\ \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) - EK\left(\frac{x - X_i}{h_n}\right) \\ \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)^2 - EK\left(\frac{x - X_i}{h_n}\right)^2 \end{bmatrix},$$

where $\eta_i = \epsilon_i^2/\sigma^2(X_i) - 1$. *Then,*

$$\sqrt{nh_n^d} T_n(x) \Rightarrow N(0, \Omega(x))$$

$$\Omega(x) = f_X(x) \begin{bmatrix} \mu_2(K)\sigma^2(x) & \mu_2(K)\sigma^3(x)m_3(x) & \mu_3(K)\sigma^3(x)m_3(x) & 0 & 0 \\ \mu_2(K)\sigma^3(x)m_3(x) & \mu_2(K)\sigma^4(x)m_4(x) & \mu_3(K)\sigma^4(x)m_4(x) & 0 & 0 \\ \mu_3(K)\sigma^3(x)m_3(x) & \mu_3(K)\sigma^4(x)m_4(x) & \mu_4(K)\sigma^4(x)m_4(x) & 0 & 0 \\ 0 & 0 & 0 & \mu_2(K) & \mu_3(K) \\ 0 & 0 & 0 & \mu_3(K) & \mu_4(K) \end{bmatrix}.$$

Proof of Lemma 3. This follows by applying the Cramér-Wold device, the Lyapunov CLT and Bochner's Lemma.

References

Ai, C. (1997). A semiparametric maximum likelihood estimator, *Econometrica* 65(4): 933–963. doi:10.2307/2171945

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59(3): 817–858. doi:10.2307/2938229

Bandi, F. M. and Phillips, P. C. B. (2003). Fully nonparametric estimation of scalar difussion models, *Econometrica* 71(1): 241–283. doi:10.1111/1468-0262.00395

Chesher, A. and Jewitt, I. (1987). The bias of the heteroskedasticity consistent covariance matrix estimator, *Econometrica* 55(5): 1217–1222. doi:10.2307/1911269

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information, *Biometrika* 65: 457–481. doi:10.1093/biomet/65.3.457

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, University of California Press, Berkeley, California, pp. 59–82.

Escanciano, J. C., Jacho-Chávez, D. T. and Lewbel, A. (2014). Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing, *Journal of Econometrics* 178: 426–443. doi:10.1016/j.jeconom.2013.06.004

Escanciano, J. C., Jacho-Chávez, D. T. and Lewbel, A. (2016). Identification and estimation of semiparametric two step models, *Quantitative Economics* 7(2): 561–589. doi:10.3982/QE328

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Vol. 66 of *Monographs on Statistics and Applied Probability*, 1 edn, Chapman and Hall.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression, *Biometrika* 85(3): 645–660. doi:10.1093/biomet/85.3.645

Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression, *The Annals of Statistics* 20(2): 695–711. doi:10.1214/aos/1176348652

Hall, P. and Yatchew, A. (2005). Unified approach to testing functional hypotheses in semiparametric contexts, *Journal of Econometrics* 127(2): 225–252. doi:10.1016/j.jeconom.2004.08.005

Härdle, W. (1990). *Applied Nonparametric Regression*, Vol. 9 of *Econometric Society Monographs*, 1 edn, Cambridge University Press.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Vol. 43 of *Monographs on Statistics and Applied Probability*, 1 edn, Chapman and Hall.

Hayfield, T. and Racine, J. S. (2008). Nonparametric econometrics: The np package, *Journal of Statistical Software* 27(5): 1–32. doi:10.18637/jss.v027.i05

Henderson, D. J. and Parmeter, C. F. (2015). *Applied Nonparametric Econometrics*, Cambridge University Press, New York.

Ho, A. T. Y., Huynh, K. P. and Jacho-Chávez, D. T. (2014). crs: A package for nonparametric spline estimation in r, *Journal of Applied Econometrics* 29(2): 348–352. doi:10.1002/jae.2381

Huynh, K. P. and Jacho-Chávez, D. T. (2009a). Growth and governance: A nonparametric analysis, *Journal of Comparative Economics* 37(1): 121–143. doi:10.1016/j.jce.2008.08.003

Huynh, K. P. and Jacho-Chávez, D. T. (2009b). A nonparametric quantile analysis of growth and goverance, in Q. Li and J. S. Racine (eds), *Advances in Econometrics: Nonparametric Econometric Methods*, Vol. 25, Emerald Group Publishing.

