INTRODUCTION TO THE VARIANCE-STABILIZING BANDWIDTH FOR THE NADARAYA-WATSON REGRESSION ESTIMATOR

Nishida, Kiheiji
Graduate School of Systems and Information Engineering, University of Tsukuba : Researcher

Kanazawa, Yuichiro
Department of Social Systems and Management, Graduate School of Systems and Information Engineering University of Tsukuba : Professor

INTRODUCTION TO THE VARIANCE-STABILIZING BANDWIDTH FOR THE NADARAYA-WATSON REGRESSION ESTIMATOR

by

Kiheiji NISHIDA   and   **Yuichiro KANAZAWA**

➵•◆•➵

FUKUOKA, JAPAN
2011

INTRODUCTION TO THE VARIANCE-STABILIZING BANDWIDTH FOR THE NADARAYA-WATSON REGRESSION ESTIMATOR

By

Kiheiji NISHIDA[*]   and   **Yuichiro KANAZAWA**[†]

Abstract

In linear regression under heteroscedastic variances, Aitken estimator is employed to account for the differences in variances. Employing the same principle, we propose the Nadaraya-Watson regression estimator with variable variance-stabilizing bandwidth (VS bandwidth) that minimizes asymptotic MISE (AMISE) while maintaining asymptotic homoscedasticity. We examine its local and global properties relative to the MISE minimizing fixed bandwidth. The proposed VS bandwidth produces the asymptotic variance smaller on some part of the support than the fixed bandwidth and may in some cases achieve smaller AMISE than its fixed counterpart. In numerical examples, we find that the proposed VS bandwidth is more serviceable when the distribution of X's are flatter.

Key Words and Phrases: Aitken estimator, Bandwidth selection, The Nadaraya-Watson estimator, Variable bandwidth, Variance-stabilization.

1. Introduction

Suppose we construct a Nadaraya-Watson regression estimator (henceforth the NW estimator, Nadaraya, 1964, 1965, 1970; Watson, 1964; Watson and Leadbetter, 1963) from a pair of observations (x_i, y_i) $i = 1, 2, \ldots, n$. This simple setup is well suited to briefly explain the mechanism of variance-stabilization as stated later in (7) and the relative magnitude of the mean integrated squared error to be determined as stated in (13). The response Y_i is influenced by the explanatory variable X_i in the form of $m(X_i)$ and the disturbance U_i as

$$Y_i = m(X_i) + U_i, \tag{1}$$

where the X_i's, $i = 1, 2, ..., n$ are i.i.d. random variables whose marginal density is defined as $f_X(x)$ on support I and U_i's, $i = 1, 2, ..., n$ are identically distributed random variables conditionally independent upon X_i and independent of X_j, $i \neq j$. We assume that their conditional moments are,

$$E_{U_i|X_i}[U_i|X_i = x] = 0, \quad E_{U_i|X_i}[U_i^2|X_i = x] = \sigma^2(x) < \infty.$$

[*] Researcher, Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Ten-noh-dai, Tsukuba, Ibaraki 305-8573, Japan. kiheiji.nishida@gmail.com

[†] Professor, Department of Social Systems and Management, Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Ten-noh-dai, Tsukuba, Ibaraki 305-8573, Japan. kanazawa@sk.tsukuba.ac.jp

Under these assumptions, the function $m(X_i)$ in (1) is expressed as the conditional expectation $m(X_i) = E_{Y_i|X_i}(Y_i|X_i) = \int y f_{Y|X}(y|x)dy = \int y f_{X,Y}(x,y)/f_X(x)dy$, where $f_{X,Y}(x,y)$ is bivariate density function of (X,Y). By replacing these $f_{X,Y}(x,y)$ and $f_X(x)$ with their kernel bivariate and univariate estimates, the former with multiplicative kernel $K_{X,Y}(x,y) = K_X(x)K_Y(y)$ of respective bandwidths h_x and h_y on X and Y and the latter of kernel $K_X(x)$ of the bandwidth h_x on X, we arrive at the usual NW estimator at $X = x$ as $\widehat{m}_{h_x}(x) = \sum_{i=1}^{n} K_X\left(\frac{x-X_i}{h_x}\right)Y_i / \sum_{i=1}^{n} K_X\left(\frac{x-X_i}{h_x}\right)$. We write the kernel function $K_X(t)$ as $K(t)$ for brevity.

Variance and bias of the NW estimator have been well known (see e.g. Pagan and Ullah 1999). With the standard set of assumptions on kernel **K1-K3** and the additional assumptions **A1-A6** on $f_X(x)$, $\sigma^2(x)$, $m(x)$ and h_x in section 2., the theoretical bias of the NW estimator $\widehat{m}_{h_x}(x)$ at x is

$$E_{\mathbf{X},\mathbf{Y}}[\widehat{m}_{h_x}(x)] - m(x) = \frac{h_x^2 \alpha(x)}{2f_X(x)}\left[\int t^2 K(t)dt\right] + O\left(\frac{1}{nh_x}\right) + o(h_x^2), \qquad (2)$$

where, for notational convenience, we write

$$\alpha(x) = 2m^{(1)}(x)f_X^{(1)}(x) + m^{(2)}(x)f_X(x).$$

Similarly the theoretical variance at x is known to be,

$$V_{\mathbf{X},\mathbf{Y}}[\widehat{m}_{h_x}(x)] = \frac{1}{nh_x} \cdot \frac{\sigma^2(x)}{f_X(x)}\left[\int K^2(t)dt\right] + O\left(\frac{1}{n}\right) + o\left(\frac{1}{nh_x}\right). \qquad (3)$$

