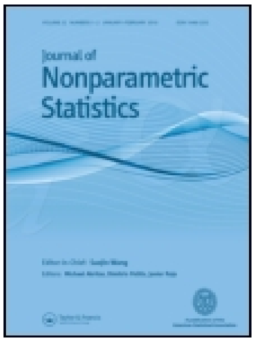

Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: https://www.tandfonline.com/loi/gnst20

On the asymptotic variance in nonparametric regression with fractional time-series errors

Yuanhua Feng

To cite this article: Yuanhua Feng (2007) On the asymptotic variance in nonparametric regression with fractional time-series errors, Nonparametric Statistics, 19:2, 63-76, DOI: 10.1080/10485250701381737

To link to this article: https://doi.org/10.1080/10485250701381737

Nonparametric Statistics
Vol. 19, No. 2, February 2007, 63–76

On the asymptotic variance in nonparametric regression with fractional time-series errors

YUANHUA FENG*

Department of Actuarial Mathematics and Statistics, Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

(Received 15 July 2005; in final form 27 March 2007)

This paper focuses on deriving explicit formulae of the asymptotic variance in nonparametric regression with fractional time-series errors. Unified formulae are obtained for fixed design models with long-memory, short-memory and antipersistent errors. It is also found that in strongly antipersistent case the Uniform kernel is no longer the minimum variance one and that for a fourth-order kernel the constant in the asymptotic variance for long-memory errors may be clearly smaller than that for i.i.d. errors. The results are applied to improving an existing data-driven algorithm. Practical performance of the proposed algorithm is illustrated with simulated and real data examples.

Keywords: Nonparametric regression; Long memory; Antipersistence; Bandwidth selection

AMS Subject Classifications: 62G08; 62G20; 62M10

1. Introduction

This paper considers fixed design nonparametric regression with fractional time-series errors including i.i.d. or short-memory, long-memory and antipersistent errors. The introduction of antipersistent errors is to solve the problem of possible over-differencing, which may occur in the difference series of some integrated time-series from finance and other fields. Asymptotic results in this context are obtained, e.g., by Hall and Hart [1], Deo [2] and Beran and Feng [3, 4]. Gao and Anh [5] proposed to approximate the nonparametric component in a partial linear regression with long-memory errors using finite-series sum and obtained similar asymptotic results as those in the context of kernel and local polynomial rgeression. Estimation and testing in a similar model with long-memory errors are recently studied by Aneiros-Pérez *et al.* [6], where it is assumed that the covariates of the linear part may spread out in some fashion from an unknown fixed design. Nonparametric regression with long-memory infinite variance errors is investigated by Peng and Yao [7].

Bandwidth cselection algorithms based on the iterative plug-in ref. [8] are proposed by Ray and Tsay [9] and Beran and Feng [3, 10] in the current context, where the asymptotic

*Tel.: +44 131 451 4365; Fax: +44 131 451 3249; Email: y.feng@ma.hw.ac.uk

Nonparametric Statistics
ISSN 1048-5252 print/ISSN 1029-0311 online
http://www.tandf.co.uk/journals
DOI: 10.1080/10485250701381737

variance are estimated using some implicit formulae obtained by Hall and Hart [1] and Beran and Feng [3], which may share great numerical problem. To our knowledge, explicit formula is only known in a very special case. This paper focuses on driving the explicit formulae of the asymptotic variance for general local polynomial estimators of $g^{(\nu)}$, the νth derivatives of the regression function g. Although known results of asymptotic variance in cases with long-memory, short-memory and antipersistent errors are given in different formulae separately, it is shown that all these formulae have a unified form, which changes smoothly from case with long-memory errors to case with antipersistent errors. Together with other findings in the literature, we can see that nonparametric regression with antipersistent errors is a necessary complement for models with long-memory and short-memory errors. Furthermore, nonparametric regression with antipersistent errors is also very useful in practice to avoid possible over-differencing. For instance, Karupiah and Los [11] found that many return series of high-frequency foreign exchange rates exhibit antipersistence instead of long memory.

The main results are helpful for deep understanding of nonparametric regression with fractional errors and useful for developing bandwidth selectors in this context. In this paper, they are applied to improve an existing data-driven algorithm for estimating g and g' based on local polynomial regression. Data examples show that this algorithm works well in practice. Since kernel and local polynomial regression is asymptotically equivalent, results in this paper can also be used to generalize the algorithm proposed by Ray and Tsay [9]. It is also worthwhile to apply the results in this paper to develop data-driven algorithm for the finite-series sum approximation proposed by Gao and Anh [5].

The paper is organized as follows. Section 2 describes the related results in the literature. The main results are described in section 3. In section 4, these results are used to improve an existing data-driven algorithm and then applied to simulated and data examples. Final remarks in section 5 conclude the paper. Proofs of results are shown in the Appendix.

2. Related results

Our discussion based on known formulae of asymptotic variance in nonparametric regression with fractional time-series errors, which will be summarized briefly. Formulae of asymptotic bias in the current context are the same as in models with i.i.d. errors and are hence omitted.

Consider the equidistant nonparametric regression

$$Y_i = g(t_i) + X_i, \tag{1}$$

where $t_i = (i/n)$ is the re-scaled time, $g : [0, 1] \to \Re$ is a smooth trend and X_i is a stationary fractionally integrated (hereafter fractional) error process defined by

$$(1 - B)^\delta X_i = Z_i, \tag{2}$$

where $\delta \in (-0.5, 0.5)$, B is the backshift operator and Z_i is a stationary time-series with absolutely summable autocovariances $\gamma_Z(k)$ so that $c_f = (1/2\pi)\sum_{k=-\infty}^{\infty}\gamma_Z(k) > 0$. Model (1) defines a nonparametric regression with long-memory ($\delta > 0$), short-memory ($\delta = 0$) and antipersistent ($\delta < 0$) errors. Here, the fractional difference $(1 - B)^\delta$ is introduced by Granger and Joyeux [12] and Hosking [13] (see also [14]). Model (1) is used for modelling a deterministic smooth trend in equidistant time-series with long memory, which is also the most well-studied model in the current context. The results obtained under model (1) can be extended to regular fixed design case but they do not apply to random design case, because the effect of long-memory errors under random design is quite different to that considered here. See, e.g., the difference between the asymptotic results on the parameter estimation given in Theorem 2.1, of Gao and Anh [5] and Theorem 1 of Aneiros-Pérez *et al.* [6] obtained under different designs.

