Kernel Dependent Functions in Nonparametric Regression with Fractional Time Series Errors

Yuanhua Feng, University of Konstanz

Abstract

This paper considers estimation of the regression function and its derivatives in nonparametric regression with fractional time series errors. We focus on investigating the properties of a kernel dependent function $V(\delta)$ in the asymptotic variance and finding closed form formula of it, where δ is the long-memory parameter. It is shown that $V(\delta)$ has a unified form for $\delta \in (-0.5, 0.5)\backslash 0$ with $V(0) := \lim_{\delta \to 0} V(\delta) = R(K)$, the kernel constant for iid errors. General solution of $V(\delta)$ for polynomial kernels is given together with a few examples. It is also found, e.g. that the Uniform kernel is no longer the minimum variance one by strongly antipersistent errors and that, for a fourth order kernel, $V(\delta)$ at some $\delta > 0$ is clearly smaller than $R(K)$. The results are used to develop a general data-driven algorithm. Data examples illustrate the practical relevance of the approach and the performance of the algorithm.

Keywords: Nonparametric regression, long memory, antipersistence, fractional difference, kernel dependent function, bandwidth selection.

1 Introduction

Nonparametric regression with fractional time series errors was introduced by Beran (1999) under the SEMIFAR (semiparametric fractional autoregressive) model, which includes nonparametric regression with iid or short-memory errors (see e.g. Altman, 1990), long-memory errors (Hall and Hart, 1990) and antipersistent errors. The motivation to introduce antipersistent errors is to quantify the phenomenon of overdifferencing. Hence, the difference series of another time series series may have a nonparametric trend (g) together with antipersistent errors (see Section 4 for an example). Asymptotic results which are unified for the three types of dependent errors, are obtained by Beran (1999) for kernel estimator $\hat{g}$ of the trend and most recently by Beran and Feng (2002a) for local polynomial estimators $\hat{g}^{(\nu)}$ ($\nu \geq 0$) of the derivatives of the trend.

Practical implementation of $\hat{g}^{(\nu)}$ requires the selection of the bandwidth. Now, this problem and the estimation of the dependence structure depend on each other. Ray and

Tsay (1997) proposed a data-driven algorithm in the case with long-memory errors. Beran (1999) and Beran and Feng (2002b) proposed data-driven SEMIFAR algorithms, which work well for the three types of dependent errors. In these algorithms, a closed form formula of a kernel dependent function $V(\delta)$ ($\delta \in (-0.5, 0.5)$) in the asymptotic variance is used, which was only known for the Uniform kernel. General formulae of $V(\delta)$ are given in the form of improper integral (Hall and Hart, 1990 and Beran and Feng, 2002a), which are difficult to use due to numerical problem. To find the closed form formulae for a given kernel is generally not easy. We are also faced with a question: What is the relationship between $V(\delta)$ for antipersistent ($\delta < 0$) and long-memory ($\delta > 0$) errors? This motivated us to study the properties of $V(\delta)$ and to search general solution of $V(\delta)$.

Although the formulae of $V(\delta)$ given in Beran and Feng (2002a) look quite different for $\delta < 0$ and $\delta > 0$, it is shown that, for any polynomial kernel, the function form of $V(\delta)$ is unified for $\delta \in (-0.5, 0.5)\backslash 0$ with $\lim_{\delta \to 0} V(\delta) = R(K)$, where $R(K) = \int K^2(x)dx$ is the kernel constant in nonparametric regression with iid or short-memory errors (i.e. $\delta = 0$). This means $V(\delta)$ is a unified, continuous function in $(-0.5, 0.5)$ by defining $V(0) = R(K)$. Based on this result a general closed form formula of $V(\delta)$ is obtained as a quadruple sum of functions of δ. More explicit solutions are given for a few simple kernels. Furthermore, we also found some interesting phenomena in nonparametric regression with fractional time series errors, e.g. the well known fact in the case with iid or short-memory errors, that is the Uniform kernel is the minimum variance kernel (see e.g. Müller, 1988), is not true for strongly antipersistent errors, furthermore, for a fourth order kernel, $V(\delta)$ at some $\delta > 0$ is clearly smaller than $R(K)$ due to some negative weights.

The main results are useful for developing or improving data-driven algorithms in the current context. As an example, one of the SEMIFAR algorithms is generalized for data-driven estimation of g and g' based on local polynomial fitting. Data examples illustrate the usefulness of the approach and the practical performance of the algorithm. Our results do not depend on special model assumptions and can be used in more general cases, e.g. in the algorithm of Ray and Tsay (1997) proposed for kernel regression.