Finding bandwidth selection rule is an important issue in nonparametric regression. Härdle et al. said, "[W]e are faced with the problem of finding a bandwidth-selection rule that has desirable theoretical properties and is applicable in practice" (Härdle et.al., 2004, p.94). One of such desirable theoretical properties is pointwise convergence in probability of the proposed estimator to the underlying regression function, which naturally leads to a bandwidth selection rule that minimizes the mean squared error (MSE) at a single point. But if we are interested in how well the entire $m(x)$ performs over the support, then we would rather use the global measure of closeness, which leads to the mean integrated squared error (MISE),

$$
\begin{aligned}
&MISE(m(x), \widehat{m}_{h_x}(x)) \\
&= \int_I \left[\int_I \cdots \int_I [m(x) - \widehat{m}_{h_x}(x)]^2 \prod_{i=1}^{n} f_{X_i,Y_i}(x_i,y_i)dx_i dy_i\right] f_X(x)dx \qquad (4)
\end{aligned}
$$

and the fixed bandwidth so obtained is

$$h_{fixed} = \left[\frac{\left[\int K^2(t)dt\right]\int_I \sigma^2(x)dx}{\left[\int t^2 K(t)dt\right]^2 \int_I \frac{\alpha^2(x)}{f_X(x)}dx}\right]^{\frac{1}{5}} n^{-\frac{1}{5}}. \qquad (5)$$

It is known that the NW estimator with this fixed bandwidth converges in mean square and therefore converges in probability pointwise to $m(x)$ with additional assumptions

K4-K5 (see e.g. Härdle et.al., 2004, pp.93-94). Substituting (5) for (3) gives the leading term of the theoretical variance (henceforth asymptotic variance or abbreviated as AV),

$$AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{fixed}}(x)\right] = \frac{\sigma^2(x)}{f_X(x)}\left[\frac{\left[\int K^2(t)dt\right]^{\frac{4}{5}}\left[\int_I \sigma^2(x)dx\right]^{-\frac{1}{5}}}{\left[\int t^2 K(t)dt\right]^{-\frac{2}{5}}\left[\int_I \frac{\alpha^2(x)}{f_X(x)}dx\right]^{-\frac{1}{5}}}\right]n^{-\frac{4}{5}}. \quad (6)$$

The MISE minimizing fixed bandwidth in (5) does not achieve asymptotic homoscedasticity of the $\widehat{m}_{h_x}(x)$ up to the leading term (henceforth asymptotic homoscedasticity) unless $f_X(x)$ and $\sigma^2(x)$ are of the same functional form on I. To numerically illustrate the point, we consider two non-linear cases where regression function is $m(x) = 4x^3$. We suppose that the data are clustered around the center of the support $I = [0,1]$ and they tapered off towards both tails with density $f_X(x)$ is the normal $N(0.5, 0.04)$ truncated on I. As a kernel function, we use the Epanechnikov kernel.

Let us first see a case where the conditional variance $\sigma^2(x)$ stays constant, say, at 0.25. See the left panel of Figure 1 for the $m(x)$, $m(x) \pm 1.96 AV_{X_i,Y_i}^{1/2}\left[\widehat{m}_{h_{fixed}}(x)\right]$ and $f_X(x)$ along with one set of randomly generated (X_i, Y_i) $i = 1, \ldots, 1000$. Consider next an often observed case where the conditional variance increases as $m(x)$ increases, say, $\sigma^2(x) = (x + 0.5)^2$ as in the left panel of Figure 2. In both of these panels, we observe that the asymptotic variances in (6) are larger at both tails.
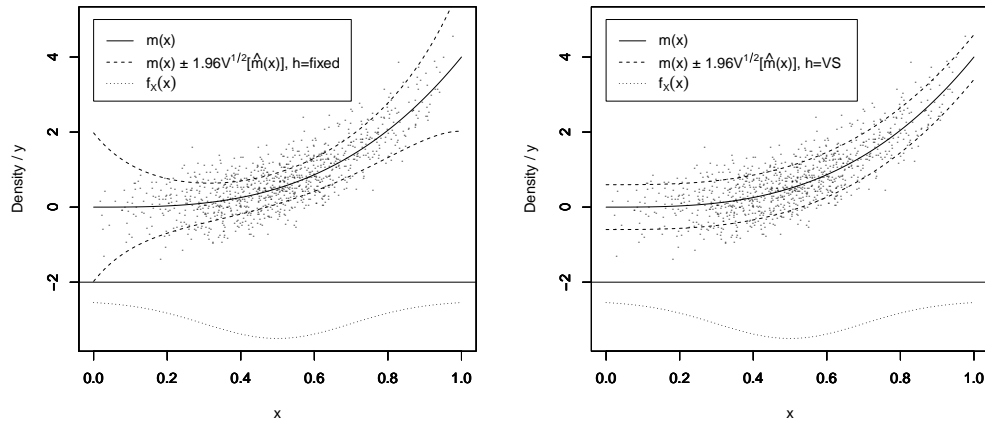

Figure 1: Plots of the asymptotic variances. Conditional variance $\sigma^2(x)$ stays constant at 0.25, $m(x) = 4x^3$ and $f_X(x)$ is a normal distribution N(0.5, 0.04) truncated on [0, 1].