Our goal is to estimate $g^{(\nu)}$ ($\nu \geq 0$) under model (1). In this paper pth order local polynomial estimators of $g^{(\nu)}$ [15] with $p-\nu$ odd, or kth order asymptotically equivalent kernel estimators with $k = p+1$, will be considered. Estimation of g' is useful to discover structural details of the data. Estimation of $g^{(k)}$ is necessary for developing plug-in algorithms. Assume that g is at least $(p+1)$ times continuously differentiable. Let $K(u)$ be a second-order kernel with compact support $[-1, 1]$ and h denote the bandwidth. The pth order local polynomial estimators of $g^{(\nu)}$ are defined as follows. Let $\mathbf{a} = (a_0, a_1, \ldots, a_p)'$ and

$$\hat{\mathbf{a}}'(x) = \arg\min_{\mathbf{a}} \sum_{i=1}^{n} \left[Y_i - a_0 - a_1(t_i - x) - \cdots - a_p(t_i - x)^p\right]^2 K\left(\frac{t_i - x}{h}\right). \tag{3}$$

Then $\hat{m}^{(\nu)}(x) := j!\hat{a}_j(x)$ estimates $m^{(\nu)}(x)$. Commonly used second-order kernels are of the form $K_\mu(x) = C_\mu(1-x^2)^\mu \mathbb{I}_{[-1,1]}$ with $\mu = 0, 1, \ldots$, where C_μ is determined by $\int_{-1}^{1} K_\mu(x)\mathrm{d}x = 1$ and μ is called the smoothness order [16]. They are the uniform, the Epanechnikov and the Bisquare kernels for $\mu = 0$, 1 and 2, respectively.

Local polynomial estimator $\hat{g}^{(\nu)}$ with $p - \nu$ odd have automatic boundary correction and the boundary effect in the mean integrated squared error is negligible. Hence, we will only discuss the asymptotic variance of $\hat{g}^{(\nu)}$ at an interior point $0 < t < 1$. A local polynomial estimator with the lowest order such that $p-\nu$ is odd, i.e. with $p = \nu + 1$, can be called the standard local polynomial estimator of $g^{(\nu)}$, which are, e.g., the local linear estimate for g, the local quadratic estimate for g' and the local cubic estimate for g''. Usually, we will propose the use of those estimators to keep the polynomial order as low as possible. Higher-order local polynomial estimators can be applied in large samples when the form of the trend function is complex. In addition, higher-order local polynomial estimators are also useful in some special contexts, e.g. as pilot estimators in double smoothing approaches.

Let $K_{(\nu,k)}(x)$ denote the equivalent kernel of order k for estimating $g^{(\nu)}$ [3, 17] obtained using K_μ as weighting function. Then $K_{(\nu,k)}$ has the same smoothness order μ and the polynomial form

$$K_{(\nu,k)}(x) = \sum_{l=0}^{q} \alpha_l x^l \mathbb{I}_{[-1,1]}(x). \tag{4}$$

Observe $x = y + (x - y)$. At a point $y \in [-1, 1]$, $K_{(\nu,k)}(x)$ can be decomposed as follows.

$$K_{(\nu,k)}(x) = \sum_{l=0}^{q} \beta_l(y)(x-y)^l = K^a(y) + K^b(x-y), \tag{5}$$

where $K^a(y)$ is independent of x and $K^b(x-y) = \sum_{l=1}^{q} \beta_l(y)(x-y)^l$ contains only powers in $(x-y)$, which are at least of first order, provided K^b does not vanish. It can be shown in particular that $K^a(y) = K_{(\nu,k)}(y)$. Let $g^{(\nu)}$ be estimated with $K_{(\nu,k)}(x)$ with bandwidth h satisfying the regular conditions given in Beran and Feng [3], then we have

$$\mathrm{var}[\hat{g}^{(\nu)}(t)] \doteq 2\pi c_f V(\delta)(nh)^{-1+2\delta} h^{-2\nu}, \tag{6}$$

where

$$V(0) = \int_{-1}^{1} K^2_{(\nu,k)}(x)\mathrm{d}x =: R(K_{(\nu,k)}), \tag{7}$$

$$V(\delta) = \frac{1}{\pi}\Gamma(1-2\delta)\sin(\pi\delta)\int_{-1}^{1} K_{(\nu,k)}(y)\int_{-1}^{1} K_{(\nu,k)}(x)|x-y|^{2\delta-1}\mathrm{d}x\,\mathrm{d}y \tag{8}$$

for $\delta > 0$ (see also [1] for results on $\hat{g}$ in this case) and

$$V(\delta) = \frac{1}{\pi}\Gamma(1-2\delta)\sin(\pi\delta)\int_{-1}^{1} K_{(\nu,k)}(y) \times \left\{\int_{-1}^{1} K^b(x-y)|x-y|^{2\delta-1}\mathrm{d}x - K_{(\nu,k)}(y)\int_{|x|>1}|x-y|^{2\delta-1}\mathrm{d}x\right\}\mathrm{d}y \tag{9}$$

for $\delta < 0$. In the following, V, as a function of δ and $K_{\nu,k}$, will be called the kernel-dependent function. Calculation of V using the above formulae may have great numerical problem, especially when $\nu > 0$ or δ is small. Furthermore, one may also be interested in understanding the relationship between the results given in equations (7)–(9) for different kinds of dependent errors.

3. The main results

In this section properties of $V(\delta)$ will be investigated in detail. A closed form formula of $V(\delta)$ is given as a quadruple sum, which allows us to calculate $V(\delta)$ quickly without numerical error. More explicit formulae are given in some special cases. Following Corollary 1 in Beran and Feng [4], kernel-dependent function $V(\delta)$ for the uniform kernel $K_0(x)$ is a unified function for $\delta < 0$ and $\delta > 0$ with $\lim_{\delta\to 0} V(\delta) = R(K_0) = 1/2$. It is expected that these results should hold in general cases. The following theorem shows that it is at least true for polynomial kernels in the form of equation (4).

THEOREM 1 *Let $K_{(\nu,k)}(x)$ be a polynomial kernel on* $[-1, 1]$ *as given in equation* (4). *Then*

(i) *the solutions of equation* (8) *and* (9) *are a unified function $V(\delta)$ for $\delta \in (-0.5, 0.5)\backslash\{0\}$; and*
(ii) $\lim_{\delta\to 0} V(\delta) = V(0) = R(K_{(\nu,k)})$, *where $R(K_{(\nu,k)})$ is as defined by equation* (7).

The proof of Theorem 1 is given in the Appendix. Following this theorem, the kernel-dependent function $V(\delta)$ is continuous in $(-0.5, 0.5)$ by defining $V(0) = R(K_{(\nu,k)})$, the kernel constant for $\delta = 0$. Theorem 1 together with the findings in Beran and Feng [3,4] shows that asymptotic properties of a nonparametric regression estimator change smoothly from case with long-memory errors to case with antipersistent errors. This fact provides theoretical evidence for nonparametric regression with antipersistent errors. The proof of (i) in Theorem 1 is based on the following lemma.