The paper is organized as follows. Section 2 summarizes some related results. The main results are described in Section 3. In Section 4 the results are used to improve a data-driven algorithm, which is then applied to data examples. Section 5 contains some conclusions. Proofs of results are put in the appendix.

2 Related results

Our discussion based on the formulae of asymptotic variance in nonparametric regression with fractional time series errors, which will be summarized briefly (for details see the references below). Some necessary concepts on kernels are also described.

Consider the equidistant nonparametric regression model

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

where $t_i = (i/n)$ is the re-scaled time, $g : [0, 1] \to \Re$ is a smooth trend function and X_i is a stationary fractionally integrated 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 $c_f = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_Z(k) > 0$. Model (1) defines a nonparametric regression with long-memory ($\delta > 0$), short-memory ($\delta = 0$) and antipersistent ($\delta < 0$) errors. Here, the fractional difference $(1 - B)^\delta$ introduced by Granger and Joyeux (1980) and Hosking (1981) (see also Beran, 1994) is defined by

$$(1 - B)^\delta = \sum_{k=0}^{\infty} \frac{\Gamma(k - \delta)}{\Gamma(k + 1)\Gamma(-\delta)} B^k. \tag{3}$$

Our goal is to estimate $g^{(\nu)}$ ($\nu \geq 0$) under model (1). In this paper k-th order kernel or p-th order local polynomial estimators of $g^{(\nu)}$ will be considered. Estimation of g' is helpful to discover more structural details of the data. Estimation of $g^{(k)}$ is necessary for developing a plug-in algorithm. Definition of these estimators may be found in Hall and Hart (1990), Beran (1999) and Beran and Feng (2002a). For nonparametric regression with iid or short-memory errors see Müller (1988), Härdle (1990), Altman (1990), Hart (1991) and Fan and Gijbels (1996) among others. For a kernel estimator, it is assumed that corresponding boundary kernels (see e.g. Müller, 1991 and Müller and Wang, 1994) are used at the boundary. For a local polynomial fitting of order p assume $p - \nu > 0$ is odd, so that $\hat{g}^{(\nu)}$ has automatic boundary correction and the same asymptotic properties as a kernel estimator with an equivalent kernel of order $k = p + 1$ (see e.g. Müller, 1987, Hastie and Loader, 1993 and Ruppert and Wand, 1994). Now the boundary effect in the MISE (mean integrated squared error) is negligible. We can hence focus on asymptotic results at interior points $0 < t < 1$.

Let $K_{(\nu,k)}(x)$ denote a kernel of order k (or an equivalent kernel of a p-th order local polynomial fitting) for estimating $g^{(\nu)}$ in the interior, which is assumed to have the polynomial form with compact support

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

Commonly used second order kernels for estimating g (i.e. kernel of order $(0,2)$) are in the form of $K_\mu(x) = C_\mu(1-x^2)^\mu \mathbb{1}_{[-1,1]}$ with $\mu = 0, 1$ and 2, which correspond to the Uniform, the Epanechnikov and the Bisquare kernels, where C_μ is determined by $\int_{-1}^{1} K_\mu(x)dx = 1$ and μ, called the smoothness order, is also an important criterion for a kernel (Müller, 1988). Higher order kernels can be derived from the kernels $K_\mu(x)$ or be obtained as equivalent kernels of a local polynomial fitting use one of them as the weighting function. A higher order kernel will take over the smoothness property of the weighting function.

Observe $x = y + (x - y)$. A useful decomposition of a kernel $K_{(\nu,k)}(x)$ at a point $y \in [-1, 1]$ introduced by Beran and Feng (2002a) is

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

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 h denote the bandwidth satisfying regular conditions, let $g^{(\nu)}$ be estimated with $K_{(\nu,k)}(x)$, then following Beran and Feng (2002a),

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

where

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

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

for $\delta > 0$ (see also Hall and Hart, 1990 for results on $\hat{g}$ in this case) and

$$\begin{aligned} V(\delta) \;=\; & \frac{1}{\pi}\Gamma(1-2\delta)\sin(\pi\delta)\int_{-1}^{1} K_{(\nu,k)}(y) \\ & \times \left\{\int_{-1}^{1} K^b(x-y)|x-y|^{2\delta-1}dx - K_{(\nu,k)}(y)\int_{|x|>1} |x-y|^{2\delta-1}dx\right\} dy \end{aligned} \tag{9}$$

for $\delta < 0$. For developing a data-driven algorithm one has to computer $V(\delta)$. However, numerical integral does not work well, especially for small δ. Furthermore, one would ask: What is the relationship between the results in (7) to (9)? Is there a unified, closed form formula for all the three cases? These questions will be answered in the next section.