In linear regression under non-spherical or heteroscedastic variances, generalized least-squares or Aitken estimator is employed to account for the differences in variances at different x's. We feel that the NW estimators can be improved by employing the same principle of variance-stabilization. If we let the bandwidth h_x vary, but in the form,

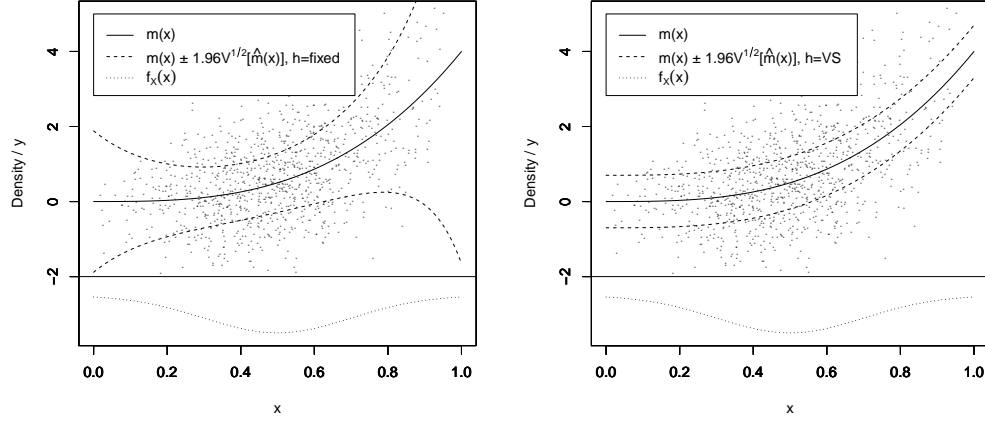$$h_{VS}(x) = h_0 \cdot \frac{\sigma^2(x)}{f_X(x)}, \quad (7)$$

Figure 2: Plots of the asymptotic variances.
Conditional variance $\sigma^2(x)$ increases as $m(x)$ increases as $\sigma^2(x) = (x+0.5)^2$, $m(x) = 4x^3$ and $f_X(x)$ is a normal distribution N(0.5, 0.04) truncated on [0, 1].

where h_0 is constant possibly dependent on the kernel $K(t)$, the regression function $m(x)$, the density $f_X(x)$ of X and the conditional variance function $\sigma^2(x)$, averaged over I, it achieves asymptotic homoscedasticity. Hence, variable bandwidth of the form (7) is variance-stabilizing (henceforth VS).

This variable VS bandwidth can be justified for two reasons: First in the domain where the density $f_X(x)$ is low, (7) forces one to choose wider bandwidth. The $f_X(x)$ being low means that there are relatively few data points on the region, so aggregating them over wider region to estimate the regression function $m(x)$ makes sense. Second in the region where the variances $\sigma^2(x)$ are low, the bandwidths should be narrower because the small variances imply that those data points are more accurate there.

To determine h_0 in (7), we choose h_0 so as to minimize the MISE among the class of bandwidths (7). The variable VS bandwidth so obtained is

$$h_{VS}(x) \;\; = \;\; \left[\frac{\left[\int K^2(t)dt \right]}{\left[\int t^2 K(t)dt \right]^2 \left[\int_I \frac{\sigma^8(x)\alpha^2(x)}{f_X^5(x)} dx \right]} \right]^{\frac{1}{5}} n^{-\frac{1}{5}} \cdot \frac{\sigma^2(x)}{f_X(x)}, \qquad (8)$$

and the corresponding asymptotic variance is

$$AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{VS}}(x) \right] = \left[\frac{\left[\int K^2(t)dt \right]^{\frac{4}{5}}}{\left[\int t^2 K(t)dt \right]^{-\frac{2}{5}} \left[\int_I \frac{\sigma^8(x)\alpha^2(x)}{f_X^5(x)} dx \right]^{-\frac{1}{5}}} \right] n^{-\frac{4}{5}}. \qquad (9)$$

The derivation of h_0 is in Appendix. It is easy to see that the proposed bandwidth enables $\widehat{m}_{h_{VS}}(x)$ to converge in probability to $m(x)$ pointwise with conditions **K1**-**K5** and **A1**-**A6** because MSE at every point of x is the same order of magnitude $n^{-4/5}$ of the MISE minimizing fixed bandwidth in (5).

To numerically illustrate the effect of this variance stabilization, see the right panels of Figure 1 where $\sigma^2(x) = 0.25$, and of Figure 2 where $\sigma^2(x) = (x + 0.5)^2$, for the aforementioned $m(x)$, $m(x) \pm 1.96AV_{X_i,Y_i}^{1/2}[\widehat{m}_{h_{VS}}(x)]$ and $f_X(x)$ along with the same set of randomly generated (X_i, Y_i) $i = 1, \ldots, 1000$. We observe that the asymptotic variance (9) stays constant on the domain.

Now we need to ask ourselves three questions: First how well the "stabilized" asymptotic variance in (9) fares relative to the asymptotic variance in (6) locally. Second how well the proposed variable VS bandwidth in (8) performs globally in comparison with the fixed bandwidth in (5) in terms of, say, the MISE. Even if these questions can be answered affirmatively, we wonder if the variance-stabilizing bandwidth in (8) is feasible in the sense that its sample-based criterion function such as the cross-validation statistic can be formulated.

In this paper, we answer the first two questions affirmatively in propositions 1, 2 and 3 in section 2.. As for the last question, it is beyond the scope of this paper given the available space. In section 3., we present two illustrative examples to demonstrate propostions 1 and 3 and to investigate how fat-tailedness of the distribution of X's affects MISE. In section 4., we give comment on the proposed VS bandwidth.

2. Asymptotic results

Suppose that the kernel function satisfy:

K1 Kernel is a density function $K(t)$ symmetric about zero.

K2 $\int t^2 K(t) dt < \infty$.

K3 $\int K^2(t) dt < \infty$.

K4 $\int |K(t)| dt < \infty$.

K5 $\lim_{t \to \pm\infty} tK(t) \to 0$.

We place the following standard set of assumptions:

A1 $h_x \to 0$ as n goes to infinity.

A2 $nh_x \to \infty$ as n goes to infinity.

A3 The density of X is $0 < f_X(x) < \infty$ on a compact support I.