LEMMA 1 *Both the integrals $\int_{-1}^{1}|x-y|^{2\delta-1}\mathrm{d}x$ for $\delta > 0$ and $-\int_{|x|>1}|x-y|^{2\delta-1}\mathrm{d}x$ for $\delta < 0$ lead to the unified solution*

$$\frac{1}{2\delta}[(1+y)^{2\delta} + (1-y)^{2\delta}]. \tag{10}$$

The proof of Lemma 1 is given in the Appendix. Note that for the uniform kernel we have $K^a(y) \equiv 1/2$ and $K^b(x-y) \equiv 0$ for $-1 < x, y < 1$. Corollary 1 in Beran and Feng [4] can be easily proved again following equations (7)–(9) and Lemma 1.

COROLLARY 1 (*Corollary* 1 *in Beran and Feng* [4]) *The kernel-dependent function* $V(\delta)$ *for the uniform kernel is*

$$V(\delta) = \frac{2^{2\delta-1}}{\pi} \frac{\Gamma(1-2\delta)\sin(\pi\delta)}{\delta(2\delta+1)} \tag{11}$$

with $V(0) := \lim_{\delta\to 0} V(\delta) = 1/2$.

The proof of Corollary 1 is omitted.

The following theorem gives a general closed form formula of $V(\delta)$, represented as a double sum of some terms related to the integral of a double binomial form, whose coefficients are determined by the kernel. Define $Vc = (1/\pi)\Gamma(1-2\delta)\sin(\pi\delta)$ and we have the following theorem.

THEOREM 2 *Let* $K_{(\nu,k)}(x)$ *be a polynomial kernel on* $[-1, 1]$ *as given in equation* (4). *Let* $m' = 2$ *for* k *even and* $m' = 3$ *for* k *odd. Then we have, for* $\delta \in (-0.5, 0.5)\backslash\{0\}$,

$$V(\delta) = Vc\left[\sum_{(l-k)\,\text{even}} \alpha_l^2 T_{l,l} + 2\sum_{\substack{l\geq m'\\(l-k)\,\text{even}}}\ \sum_{\substack{m<l\\(m-k)\,\text{even}}} \alpha_l\alpha_m T_{l,m}\right], \tag{12}$$

where for $l, m = 0, 1, \ldots, q$, *such that* $(l-k)$ *and* $(m-k)$ *are both even*

$$\begin{aligned} T_{l,m} &= \int_{-1}^{1} y^l \int_{-1}^{1} x^m |x-y|^{2\delta-1}\mathrm{d}x\,\mathrm{d}y \\ &= 2\sum_{i=0}^{m}\binom{m}{i}\frac{1}{2\delta+i}\sum_{j=0}^{l+m-i}(-1)^j\binom{l+m-i}{j}\frac{2^{2\delta+j+i+1}}{2\delta+j+i+1}. \end{aligned} \tag{13}$$

Proof of Theorem 2 is given in the Appendix. Note that $V(\delta)$ depends on the kernel function only through the obvious relationship in equation (12). This representation based on the facts that $\alpha_l = 0$ in equation (4) for coefficients α_l such that $(l-k)$ is odd and that $T_{l,m} = T_{m,l}$. Hence, $V(\delta)$ only contains $T_{l,m}$ with l and m both even (for even k) or both odd (for odd k). Consequently, $l+m$ is always even. The term $T_{l,m}$ is a function of l, m and δ, independent of the kernel. Although $T_{l,m} = T_{m,l}$, equation (13) suggests that, for $l > m$, $T_{l,m}$ is easier to calculate than $T_{m,l}$. The following lemma simplifies the calculation of $T_{l,m}$.

LEMMA 2 *Assume that* $l, m = 0, 1, \ldots, q$, *are both even or both odd such that* $l+m$ *is even, then we have*

$$\int_{-1}^{1} y^l \int_{-1}^{y} x^m (y-x)^{2\delta-1}\mathrm{d}x\,\mathrm{d}y = \int_{-1}^{1} y^l \int_{y}^{1} x^m (x-y)^{2\delta-1}\mathrm{d}x\,\mathrm{d}y, \tag{14}$$

and hence

$$T_{l,m} = 2\int_{-1}^{1} y^l \int_{y}^{1} x^m (x-y)^{2\delta-1}\mathrm{d}x\,\mathrm{d}y. \tag{15}$$

The proof of Lemma 2 is given in the Appendix.

From equation (13) we can obtain $T_{0,0} = 1/[\delta(2\delta+1)]2^{2\delta+1}$, which can also be used to show Corollary 1. To illustrate Theorem 2 clearly, the solution of $V(\delta)$ for another simple kernel $K_{(1,3)}(x) = -(3/2)x\,\mathbb{I}_{[-1,1]}(x)$, i.e. the kernel of order $(1, 3)$ for estimating g' with $\mu = 0$ (see table 5.7 in [16]) is given.

COROLLARY 2 *The kernel-dependent function for the kernel* $K_{(1,3)}(x) = -(3/2)x\,\mathbb{I}_{[-1,1]}(x)$ *is*

$$V(\delta) = \frac{9}{\pi}\Gamma(1-2\delta)\sin(\pi\delta)\frac{(1-2\delta)2^{2\delta-1}}{\delta(2\delta+1)(2\delta+3)} \tag{16}$$

with $V(0) := \lim_{\delta\to 0} V(\delta) = 3/2$.

The proof of Corollary 2 is given in the Appendix. From the proof of Corollary 2 we can see, given Theorem 2, it is still not easy to obtain a more explicit formula of $V(\delta)$ for a kernel function with higher-order powers. However, it is not difficult to computer $T_{l,m}$ and $V(\delta)$ following equations (12) and (13). For this purpose, a function in R, called kdf.t(l, m, d) is developed, where 'd' stands for δ. Following equation (12), $V(\delta)$ can be easily calculated by means of kdf.t(l, m, d). For instance, the R command to calculate $V(\delta)$ corresponds to the Epanechnikov kernel $K_{(0,2)}(x) = (3/4)(1-x^2)\,\mathbb{I}_{[-1,1]}$ with $\mu = 1$ at d, denoted by V021d, is

$$\text{V021d} = (3/4)^{**}2^{*}\text{Vc}^{*}(\text{kdf.t}(0,0,d) - 2^{*}\text{kdf.t}(2,0,d) + \text{kdf.t}(2,2,d)).$$

Figure 1 displays $V(\delta)$ on $[-0.45, 0.45]$ for the uniform, the Epanechnikov and the Bisquare kernels, the three corresponding kernels of order $(0,4)$, the three corresponding kernels of order $(1,3)$ for estimating g' and the three corresponding kernels of order $(2,4)$ for estimating g''. See table 5.7 in [16] for the formulae of these kernels. They are at the same time the equivalent kernels for local linear and local cubic fitting of g, local quadratic fitting of g' as well as local cubic fitting of g'' using $K_\mu(x)$, $\mu = 0$, 1 and 2, as weighting functions, respectively. In particular, the functions given in equations (11) and (16) are those shown in figures 1(a) and (g). The dashed line in the figure shows the corresponding value $V(0) = R(K_{(\nu,k)})$.