3 The main results

In this section properties of $V(\delta)$ will be investigated in detail. A closed form formula of $V(\delta)$ is found and represented as a complex quadruple sum, which is however enough for computational purpose and allows us to calculate $V(\delta)$ quickly without numerical error. More explicit formulae are only given for a few simple kernels.

Corollary 1 in Beran (1999) (see also Beran and Feng, 2002c and Corollary 1 below) shows that $V(\delta)$ for the Uniform kernel $K_0(x)$ is a unified function for $\delta < 0$ and $\delta > 0$ with $\lim_{\delta\to 0} V(\delta) = R(K_0) = \frac{1}{2}$. It is expected that these results should hold in general cases. The following theorem shows that it is at least true for kernels in the form of (4).

Theorem 1 *Let $K_{(\nu,k)}(x)$ be a polynomial kernel on $[-1,1]$ as given in (4). Then*
i) The solutions of (8) and (9) are a unified function $V(\delta)$ for $\delta \in (-0.5, 0.5)\backslash 0$.
ii) $\lim_{\delta\to 0} V(\delta) = V(0) = R(K_{(\nu,k)})$, where $R(K_{(\nu,k)})$ is as defined by (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 results in Beran (1999) and Beran and Feng (2002a) shows that asymptotic properties of a nonparametric regression estimator change *smoothly* from case with antipersistent errors to case with long-memory errors. This fact provides theoretical evidence for nonparametric regression with antipersistent errors. The proof of *i*) in Theorem 1 is based on the following basic result.

Lemma 1 *Both of the integrals $\int_{-1}^{1} |x-y|^{2\delta-1}dx$ for $\delta > 0$ and $-\int_{|x|>1} |x-y|^{2\delta-1}dx$ for $\delta < 0$ leads to the unified formula*

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

The proof of Lemma 1 is given in the appendix. Corollary 1 in Beran (1999) can be easily proved again following (7) to (9) and Lemma 1.

Corollary 1 *(Corollary 1 in Beran, 1999) The kernel dependent function $V(\delta)$ for the Uniform kernel is*

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

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

The proof of Corollary 1 is straightforward and 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 $Vc = \frac{1}{\pi}\Gamma(1-2\delta)\sin(\pi\delta)$ we have

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

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

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

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

Proof of Theorem 2 is given in the appendix. Note that $V(\delta)$ depends on the kernel function only through the obvious relationship (12). This representation based on the facts that $\alpha_l = 0$ in (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, we have $l+m$ is always even. The term $T_{l,m}$ is a function of l, m and δ, independent of the kernel. Although $T_{l,m} = T_{m,l}$, (13) shows 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*

$$\int_{-1}^{1} y^l \int_{-1}^{y} x^m (y-x)^{2\delta-1} dx dy = \int_{-1}^{1} y^l \int_{y}^{1} x^m (x-y)^{2\delta-1} dx dy \tag{14}$$

and hence

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

The proof of Lemma 2 is given in the appendix.

From (13) we can obtain $T_{0,0} = \frac{1}{\delta(2\delta+1)}2^{2\delta+1}$, which can also be used to prove Corollary 1. To illustrate Theorem 2 clearly, we will give a more explicit solution of $V(\delta)$ for another simple kernel $K_{(1,3)}(x) = -\frac{3}{2}x\mathbb{1}_{[-1,1]}$, i.e. the kernel of order $(1,3)$ for estimating g' with $\mu = 0$ (see Table 5.7 in Müller, 1988).

Corollary 2 *The kernel dependent function for the kernel* $K_{(1,3)}(x) = -\frac{3}{2}x\mathbb{1}_{[-1,1]}$ *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 $V(0) := \lim_{\delta\to 0} V(\delta) = \frac{3}{2}$.

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 $V(\delta)$. For this purpose, an S-Plus function, called kdf.t(l, m, d) is developed, where the variable d stands for δ. Following (12), $V(\delta)$ can be easily calculated by means of the S-Plus function kdf.t(l, m, d). For instance, the S-Plus commend to calculate $V(\delta)$ corresponds to the Epanechnikov kernel $K_{(0,2)}(x) = \frac{3}{4}(1-x^2)\mathbb{1}_{[-1,1]}$ with $\mu = 1$ at d, denoted by V021d, is

V021d_(3/4)**2*Vc*(kdf.t(0,0,d)-2*kdf.t(2,0,d)+kdf.t(2,2,d))

Figure 1 shows the functions $V(\delta)$ on $[-0.45, 0.45]$ for commonly used kernels, i.e. for the Uniform, the Epanechnikov and the Bisquare kernels, the three corresponding kernels of order $(0,4)$ for estimating g, 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 Müller (1988) 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'' with $K_\mu(x)$, $\mu = 0$, 1 and 2, as weighting functions, respectively. In particular, the functions given in (11) and (16) are those shown in Figures 1a and 1g. The dashed line in a figure shows the corresponding value $R(K_{(\nu,k)})$.