A4 The $f_X(x)$ is bounded continuously differentiable on I. [1]

A5 The conditional variance $\sigma^2(x)$ is continuous with respect to x and $0 < \sigma^2(x) < \infty$ on I.

A6 The regression function $m(x)$ is twice bounded continuously differentiable [2] and is bounded and not identically constant on I.

---

[1] At the left (right) boundary, $f_X(x)$ is right (left) differentiable.
[2] At the left (right) boundary, $m(x)$ is twice right (left) differentiable.

Define the ratio of two "densities" of the $\sigma^2(x)$ to the $f_X(x)$ or

$$\gamma(x) = \frac{\sigma^2(x)}{\int_I \sigma^2(x)dx} \Big/ \frac{f_X(x)}{\int_I f_X(x)dx} \ . \tag{10}$$

Consequently, by **A3**, **A4** and **A5**, we can define $x_{max} = \arg\max_{x \in I} \gamma(x)$ and $x_{min} = \arg\min_{x \in I} \gamma(x)$. Note that the quantity $\gamma(x)$, in (10), by virtue of the fact that it is a ratio of two "densities," $\gamma(x_{min}) \leq 1 \leq \gamma(x_{max})$. It is easy to see that (6) is expressed as a function of $\gamma(x)$,

$$AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{fixed}}(x)\right] = \gamma(x) \cdot \left[\frac{\left[\int K^2(t)dt\right]^{\frac{4}{5}} \left[\int_I \sigma^2(x)dx\right]^{\frac{4}{5}}}{\left[\int t^2 K(t)dt\right]^{-\frac{2}{5}} \left[\int_I \frac{\alpha^2(x)}{f_X(x)}dx\right]^{-\frac{1}{5}}} \right] n^{-\frac{4}{5}},$$

where the term on the right hand side other than $\gamma(x)$ is constant. Then, we show in proposition 1 that the asymptotic homoscedastic variance in (9) is bounded above and below by its heteroscedastic counterparts in (6).

Proposition 1. *Asymptotic homoscedastic variance* $AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{VS}}(x)\right]$ *in* (9) *is bounded above and below by the maximal and the minimal heteroscedastic counterparts in* (6) *for all* $x \in I$, *or*

$$AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{fixed}}(x_{min})\right] \leq AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{VS}}(x)\right] \leq AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{fixed}}(x_{max})\right]. \tag{11}$$

Proof. Subtracting the leading term of (9) raised to the fifth from that of (6) obtains

$$AV^5_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{fixed}}(x)\right] - AV^5_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{VS}}(x)\right]$$
$$= C_0 \cdot n^{-4}\left[\gamma^5(x)\int_I \frac{\alpha^2(z)}{f_X(z)}dz - \int_I \gamma^4(z)\frac{\alpha^2(z)}{f_X(z)}dz\right], \tag{12}$$

where $C_0 = \left[\int K^2(t)dt\right]^4 \left[\int t^2 K(t)dt\right]^2 \left[\int_I \sigma^2(z)dz\right]^4 > 0$. Evaluating (12) at $x = x_{max}$, we obtain

$$AV^5_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{fixed}}(x_{max})\right] - AV^5_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{VS}}(x_{max})\right]$$
$$= C_0 \cdot n^{-4}\int_I \left[\gamma^5(x_{max}) - \gamma^4(z)\right]\frac{\alpha^2(z)}{f_X(z)}dz \geq 0,$$

because $\gamma^5(x_{max}) \geq \gamma^4(x_{max}) \geq \gamma^4(z)$ and $\alpha^2(z)/f_X(z) \geq 0$. Similarly evaluating (12) at $x = x_{min}$, we obtain

$$AV^5_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{fixed}}(x_{min})\right] - AV^5_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{VS}}(x_{min})\right]$$
$$= C_0 \cdot n^{-4}\int_I \left[\gamma^5(x_{min}) - \gamma^4(z)\right]\frac{\alpha^2(z)}{f_X(z)}dz \leq 0,$$

because $\gamma^5(x_{min}) \leq \gamma^4(x_{min}) \leq \gamma^4(z)$ and $\alpha^2(z)/f_X(z) \geq 0$. □

Remark 1. We mentioned before that if $f_X(x)$ and $\sigma^2(x)$ are of the same functional

form, $AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{fixed}}(x)\right]$ is constant. If this is the case, $f_X(x)/\int_I f_X(x)dx$ coincides with $\sigma^2(x)/\int_I \sigma^2(x)dx$, and, the asymptotic heteroscedastic variance in (6) is identical to its homoscedastic counterpart in (9). In other words, our proposed VS bandwidth h_{VS} is the generalization of the h_{fixed}.

Let us call the leading term of the $MISE(m(x),\widehat{m}_{h_x}(x))$ in (4) as asymptotic MISE or the $AMISE(m(x),\widehat{m}_{h_x}(x))$ and call the leading term of the bias in (2) as asymptotic bias. Similarly, we call the asymptotic variance (bias squared) integrated with respect to $f_X(x)$ on I as integrated asymptotic variance (bias squared). Then, we have the following result on global performance of the proposed VS bandwidth in (8) and its fixed counterpart in (5).