Some findings in nonparametric regression with fractional time-series errors can be drawn by means of figure 1. We see, all of the functions $V(\delta)$ are convex and tends to infinite as $\delta \to -1/2$. For estimating the regression function g, $V(\delta)$ also tends to infinite as $\delta \to 1/2$. For kernels of order $(0,2)$ (i.e. symmetric densities), the minimum of $V(\delta)$ occurs at some negative δ near the origin. The difference between the minimum and $V(0)$ is not clear. For kernels of order $(0,4)$ the weights are sometimes negative. This causes the phenomenon as shown in figures 1(d)–(f), i.e. $V(\delta)$ for some $\delta > 0$ (with long memory) is clearly smaller than that for i.i.d. errors. Note, however that, the value of $V(\delta)$ does not change the order of the asymptotic variance. For estimating g' or g'' we have $\int K_{\nu,k}(x)\mathrm{dx} = 0$. As a consequence, $V(\delta)$ decreases monotonically now. Furthermore, in nonparametric regression with i.i.d. or short-memory errors it is well known that the uniform kernel is the minimum variance kernel. From figure 1, we can see that, this is no longer true for strongly antipersistent errors. Now the Epanechnikov or corresponding kernels will have smaller variance. Observe further that estimates obtained using the uniform or corresponding higher-order kernels are discontinuous but those obtained using the Epanechnikov or corresponding kernels are continuous. Hence, we propose the use of the Epanechnikov kernel as weighting function.

4. Applications

Results obtained in the last section can be used to develop data-driven algorithms in the current context. In this section they are employed to improve the numerical performance of the general bandwidth selector proposed in Beran and Feng [3]. It is assumed that X_i in equation (1) follow a fractional autoregressive model for simplicity, as used in the SEMIFAR model (semiparametric fractional autoregressive [4, 18]). Now, the dependence structure of

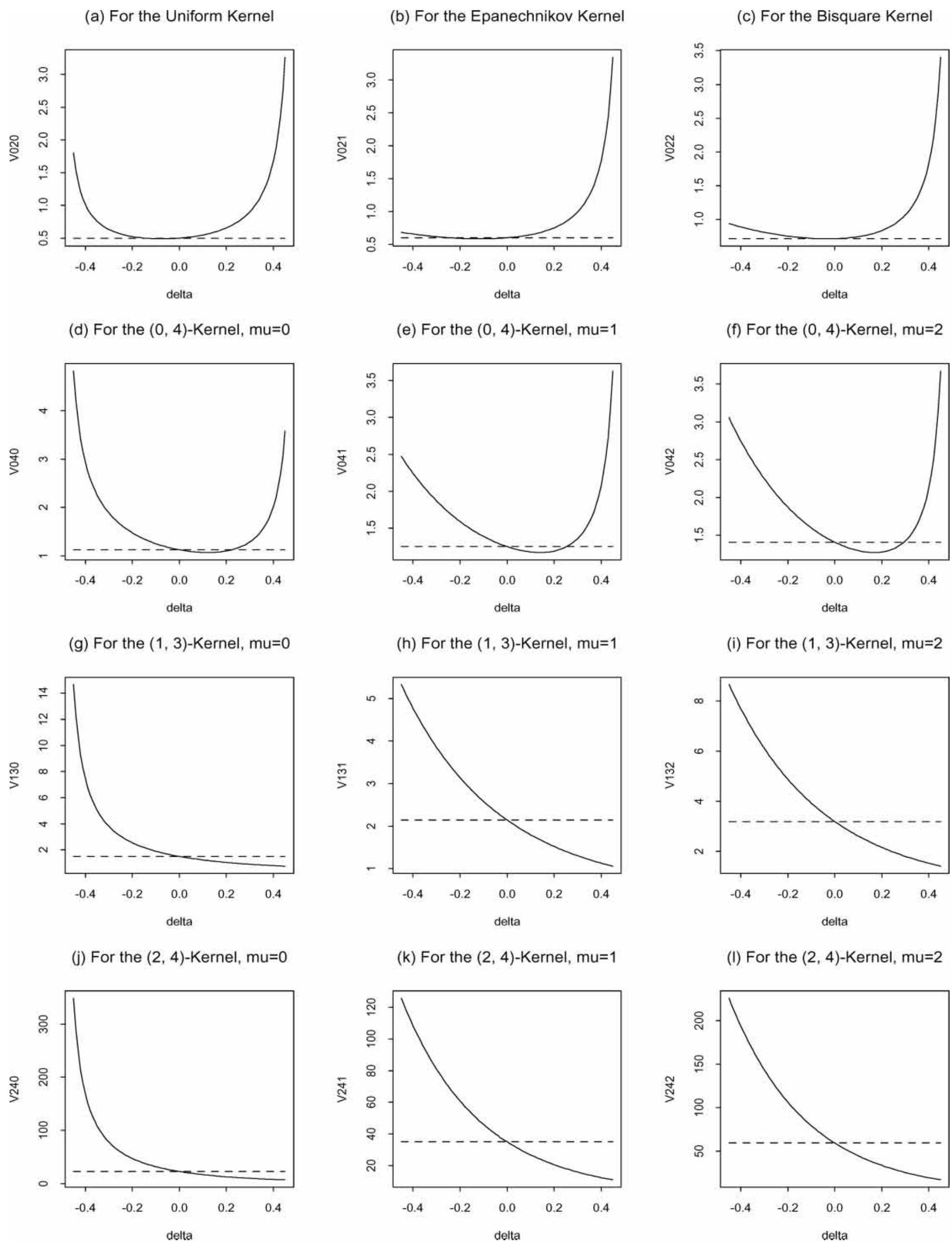

Figure 1. Kernel-dependent functions $V(\delta)$ on $[-0.45, 0.45]$ (solid curves), whereas the dashed lines show the corresponding values $V(0) = R(K_{(\nu,k)})$.

X_i is determined by an unknown parameter vector θ. The improved data-driven algorithm processes similarly to the original algorithm and has the same asymptotic properties. Hence, this will not be discussed here to same space.