Some empirical findings in nonparametric regression with fractional time series errors can be drawn from Figure 1. We see, all of the functions $V(\delta)$ are convex and tends to infinite as $\delta \to -\frac{1}{2}$. By kernels for estimating g, $V(\delta)$ also tends to infinite as $\delta \to \frac{1}{2}$. By 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. By kernels of order $(0,4)$ the weights are sometimes negative. This causes the phenomenon as shown in Figures 1d to 1f, i.e. the constant in the asymptotic variance for some $\delta > 0$ is clearly smaller than that for iid errors. Note however that, the value of $V(\delta)$ does not change the order of the asymptotic variance. Furthermore, note that $\int K(x)dx = 0$ for a kernel for estimating g' or g''. As a consequence, $V(\delta)$ decreases monotonously now. In nonparametric regression with iid or short-memory errors it is well known that the Uniform kernel is the minimum variance kernel (see e.g. Müller,1988). This is not true, if the errors are strongly antipersistent. For instance, if $\delta < -0.33$, $V(\delta)$ of the Epanichnikov kernel is smaller than that of the Uniform kernel. Furthermore, note 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 to use the Epanechnikov kernel as the weighting function. Figures 1c and 1f show that, the Bisquare kernel is also a good candidate for a weighting function, because now we will obtain more smoother estimates with just slight increase in the asymptotic variance.

4 Applications

Results obtained in the last section can be used to generalize a data-driven algorithm in nonparametric regression with fractional time series errors. As an example one of the data-driven SEMIFAR algorithms (AlgB in Beran and Feng, 2002b) is generalized. In this generalization local linear and local cubic estimation of g, and local quadratic estimation of g' are included. For carrying out $\hat{g}'$, $\hat{g}$ with $p = 1$ or $p = 3$ is used as a pilot smoothing for estimating the dependence structure. The three second order kernels $K_\mu(x) = C_\mu(1 - x^2)^\mu \mathbb{I}_{[-1,1]}$, $\mu = 0, 1$ or 2, are built-in as alternative weighting functions. Detailed description about the algorithm will be omitted to save space, because it is quite similar to the original one (see Beran and Feng, 2002b). The consistency of such a general algorithm was shown by Beran and Feng (2002a).

In the following two examples will be given to show the practical performance of the generalized SEMIFAR algorithm and to show the practical relevance of the approach discussed in Section 2. The data sets are 1. The monthly Northern Hemisphere temperature (called Temp NH), from January 1880 to December 2002, anomalies (in C$^\text{o}$) w.r.t. the monthly averages during 1961 to 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 (called GISP 2B following the name of the data set) between 1270 and 1988, downloaded from the web-page of the Arctic System Science, Colorado.

Figure 2 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. $\hat{g}'$ is also shown in Figures 2c and 2f for these examples respectively against the re-scaled time $t \in [0, 1]$, where the dependence structure was estimated using the pilot smoothing with $p = 3$. For all results the Epanechnikov kernel was used as the weighting function. The selected bandwidth $\hat{h}$, the estimated long-memory parameter $\hat{\delta}$, the estimated AR model and the answer to the question, if there is a significant trend in the data, are listed in Table 1, where the AR model was selected from AR(r) models of orders $r = 0, 1, ..., 5$. Also given in Table 1 are the selected bandwidths $\hat{h}_d$ for estimating g' in all cases. An additional parameter $m \in \{0, 1\}$ to determine if the series is integrated ($m = 1$) or non-integrated ($m = 0$) was also estimated by the SEMIFAR algorithm. Here we have $\hat{m} = 0$ in all cases, i.e. the two series are non-integrated. From Figure 2 and Table 1 we see that the generalized SEMIFAR algorithm works well in practice. The selected optimal bandwidths differ from case to case and are quite reasonable. 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 smoothing.

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. However, the dependence structure will be wrongly estimated without pre-eliminating of the trend. The significance of the fitted trend means that there is a clear global warming during this time. $\hat{g}'$ together with the horizon line $y = 0$ in Figure 2c divides the temperature change into four periods according to the sign of $\hat{g}'$, which correspond about to I. 1880 - 1906; II. 1907 - 1949; III. 1950 - 1966 and IV. 1967 - 2002. In the first and in particular in the third periods the averaged temperature decreased slightly. But the decrease in the averaged temperature was relatively very small. The averaged temperature increased clearly in the other two periods. By integrating $\hat{g}'$ we can find: 1. The total amount of temperature increase during this time was 0.95 C^{o}; 2. The amount of temperature increase since 1967 was 0.70 C^{o}. Furthermore, from Figure 2c we can see that the averaged temperature increases stronger and stronger since 1967, i.e. the increasing rate increases.