Proposition 2. *The ratio of the integrated asymptotic variance to the integrated asymptotic bias squared under the proposed VS bandwidth $h_{VS}(x)$ remains the same 4 to 1 as the ratio under the MISE minimizing fixed bandwidth h_{fixed}.*

Proof. It is well-known that the relation,

$$AMISE(m(x),\widehat{m}_{h_{fixed}}(x)) = \frac{5}{4}\int_I AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{fixed}}(x)\right] f_X(x)dx,$$

is obtained because the AMISE is written as the function of a bandwidth h_x, $(1/nh_x)C_1 + (h_x^4/4)C_2$, where C_1 and C_2 are respectively the integrated asymptotic variance and the integrated asymptotic bias squared under the bandwidth h_x. As seen from (A.2), the same ratio of 4 to 1 is maintained between the integrated asymptotic variance and the integrated asymptotic bias squared for h_0, so

$$AMISE(m(x),\widehat{m}_{h_{VS}}(x)) = \frac{5}{4}AV_{\mathbf{X},\mathbf{Y}}\left[\widehat{m}_{h_{VS}}(x)\right]. \qquad \square$$

Proposition 2 says that the AMISE can be measured by the integrated asymptotic variance. We would expect from this and proposition 1 that the resulting AMISE under the $\widehat{m}_{h_{VS}}(x)$ cannot be always larger than the AMISE under the $\widehat{m}_{h_{fixed}}(x)$.

Proposition 3. *$AMISE(m(x),\widehat{m}_{h_{VS}}(x))$ is not larger than $AMISE(m(x),\widehat{m}_{h_{fixed}}(x))$ for all choices of $(f_X(x),\sigma^2(x),m(x))$.*

Proof. It suffices to show that there are cases

$$AMISE^5(m(x),\widehat{m}_{h_{fixed}}(x)) - AMISE^5(m(x),\widehat{m}_{h_{VS}}(x))$$

$$= C_0\left(\frac{5}{4}\right)^5 \cdot n^{-4}\int_I \left[1-\gamma^4(x)\right]\cdot\frac{\alpha^2(x)}{f_X(x)}dx > 0. \qquad (13)$$

Since the $\gamma(x)$ in (10) is the ratio of two "densities," $\gamma(x)$ cannot be greater than 1 for all x. Then, the term $\alpha^2(x)/f_X(x)$ can be found to hold (13) because the term is independent of $1-\gamma^4(x)$, given $f_X(x)$. $\qquad \square$

Remark 2. The inequality (13) holds when the term $\alpha^2(x)/f_X(x)$ is large in the area where $\gamma(x) < 1$, while the $\alpha^2(x)/f_X(x)$ is small in the area $\gamma(x) > 1$.

3.   Illustrative examples

Illustrative example 1. To demonstrate propositions 1 and 3, we give illustrative example 1. Set $f_X(x) = 2(2 - x)/3$ and $m(x) = (x - 2)^{-2}$ on the domain $I = [0, 1]$ as in the left panel of Figure 3. This will determine $\alpha^2(x)/f_X(x) = (8/3)(2 - x)^{-7}$ to be a monotone increasing function on I as seen in the right panel of Figure 3. At the same time, this choice of $f_X(x)$ will dictate the choice of $\sigma^2(x)$ to be a function of $(2 - x)$ in order for $\gamma(x) = 1$. Let us set $\sigma^2(x) = 0.1(2 - x)$ or $\sigma^2(x)/\int_I \sigma^2(x)dx = 2(2 - x)/3$. It is easy to see from (10) that $\gamma(x) = 1$ and from (13) that for these cases the $AMISE(m(x), \widehat{m}_{h_{VS}}(x))$ is equal to $AMISE(m(x), \widehat{m}_{h_{fixed}}(x))$. In view of Remark 2 and the fact that $\alpha^2(x)/f_X(x)$ is monotone increasing, we realize $AMISE(m(x), \widehat{m}_{h_{fixed}}(x))$ is larger than $AMISE(m(x), \widehat{m}_{h_{VS}}(x))$ when $\sigma^2(x)$ is rotated about $x = 0.5$ clockwise as in the left panel of Figure 4, say to $\sigma^2(x) = 0.1(3 - 2x)$ or $\sigma^2(x)/\int_I \sigma^2(x)dx = (3 - 2x)/2$, and the reverse occurs when $\sigma^2(x)$ is rotated counterclockwise as in the same panel, say to $\sigma^2(x) = 0.1(x + 1)$ or $\sigma^2(x)/\int_I \sigma^2(x)dx = 2(x+1)/3$. Additionally, we see how the relative magnitude of these two measures when the $\sigma^2(x)$ is constant as $\sigma^2(x) = 0.1$ or $\sigma^2(x)/\int_I \sigma^2(x)dx = 1$.

For each conditional variance function $\sigma^2(x)$ above, we show in Table 1 the asymptotic homoscedastic variance in (9), the maximal and minimal values of the asymptotic heteroscedastic variance in (6), and the ratio of the two AMISE's,

$$r(f_X(x), \sigma^2(x), m(x)) = \frac{AMISE(m(x), \widehat{m}_{h_{VS}}(x))}{AMISE(m(x), \widehat{m}_{h_{fixed}}(x))}. \tag{14}$$

It is easy to see that the asymptotic homoscedastic variance in (9) is sandwitched by its maximal and minimal counterparts in (6) as proposition 1 claims for a given $\sigma^2(x)$. We also observe the ratio r moves from 1.51 to 0.89 as proposition 3 implies. From the two right panels in Figures 3 and 4, we see that $\sigma^2(x) = 0.1(3 - 2x)$ creates the situation described in Remark 2. [3]

In addition, we present Figure 5 to understand how accurately the NW estimators are locally estimated by the two bandwidths for the four conditional variance functions. In each panel of the figure, we plot asymptotic mean squared errors (AMSE) of the NW estimators for the two bandwidths at every point x. Simply adding the leading term of the variance in (3) to the squared leading term of the bias in (2), we obtain AMSE. Our proposed VS bandwidth outperforms MISE minimizing fixed bandwidth in terms of AMSE at the point x smaller than approximately $x = 0.8$ for $\sigma^2(x) = 0.1(x + 1)$, and approximately $x = 0.9$ for $\sigma^2(x) = 0.1$. When $\sigma^2(x) = 0.1(2 - x)$ is employed, AMSE's of the two bandwidths coincide at every point. When $\sigma^2(x) = 0.1(3 - 2x)$ is employed, VS bandwidth outperforms MISE minimizing fixed bandwidth at the point x greater than approximately $x = 0.8$. In this situation, we notice that, at these points greater than approximately $x = 0.8$, AMSE's are greatly improved by VS bandwidth while we see small deterioration of AMSE's at the rest of the points.