Note that the estimation of the long-memory parameter and that of the trend depend strongly on each other. On the one hand, a deterministic trend in a time-series will cause erroneous long memory, if it is not estimated and removed properly. On the other hand, as indicated by a referee, the assumption of the existence of a trend in equation (1) may also cause erroneous long-range dependence when there is indeed no trend in the time-series. Both problems can be solved, if a good data—driven algorithm is used. To show this, data-driven estimation results of three typical simulated FARIMA(0, δ, 0) series with $\delta = -0.25$, 0 and 0.25, and the underlying trend

$$g_1(x) = 2\tanh(5(x-0.5)) - 2\sin(2.5(x-0.5)\pi)$$

are displayed in figure 2(a), (c) and (e), respectively, where standard normal innovations are used. Figure 2(b), (d) and (f), respectively, show the data-driven estimates for those time-series without trend (or with a trend $g_0 \equiv 0$, say). In all cases, local linear estimators with the Epanechnikov kernel as weighting function are used. The selected bandwidths and the estimated long-memory parameters $\hat{\delta}_r$ from the residuals are listed in table 1. The estimated long-memory parameters $\hat{\delta}_d$ from the data directly are also given as comparisons. Consider first the cases with a trend. Figure 2 and table 1 confirm the facts that the larger the δ, the more difficult the estimation of the mean function and the larger the δ, the larger the optimal bandwidth for estimating g. The estimated $\hat{\delta}_r$ from the residuals are satisfactory. But if the

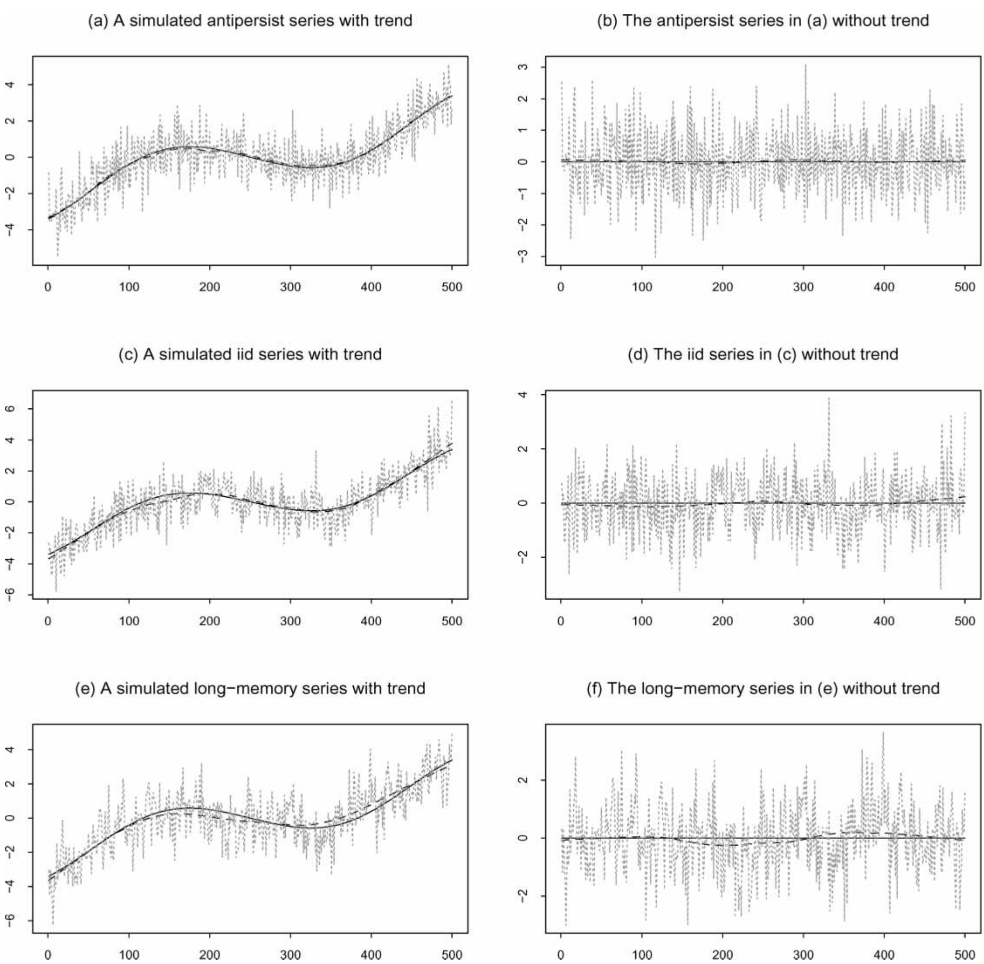

Figure 2. The data (short dashes) together with the mean functions (solid lines) and their estimates (long dashes) for six simulated time-series with or without trend.

Table 1. Estimation results for all examples.

	$\delta = -0.25$			$\delta = 0$			$\delta = 0.25$		
Trend	$\hat{h}$	$\hat{\delta}_r$	$\hat{\delta}_d$	$\hat{h}$	$\hat{\delta}_r$	$\hat{\delta}_d$	$\hat{h}$	$\hat{\delta}_r$	$\hat{\delta}_d$
g_1	0.0645	−0.3051	0.3428	0.0974	−0.0040	0.3845	0.1426	0.2106	0.4257
g_0	0.1534	−0.2936	−0.2465	0.2065	0.0167	0.0358	0.1906	0.2182	0.2290

trend is not removed, the estimated long-memory parameter is clearly wrong, in particular in the first two cases. If the series themselves are stationary but model (1) is used, the theoretically optimal bandwidth should be $h_0 = 0.5$, i.e. all observations should be used. We can see the selected bandwidths now are all clearly larger than in the corresponding cases with trend. The estimated long-memory parameters from the residuals perform well, which are however usually worse than those obtained under the knowledge that there is no trend. The case without trend was also included in the large simulation study in Beran and Feng [10] using algorithms without the improvements discussed in this paper. Their results indicated that, in the case without trend, the performance of the iterative plug-in algorithm is worse than in cases with trend. In summary, although model (1) works in cases without trend, a parametric model should be chosen automatically, if we can show that there is no significant trend in the data. For this purpose, a proper test approach should be developed.

In the following, two data examples will be given to show the practical performance of the proposed algorithm. The data sets used are: (1) the monthly Northern Hemisphere temperature (called Temp NH), from January 1880 to December 2002, anomalies (in C$^{\mathrm{o}}$) with respect to the monthly averages during 1961–1990, downloaded from the data release of the Climatic Research Unit at the University of East Anglia, and (2) the annual layer ice thickness at Arctic (GISP 2B) between 1270 and 1988, downloaded from the web page of the Arctic System Science, Colorado.