Jeffrey, A., Kristensen, D., Linton, O., Nguyen, T. and Phillips, P. C. B. (2004). Nonparametric estimation of a multifactor heath-jarrow-morton model: An integrated approach, *Journal of Financial Econometrics* 2: 251–289. doi:10.1093/jjfinec/nbh010

Kauermann, G. and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation, *Journal of the American Statistical Association* 96(456): 1387–1396. doi:10.1198/016214501753382309

Li, D., Phillips, P. C. B. and Gao, J. (2016). Uniform consistency of nonstationary kernel-weighted sample covariances for nonparametric regression, *Econometric Theory* 32(3): 655–685. doi:10.1017/S0266466615000109

Linton, O. B. and Jacho-Chávez, D. T. (2010). On internally corrected and symmetrized kernel estimators for nonparametric regression, *Test* 19(1): 166–186. doi:10.1007/s11749-009-0145-y

Mack, Y. P. and Müller, H.-G. (1989). Derivative estimation in nonparametric regression with random predictor variable, *Sankhya: The Indian Journal of Statistics* 51(Series A): 59–72.

Mammen, E., Rothe, C. and Schienle, M. (2012). Nonparamtric regression with nonparametrically generated covariates, *Annals of Statistics* 40(2): 1132–1170. doi:10.1214/12-AOS995

Newey, W. K. and West, K. D. (1987). A simple positive definite heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55(3): 703–708. doi:10.2307/1913610

Phillips, P. C. B. (2005). Hac estimation by automated regression, *Econometric Theory* 21(1): 116–142. doi:10.1017/S0266466605050085

Phillips, P. C. B. (2007). Regression with slowly varying regressors and nonlinear trends, *Econometric Theory* 23(4): 557–614. doi:10.1017/S0266466607070260
Phillips, P. C. B. (2009). Local limit theory and spurious nonparametric regression, *Econometric Theory* 25(6): 1466–1497. doi:10.1017/S0266466609990223
Phillips, P. C. B. and Su, L. (2011). Nonparametric regression under location shifts, *Econometrics Journal* 14(3): 457–486. doi:10.1111/j.1368-423X.2011.00344.x
Phillips, P. C. B., Sun, Y. and Jin, S. (2006). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation, *International Economic Review* 47(3): 837–894. doi:10.1111/j.1468-2354.2006.00398.x
Phillips, P. C. B., Sun, Y. and Jin, S. (2007). Run variance estimation and robust regression testing using sharp origin kernels with no truncation, *Journal of Statistical Planning and Inference* 137(3): 985–1023. doi:10.1016/j.jspi.2006.06.033
Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation*, Academic Press, New York.
Robinson, P. M. (1983). Nonparametric estimators for time series, *Journal of Time Series Analysis* 4: 185–206. doi:10.1111/j.1467-9892.1983.tb00368.x
Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*, Birkhäuser, Boston, Basel, Stuttgart.
Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, 1 edn, Chapman and Hall, London.
Sul, D., Phillips, P. C. B. and Choi, C.-Y. (2005). Prewhitening bias in hac estimation, *Oxford Bulletin of Economics and Statistics* 67(4): 517–546. doi:10.1111/j.1468-0084.2005.00130.x
Sun, Y., Phillips, P. C. B. and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing, *Econometrica* 76(1): 175–194. doi:10.1111/j.0012-9682.2008.00822.x
van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.
Vogt, M. (2012). Nonparametric regression for locally stationary time series, *Annals of Statistics* 40(5): 2601–2633. doi:10.1214/12-AOS1043
Wang, Q. and Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression, *Econometric Theory* 25(3): 710–738. doi:10.1017/S0266466608090269
Wang, Q. and Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression, *Econometrica* 77(6): 1901–1948.
Wang, Q. and Phillips, P. C. B. (2011). Asymptotic theory for zero energy density estimation with nonparametric regression applications, *Econometric Theory* 27(2): 235–259. doi:10.1017/S0266466610000277
White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48(4): 817–38. doi:10.2307/1912934
Xu, K.-L. and Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances, *Journal of Econometrics* 142(1): 265–280. doi:10.1016/j.jeconom.2007.06.001
Xu, K.-L. and Phillips, P. C. B. (2011). Tilted nonparametric estimation of volatility functions with empirical applications, *Journal of Business and Economic Statistics* 29(4): 518–528. doi:10.1198/jbes.2011.09012
Yatchew, A. (2003). *Semiparametric Regression for the Applied Econometrician*, Themes in Modern Econometrics, 1 edn, Cambridge University Press.
Yu, K. and Jones, M. C. (2004). Likelihood-based local linear estimation of the conditional variance function, *Journal of the American Statistical Association* 99(465): 139–144. doi:10.1198/016214504000000133