---

[3] We are sometimes asked a question if the data characterized by $f_X(x) = 2(2-x)/3$, $m(x) = (x-2)^{-2}$ and $\sigma^2(x) = 0.1(3 - 2x)$ is often observed in realistic data analysis. We can find a typical example illustrating the situation when we explain test scores of students by their preparation time spent for the test. Those students who spent a lot of preparation time can get higher test score with small variance and the number of those students are small in general.
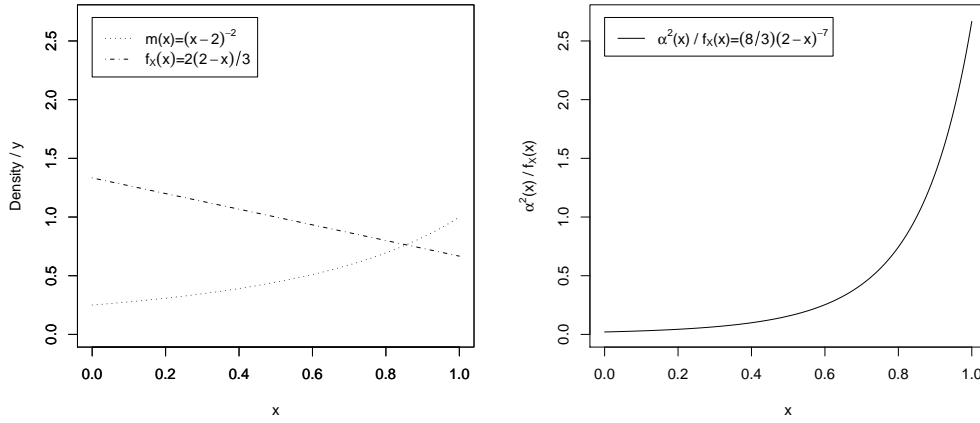
Figure 3: Functions in illustrative example 1.
The regression function $m(x) = (x - 2)^{-2}$ and its associated density of X are illustrated in the left panel. The function $\alpha^2(x)/f_X(x) = (8/3)(2 - x)^{-7}$ in the integrand in (13) is in the right panel.
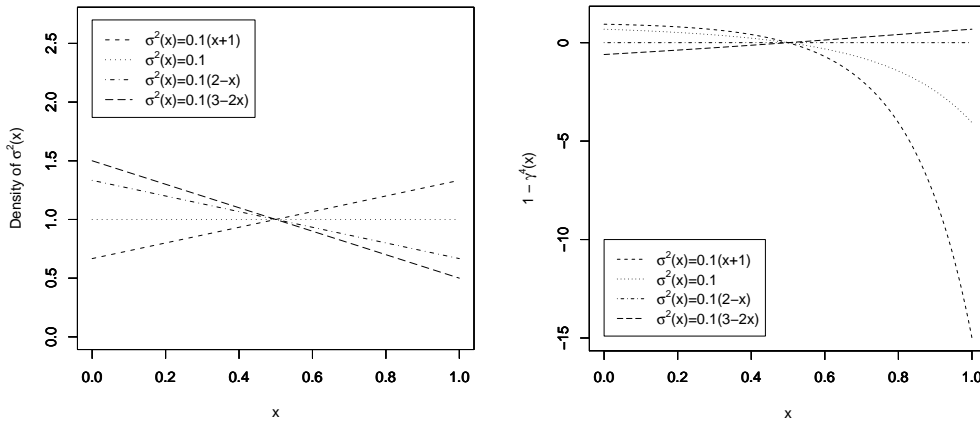