Figure 3 shows the data together with $\hat{g}$ using local linear (upper) and local cubic (middle) fitting for the Temp NH (left) and GISP 2B (right) series, where the observations of the GISP 2B series are shown from 1988 back to 1270 due to the nature of this data set. The estimated first derivatives using $p = 2$ are also shown in figures 2(c) and (f) for both examples, respectively, where the dependence structure was estimated using the pilot smoothing with $p = 3$. Again, the Epanechnikov kernel is used as weighting function. The selected bandwidth $\hat{h}$, the estimated long-memory parameter $\hat{\delta}$, the estimated AR model and the answer to the question, whether the trend is significant or not, are listed in table 2, where the AR model was selected from AR models of orders $0, 1, \ldots, 5$ following the Bayesian information criterion. The selected bandwidths $\hat{h}_d$ for estimating g' using both pilot estimations with $p = 1$ and $p = 3$ are also given in table 2. From figure 3 and table 2 we can see that the proposed algorithm works well in practice. The estimated dependence structure with $p = 1$ and $p = 3$ was about the same. This resulting in the fact that $\hat{g}'$ is almost independent of the pilot polynomial order.

The estimates of g for Temp NH series with $p = 1$ and $p = 3$ look quite similar, although the selected bandwidths in these two cases are clearly different. The results show that there are simultaneously significant trend, short memory and long memory in this time-series. Furthermore, there is clearly global warming in the last years. The GISP 2B is indeed a difference series of the total ice thickness, for which significant trend together with an antipersistent error process was found. However, the antipersistence cannot be correctly discovered, if the trend is not eliminated. For, if a FARIMA(0, δ, 0) model is fitted to the data, we will have $\hat{\delta} = 0.277$.

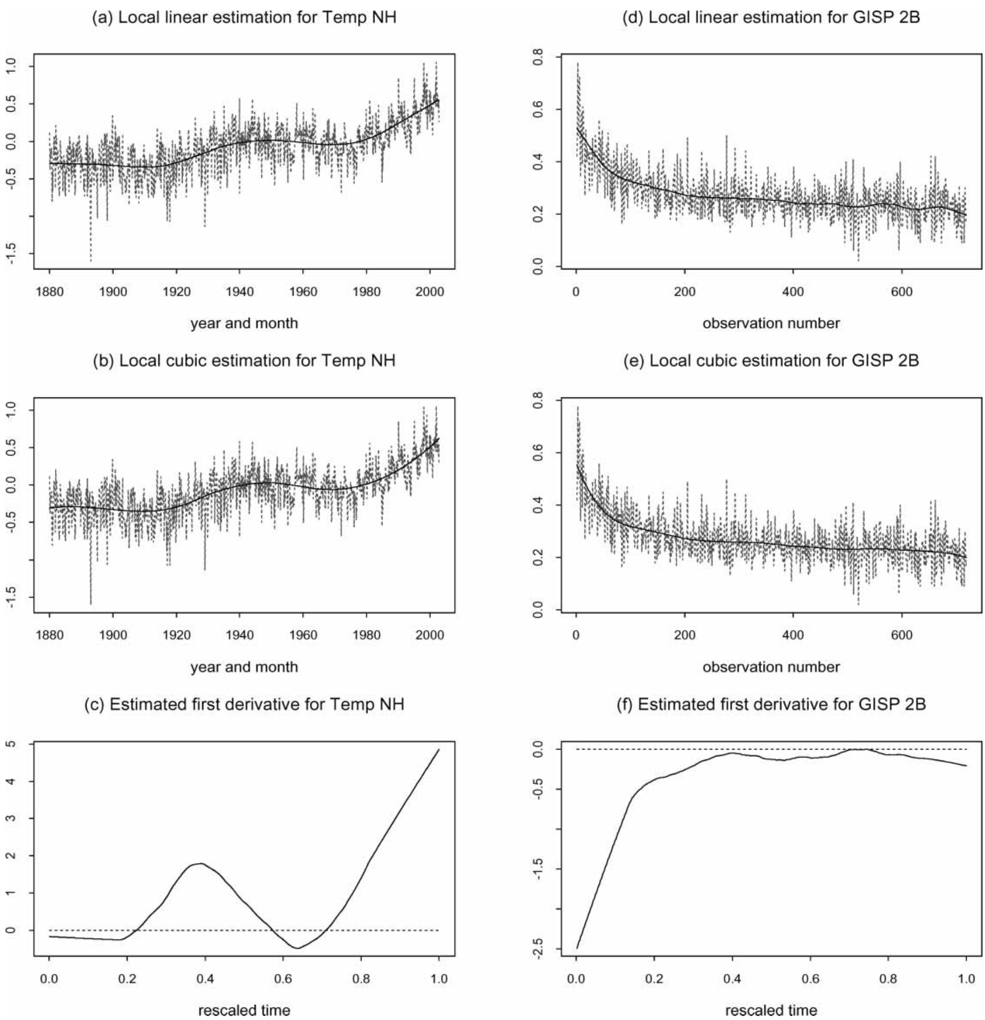

Figure 3. Data-driven local linear (upper) and local cubic (middle) estimates of g, and local quadratic estimate of g' (below) for the Temp NH (left) and GISP 2B (right) series.

Table 2. Estimation results for all examples.

Series	p	$\hat{h}$	$\hat{\delta}$	95% CI for δ	$\hat{r}$	$\hat{\phi}_1$	95% CI for ϕ_1	g-sig.	$\hat{h}_d$
Temp	1	0.126	0.208	[0.128, 0.289]	1	0.239	[0.139, 0.339]	Y	0.182
NH	3	0.237	0.205	[0.124, 0.286]	1	0.243	[0.143, 0.342]	Y	0.182
GISP	1	0.060	−0.128	[−0.185, −0.071]	0	–	–	Y	0.134
2B	3	0.153	−0.125	[−0.183, −0.068]	0	–	–	Y	0.134

5. Final remarks

In this paper, explicit formulae of asymptotic variance in nonparametric regression with fractional time-series errors are derived. Some interesting phenomena are found by means of these results. The main results are then used to improve an existing data-driven algorithm for estimating g and g'. Data examples show that the current algorithm works well in different cases. The main results of this paper can also be used for developing new data-driven algorithms in this context or be used in other exisitng proposals, e.g. that of Ray and Tsay [9]. Our results show in particular that nonparametric regression with long-memory, short-memory and antipersistent errors should be considered as a whole.

Acknowledgements

The paper was partly supported by the Center of Finance and Econometrics (CoFE), University of Konstanz, Germany. Our thanks go first to the Editor and a referee for useful comments and suggestions, which led to a significant improvement of the paper. We are very grateful to Professor Jan Beran, Department of Mathematics and Statistics, University of Konstanz, for helpful advices and comments. We would also like to thank the Climatic Research Unit at the University of East Anglia and the Arctic System Science, Colorado, for publishing the data.