Figure 4: Functions in illustrative example 1.
The "densities" $\sigma^2(x)/\int_I \sigma^2(x)dx$ for the four $\sigma^2(x)$'s are illustrated in the left panel, while the functions $1 - \gamma^4(x)$ for the four $\sigma^2(x)$'s are in the right panel.
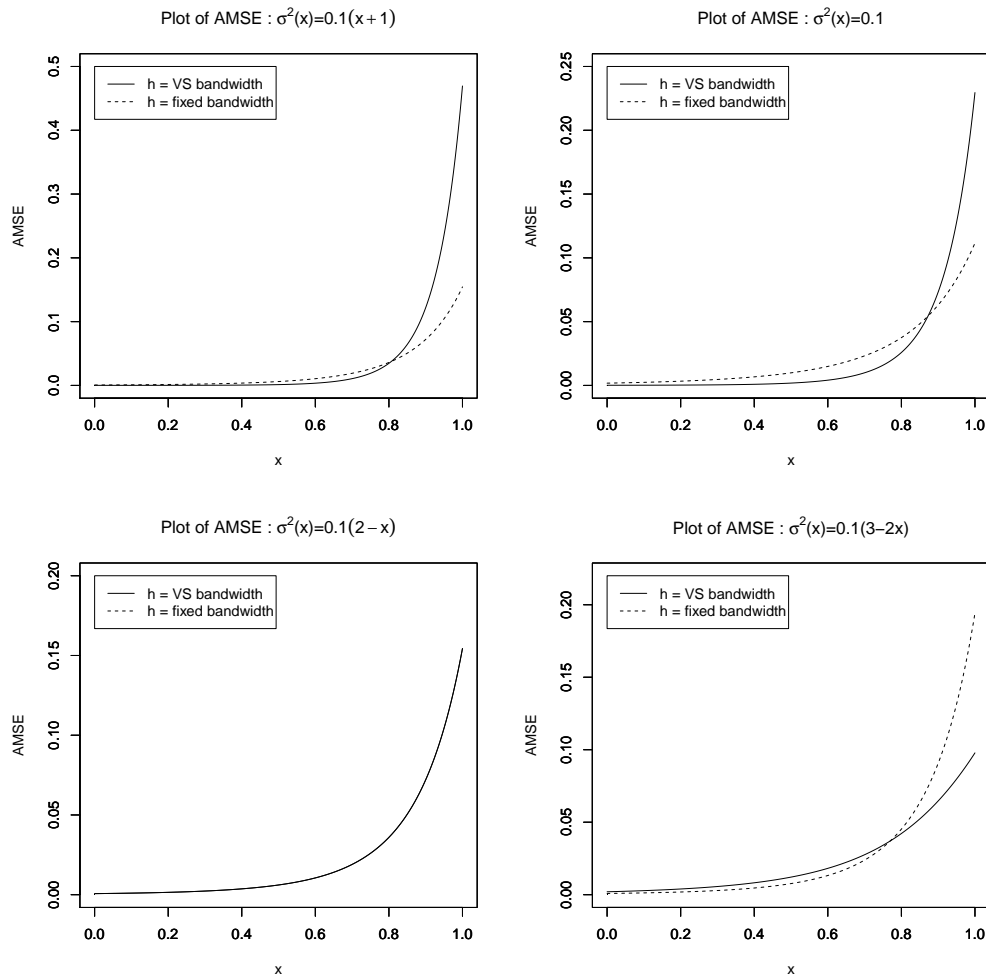
Figure 5: Plots of AMSE's for the four types of $\sigma^2(x)$.
In each panel, we plot AMSE's of the NW estimators for the two bandwidths at every point x. Please note that, when employed $\sigma^2(x) = 0.1(2 - x)$, plots of the two AMSE's coincide.

$\sigma^2(x)$	homo.var. in (9)	hetero.var. in (6)		r in (14)
		min. (arg.min)	max. (arg.max)	
$0.1(x+1)$	0.0981	0.0324 ($x=0$)	0.1297 ($x=1$)	1.5133
0.1	0.0587	0.0351 ($x=0$)	0.0703 ($x=1$)	1.2525
$0.1(2-x)$	0.0648	0.0648	0.0648	1.0000
$0.1(3-2x)$	0.0727	0.0612 ($x=1$)	0.0918 ($x=0$)	0.8904

Table 1: The result of numerical calculation for illustrative example 1.
The values of the homoscedastic variance in (9), the maximal and the minimal of the asymptotic heteroscedastic variance in (6) and the ratio r of the AMISE's in (14) are respectively calculated numerically for four $\sigma^2(x)$ for a given ($f_X(x) = 2(2-x)/3$, $\sigma^2(x)$, $m(x) = (2-x)^{-2}$). In calculating variances, n is set to be 1 and the Epanechnikov kernel is employed. The calculated numbers are rounded to the fourth decimal place.

Illustrative example 2. We give illustrative example 2 to investigate how fat-tailedness of the distribution of X's affects the ratio r of the AMISE's in (14). This is because we expect low density $f_X(x)$ at and near the boundaries of $I = [0,1]$ makes the proposed VS bandwidth very wide, rendering the bias not to be ignorable. For this, we retain $m(x) = (x-2)^{-2}$ on $I = [0,1]$ from illustrative example 1. We then set the conditional variance $\sigma^2(x)$ to be fixed at 0.1 as the regression function $m(x)$ increases. This constant $\sigma^2(x)$ over the support corresponds to the second row in Table 1. If the $\sigma^2(x)$ varies, it is more likely to increase as the $m(x)$ increases. Thus we additionally set in the form of increasing $\sigma^2(x) = 0.1(x+1)$, which corresponds to first row in Table 1. Now, on Table 2, we set up 17 cases all truncated normal with mean 0.5 at the center of support I and with the original (that is, before truncation) normal standard deviation s ranging from 0.25 to 1 to infinity. The second column of Table 2 shows that how many data are concentrated on I if the normal is not truncated on I. As s increases, the densities of X's are more flat. Especially when s approaches infinity, the distribution of X's is uniform. The third and fourth columns on Table 2 give the resulting r's in (14) for the aforementioned two cases of $\sigma^2(x)$. In Figure 6, we plotted the points (s, r) shown in Table 2 for the two $\sigma^2(x)$'s.

For $\sigma^2(x) = 0.1$, from Table 2 and Figure 6, the ratio r in (14) rapidly decreases in the begining but more slowly later from 2.2751 to 0.9956 as s increases from 0.25 to 0.6. As s increases further from 0.6 to 0.85, the ratio r increases but barely to 1.0220. As s grows from 0.85 to infinity, the ratio r decreases again slowly to 1.0000. Similarly, in the case of $\sigma^2(x) = 0.1(x+1)$, the ratio r rapidly decreases from 2.8074 to 0.9976 as s increases from 0.25 to 0.6. As s increases further from 0.6 to 0.95, the ratio r increases to 1.1934. As s grows from 0.95 to infinity, the ratio r decreases again slowly to 1.1883. We notice that the ratios r in both cases are quite high when X's are concentrated in the middle, but barely affected when X's has flatter distribution.