References

[1] Hall, P. and Hart, J.D., 1990, Nonparametric regression with long-range dependence. *Stochastic Process Application*, **36**, 339–351.
[2] Deo, R.S., 1997, Nonparametric regression with long-memory errors. *Statistics Probability Letters*, **33**, 89–94.
[3] Beran, J. and Feng, Y., 2002, Local polynomial fitting with long-memory, short-memory and antipersistant errors. *Annals of the Institut Statistical Mathematics*, **54**, 291–311.
[4] Beran, J. and Feng, Y., 2002, SEMIFAR models – a semiparametric framework for modelling trends, long-range dependence and nonstationarity. *Computional Statistational and Data Analysis*, **40**, 393–419.
[5] Gao, J.T. and Anh, V., 1999, Semiparametric regression under long-range dependent errors. *Journal Statistics Planing Information*, **80**, 37–57.
[6] Aneiros-Pérez, G., González-Manteiga, W. and Vieu, P., 2004, Estimation and testing in a partial linear regression model under long-memory dependence. *Bernoulli*, **10**, 49–78.
[7] Peng, L. and Yao, Q., 2004, Nonparametric regression under dependent errors with infinite variance. *Annals of the Institute of Statistical Mathematics*, **56**, 73–86.
[8] Gasser, T., Kneip, A. and Köhler, W., 1991, A flexible and fast method for automatic smoothing. *Journal of the American Statistics Assocociation*, **86**, 643–652.
[9] Ray, B.K. and Tsay, R.S., 1997, Bandwidth selection for kernel regression with long-range dependence. *Biometrika*, **84**, 791–802.
[10] Beran, J. and Feng, Y., 2002, Iterative plug-in algorithms for SEMIFAR models – definition, convergence and asymptotic properties. *Journal of Computational and Graphical Statistics*, **11**, 690–713.
[11] Karuppiah, J. and Los, C.A., 2005, Wavelet multiresolution analysis of high-frequency Asian FX rates, Summer 1997. *International Review Financial Analysis*, **14**, 211–246.
[12] Granger, C.W.J. and Joyeux, R., 1980, An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15–30.
[13] Hosking, J.R.M., 1981, Fractional differencing. *Biometrika*, **68**, 165–176.
[14] Beran, J., 1994, *Statistics for Long-Memory Processes* (New York: Chapman & Hall).
[15] Fan, J. and Gijbels, I., 1996, *Local Polynomial Modeling and its Applications* (London: Chapman & Hall).
[16] Müller, H.G., 1988, *Nonparametric Analysis of Longitudinal Data* (Berlin: Springer–Verlag).
[17] Ruppert, D. and Wand, M.P., 1994, Multivariate locally weighted least squares regression. *Annals of Statistics*, **22**, 1346–1370.
[18] Beran, J., 1999, SEMIFAR models – A semiparametric framework for modelling trends, long range dependence and nonstationarity. Discussion paper No. 99/16, Center of Finance and Econometrics, University of Konstanz.

Appendix A

Proofs of the results are as follows.

Proof of Lemma 1 (i) For $\delta > 0$, set $z = y - x$ and $u = x - y$,

$$\int_{-1}^{1} |x-y|^{2\delta-1}\,\mathrm{d}x = \int_{-1}^{y} (y-x)^{2\delta-1}\,\mathrm{d}x + \int_{y}^{1} (x-y)^{2\delta-1}\,\mathrm{d}x$$

$$= -\int_{1+y}^{0} z^{2\delta-1}\,\mathrm{d}z + \int_{0}^{1-y} u^{2\delta-1}\,\mathrm{d}u$$

$$= \frac{1}{2\delta}\left[(1+y)^{2\delta} + (1-y)^{2\delta}\right]. \tag{A1}$$

(ii) For $\delta < 0$, set $z = y - x$ and $u = x - y$ again,

$$-\int_{|x|>1} |x-y|^{2\delta-1}\,\mathrm{d}x = -\int_{-\infty}^{-1} (y-x)^{2\delta-1}\,\mathrm{d}x + \int_{1}^{\infty} (x-y)^{2\delta-1}\,\mathrm{d}x$$

$$= \int_{\infty}^{1+y} z^{2\delta-1}\,\mathrm{d}z - \int_{1-y}^{\infty} u^{2\delta-1}\,\mathrm{d}u$$

$$= \frac{1}{2\delta}\left[(1+y)^{2\delta} + (1-y)^{2\delta}\right]. \tag{A2}$$

Lemma 1 is proved. ■

Proof of Theorem 1 (i) By means of the kernel decomposition given in equation (5), $V(\delta)$ in equation (8) for $\delta > 0$ can also be rewritten in a similar way as that given in equation (9) for $\delta < 0$,

$$V(\delta) = \frac{1}{\pi}\Gamma(1-2\delta)\sin(\pi\delta)\int_{-1}^{1} K_{(\nu,k)}(y)$$
$$\times\left\{\int_{-1}^{1} K^b(x-y)|x-y|^{2\delta-1}\,\mathrm{d}x + K_{(\nu,k)}(y)\int_{-1}^{1}|x-y|^{2\delta-1}\mathrm{d}x\right\}\,\mathrm{d}y. \tag{A3}$$

Comparing (A3) and equation (9), we see that $V(\delta)$ has a unified function form for $\delta > 0$ and $\delta < 0$, if and only if $\int_{-1}^{1}|x-y|^{2\delta-1}\,\mathrm{d}x$ for $\delta > 0$ and $-\int_{|x|>1}|x-y|^{2\delta-1}\mathrm{d}x$ for $\delta < 0$ have the same function form. This follows from Lemma 1.

(ii) Now we will show $\lim_{\delta\to 0} V(\delta) = \int_{-1}^{1} K^2(x)\mathrm{d}x$. Observing $\lim_{\delta\to 0}\Gamma(1-2\delta) = \lim_{\delta\to 0}\sin(\pi\delta)/\pi\delta = 1$, we have, following (A1) and (A3),

$$\lim_{\delta\to 0^+} V(\delta) = \lim_{\delta\to 0^+}\delta\int_{-1}^{1} K_{(\nu,k)}(y)$$
$$\times\left\{\int_{-1}^{1} K^b(y-x)|x-y|^{2\delta-1}\,\mathrm{d}x + K_{(\nu,k)}(y)\int_{-1}^{1}|x-y|^{2\delta-1}\,\mathrm{d}x\right\}\mathrm{d}y$$
$$= \frac{1}{2}\lim_{\delta\to 0^+}\int_{-1}^{1} K^2_{(\nu,k)}(y)[(1+y)^{2\delta} + (1-y)^{2\delta}]\mathrm{d}y. \tag{A4}$$

The last part in equation (A4) is due to that $\int_{-1}^{1} K_{(\nu,k)}(y)\int_{-1}^{1} K^b(y-x)|x-y|^{2\delta-1}\,\mathrm{d}x\mathrm{d}y$ is bounded for all δ. Furthermore, it is easy to show that

$$\lim_{\delta\to 0^+}\int_{-1}^{1} K^2_{(\nu,k)}(y)(1+y)^{2\delta}\,\mathrm{dy} = \int_{-1}^{1} K^2_{(\nu,k)}(y)\,\mathrm{d}y.$$

and

$$\lim_{\delta\to 0^+}\int_{-1}^{1} K^2_{(\nu,k)}(y)(1-y)^{2\delta}\mathrm{dy} = \int_{-1}^{1} K^2_{(\nu,k)}(y)\mathrm{d}y.$$

Following equation (A4) we have $\lim_{\delta\to 0^+} V(\delta) = \int_{-1}^{1} K^2_{(\nu,k)}(y)\mathrm{dy}$.

Following Lemma 1 and analogous analysis, we can obtain $\lim_{\delta\to 0^-} V(\delta) = \int_{-1}^{1} K^2_{(\nu,k)}(y)\mathrm{d}y$, too. This finishes the proof of Theorem 1. ■

Proof of Lemma 2 (i) Set $z = y - x$, we have

$$\begin{aligned}\int_{-1}^{1} y^l \int_{-1}^{y} x^m (x-y)^{2\delta-1}\,\mathrm{d}x\mathrm{d}y &= -\int_{-1}^{1} y^l \int_{y+1}^{0} (y-z)^m z^{2\delta-1}\,\mathrm{d}z\mathrm{d}y\\ &= \int_{-1}^{1} y^l \left\{\sum_{i=0}^{m}(-1)^i \binom{m}{i} y^{m-i}\int_{0}^{1+y} z^{2\delta+i-1}\,\mathrm{d}z\right\}\mathrm{d}y\\ &= \sum_{i=0}^{m}(-1)^i\binom{m}{i}\frac{1}{2\delta+i}\left\{\int_{-1}^{1} y^{l+m-i}(1+y)^{2\delta+i}\,\mathrm{d}y\right\}.\end{aligned} \tag{A5}$$

(ii) Set $z = x - y$, we have

$$\begin{aligned}\int_{-1}^{1} y^l \int_{y}^{1} x^m (x-y)^{2\delta-1}\,\mathrm{d}x\mathrm{d}y &= \int_{-1}^{1} y^l \int_{0}^{1-y} (y+z)^m z^{2\delta-1}\mathrm{d}z\mathrm{d}y\\ &= \int_{-1}^{1} y^l \left\{\sum_{i=0}^{m}\binom{m}{i} y^{m-i}\int_{0}^{1-y} z^{2\delta+i-1}\mathrm{d}z\right\}\mathrm{d}y\\ &= \sum_{i=0}^{m}\binom{m}{i}\frac{1}{2\delta+i}\left\{\int_{-1}^{1} y^{l+m-i}(1-y)^{2\delta+i}\mathrm{d}y\right\}.\end{aligned} \tag{A6}$$

Now set $z = -y$ and observe that $l + m$ is even, we have

$$\begin{aligned}\int_{-1}^{1} y^{l+m-i}(1-y)^{2\delta+i}\,\mathrm{d}y &= -\int_{1}^{-1} (-z)^{l+m-i}(1+z)^{2\delta+i}\,\mathrm{d}z\\ &= (-1)^i\int_{-1}^{1} z^{l+m-i}(1+z)^{2\delta+i}\,\mathrm{d}z.\end{aligned} \tag{A7}$$

The proof of Lemma 2 is finished by inserting equation (A7) into equation (A6). ■

Proof of Theorem 2 (i) The proof of equation (12) is straightforward and is omitted.

(ii) Following Theorem 1 we only have to calculate $T_{l,m}$ for $\delta > 0$. In the following we will continue the calculation of (A6). Set $z = 1 - y$, we have

$$\begin{aligned}\int_{-1}^{1} y^{l+m-i}(1-y)^{2\delta+i}\,\mathrm{d}y &= -\int_{2}^{0} (1-z)^{l+m-i} z^{2\delta+i}\,\mathrm{d}z\\ &= \left\{\sum_{j=0}^{l+m-i}(-1)^j\binom{l+m-i}{j}\int_{0}^{2} z^{2\delta+i+j}\,\mathrm{d}z\right\}\\ &= \sum_{j=0}^{l+m-i}(-1)^j\binom{l+m-i}{j}\frac{2^{2\delta+i+j+1}}{2\delta+i+j+1}.\end{aligned} \tag{A8}$$

Combining equations (A6) and (A8), we obtain

$$\int_{-1}^{1} y^l \int_{y}^{1} x^m (x-y)^{2\delta-1}\,\mathrm{d}x\mathrm{d}y = \sum_{i=0}^{m}\binom{m}{i}\frac{1}{2\delta+i}\sum_{j=0}^{l+m-i}(-1)^j\binom{l+m-i}{j}\frac{2^{2\delta+i+j+1}}{2\delta+i+j+1}.$$

Following Lemma 2, we have

$$
\begin{aligned}
T_{l,m} &= 2\int_{-1}^{1} y^l \int_{y}^{1} x^m (x-y)^{2\delta-1}\,\mathrm{d}x\,\mathrm{d}y \\
&= 2\sum_{i=0}^{m}\binom{m}{i}\frac{1}{2\delta+i}\sum_{j=0}^{l+m-i}(-1)^j\binom{l+m-i}{j}\frac{2^{2\delta+i+j+1}}{2\delta+i+j+1}.
\end{aligned}
\tag{A9}
$$

Theorem 2 is proved. ■

Proof of Corollary 2 Following equation (12), we have

$$
V(\delta) = \frac{9}{4\pi}\Gamma(1-2\delta)\sin(\pi\delta)T_{1,1}. \tag{A10}
$$

Furthermore, following equation (13)

$$
\begin{aligned}
T_{1,1} &= 2\sum_{i=0}^{1}\frac{1}{2\delta+i}\sum_{j=0}^{2-i}(-1)^j\binom{2-i}{j}\frac{2^{2\delta+j+i+1}}{2\delta+j+i+1} \\
&= \frac{1}{\delta}\left[\frac{2^{2\delta+1}}{2\delta+1} - 2\frac{2^{2\delta+2}}{2\delta+2} + \frac{2^{2\delta+3}}{2\delta+3}\right] \\
&\quad + \frac{2}{2\delta+1}\left[\frac{2^{2\delta+2}}{2\delta+2} - \frac{2^{2\delta+3}}{2\delta+3}\right].
\end{aligned}
\tag{A11}
$$

Straightforward calculation leads to

$$
T_{1,1} = \frac{(1-2\delta)2^{2\delta+1}}{\delta(2\delta+1)(2\delta+3)} \tag{A12}
$$

and equation (16) holds. It is clear that $\lim_{\delta\to 0} V(\delta) = 3/2 = R(K_{(1,3)})$. ■