4. Discussion

In this paper, we first show, in proposition 1, that the proposed VS bandwidth produces the asymptotic variance smaller than that proposed by the standard fixed bandwidth on some part of the support I for the NW estimator. We then show, in

s	$\int_I \phi(x;s)dx$	r in (14)	
		$\sigma^2(x) = 0.1$	$\sigma^2(x) = 0.1(x+1)$
0.25	0.9545	2.2751	2.8077
0.3	0.9044	1.7000	2.0834
0.35	0.8468	1.4347	1.7408
0.4	0.7887	1.2762	1.5202
0.45	0.7334	1.1543	1.3178
0.5	0.6826	1.0519	1.0859
0.55	0.4721	0.9980	0.9976
0.6	0.4611	0.9956	0.9976
0.65	0.4500	1.0071	1.0817
0.7	0.4392	1.0157	1.1317
0.75	0.4950	1.0201	1.1596
0.8	0.4186	1.0218	1.1755
0.85	0.4090	1.0220	1.1847
0.9	0.3998	1.0214	1.1901
0.95	0.3911	1.0204	1.1934
1.0	0.3829	1.0193	1.1953
∞	—	1.0000	1.1883

Table 2: The result of numerical calculation for illustrative example 2.
The function $\phi(x;s)$ in the second column is the density function of $N(0.5, s^2)$. The ratio in (14) is calculated numerically for different values of the parameter s. The calculated numbers are rounded to the fourth decimal place.
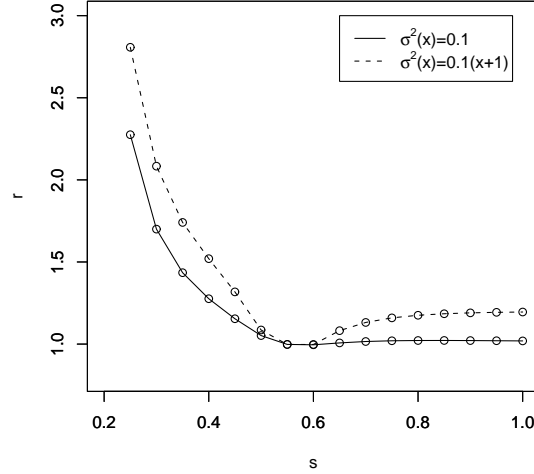

Figure 6: The plots of (s, r) in Table 2.
We plot the parameter s and the cerresponding r shown in Table 2 for the two conditional variance functions $\sigma^2(x)$.

proposition 3, that the proposed bandwidth may in some cases achieve even smaller AMISE while maintaining homoscedasticity. To demonstrate the propositions, we give illustrative example 1 to show that the penalty for homoscedasticity in terms of the AMISE ranges from -11% (that is, the proposed bandwidth is preferable) to 51% in Table 1.

In illustrative example 2, we examine how fat-tailedness of the distribution of X's affects the AMISE. It shows that the proposed VS bandwidth in (7) is more serviceable when the distribution X's are somewhat flatter because only several percentage point of penalty in the form of the increased AMISE is imposed in exchange for homoscedastic $\widehat{m}_{h_x}(x)$.

In Table 2, we are surprised to find that the ratios r go below unity when $s = 0.55$ and 0.60 in both $\sigma^2(x)$. Since the result is based on the limited number of numerical calculations, however, it is difficult to generalize the result to conclude that there will always be the cases where the proposed VS bandwidth in (7) outperforms the conventional fixed bandwidth in (5). Further extensive numerical investigation of this matter will give us more insight into how the proposed VS bandwidth will perform.

Appendix

Integrating square of the bias in (2) and variance in (3) over the support I, MISE between $\widehat{m}_{h_x}(x)$ and $m(x)$ is

$$\int_I \left[\frac{1}{nh_x} \frac{\sigma^2(x)}{f_X(x)} \left[\int K^2(t)dt \right] + o\left(\frac{1}{nh_x} \right) + O\left(\frac{1}{n} \right) \right.$$
$$\left. + \frac{h_x^4}{4} \frac{\alpha^2(x)}{f_X^2(x)} \left[\int t^2 K(t)dt \right]^2 + o\left(h_x^4 \right) + O\left(\frac{1}{n^2 h_x^2} \right) \right] f_X(x)dx. \qquad \text{(A.1)}$$

Substituting h_x in (A.1) for $h_{VS}(x)$ in (7),

$$\frac{1}{nh_0} \int_I \left[\int K^2(t)dt \right] f_X(x)dx + o\left(\frac{1}{nh_0} \right) + O\left(\frac{1}{n} \right)$$
$$+ \frac{h_0^4}{4} \int_I \frac{\sigma^8(x)\alpha^2(x)}{f_X^6(x)} \left[\int t^2 K(t)dt \right]^2 f_X(x)dx + o\left(h_0^4 \right) + O\left(\frac{1}{n^2 h_0^2} \right), \qquad \text{(A.2)}$$

differentiating the two leading terms in (A.2) with respect to h_0 and equating the outcome to zero,

$$-\frac{1}{nh_0^2} \left[\int K^2(t)dt \right] + h_0^3 \left[\int t^2 K(t)dt \right]^2 \int_I \frac{\sigma^8(x)\alpha^2(x)}{f_X^5(x)} dx = 0. \qquad \text{(A.3)}$$

The constant term in (8) is obtained by solving (A.3) with respect to h_0. $\qquad \square$

Acknowledgement

References

Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability and Its Applications* 9, 141-142.

Nadaraya, E.A. (1965). On nonparametric estimation of density functions and regression curves, *Theory of Probability and Its Applications* 10, 186-190.

Nadaraya, E.A. (1970). Remarks on nonparametric estimates for density functions and regression curves, *Theory of Probability and Its Applications* 15, 134-137.

Watson, G.S. and Leadbetter, M.R. (1963). On the estimation of probability density, I, *Annals of Mathematical Statistics* 34, 480-491.

Watson, G.S. (1964). Smooth regression analysis, *Sankhyā Series A* 26, 359-372.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Härdle, W. and Müller, M. and Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*, Springer, Berlin and Heidelberg.