The GISP 2B is indeed a difference series of the total ice thickness, for which we found significant trend together with an antipersistent error process. However, the antipersistence cannot be correctly discovered, if the trend is not eliminated. To show this, we could fit a FARIMA(0, δ, 0) model to the original data. Now, we would obtain $\hat{\delta} = 0.277$! This wrong conclusion was due to the trend in this time series. It is clear, if only error processes with short or long memory were allowed, the dependence structure in this time series would also be wrongly estimated. Hence, nonparametric regression with antipersistent errors is not only theoretically important but also relevant in practice. For this example, $\hat{g}$ with $p = 3$ looks better than that with $p = 1$, because it is more stable at the right boundary. $\hat{g}'$ for this time series is almost always negative, i.e. the trend in this series decreases monotonously. The decrease is very clear at the beginning. We think such a decreasing trend is mainly due to the strength and duration of the press.

5 Conclusions

In this paper kernel dependent functions in nonparametric regression with fractional time series errors are obtained. Some interesting phenomena are found by means of these results. The main results are then used to generalize an existing data-driven algorithm for estimating g and g'. Data examples illustrate the practical usefulness of the proposal. Although the data-driven algorithm was developed based on the SEMIFAR model, the main results of this paper do not rely on the special model assumptions and are hence applicable in general cases. It is worthwhile to discuss the use of these results in other proposals, e.g. in the algorithm proposed by Ray and Tsay (1997) or in the case, when δ and c_f are estimated by other approaches.

Acknowledgments

This paper is financially supported by the Center of Finance and Econometrics, University of Konstanz, Germany. We are grateful to Prof. Jan Beran, Department of Mathematics and Statistics, University of Konstanz for providing us the original SEMIFAR program. We would like to thank the Climatic Research Unit at the University of East Anglia and the Arctic System Science, Colorado for publishing the data.

Appendix Proofs of the results

Proof of Lemma 1.

i) For $\delta > 0$, set $z = y - x$ and $u = x - y$,

$$\begin{aligned}\int_{-1}^{1} |x-y|^{2\delta-1}dx &= \int_{-1}^{y}(y-x)^{2\delta-1}dx + \int_{y}^{1}(x-y)^{2\delta-1}dx \\ &= -\int_{1+y}^{0} z^{2\delta-1}dz + \int_{0}^{1-y} u^{2\delta-1}du \\ &= \frac{1}{2\delta}\left[(1+y)^{2\delta} + (1-y)^{2\delta}\right]. \qquad \text{(A.1)}\end{aligned}$$

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

$$\begin{aligned}-\int_{|x|>1} |x-y|^{2\delta-1}dx &= -\int_{-\infty}^{-1}(y-x)^{2\delta-1}dx + \int_{1}^{\infty}(x-y)^{2\delta-1}dx \\ &= \int_{\infty}^{1+y} z^{2\delta-1}dz - \int_{1-y}^{\infty} u^{2\delta-1}du \\ &= \frac{1}{2\delta}\left[(1+y)^{2\delta} + (1-y)^{2\delta}\right]. \qquad \text{(A.2)}\end{aligned}$$

Lemma 1 is proved. ◇

Proof of Theorem 1.

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

$$\begin{aligned}V(\delta) &= \frac{1}{\pi}\Gamma(1-2\delta)\sin(\pi\delta)\int_{-1}^{1} K_{(\nu,k)}(y) \\ &\quad \times \left\{\int_{-1}^{1} K^b(x-y)|x-y|^{2\delta-1}dx + K_{(\nu,k)}(y)\int_{-1}^{1}|x-y|^{2\delta-1}dx\right\}dy. \qquad \text{(A.3)}\end{aligned}$$

Comparing (A.3) and (9), we see that $V(\delta)$ has a unified function form for $\delta > 0$ and $\delta < 0$, if and only if $\int_{-1}^{1}|x-y|^{2\delta-1}dx$ for $\delta > 0$ and $-\int_{|x|>1}|x-y|^{2\delta-1}dx$ for $\delta < 0$ have the same function form. This follows from Lemma 1.

ii) Now we will show $\lim\limits_{\delta\to 0} V(\delta) = \int_{-1}^{1} K^2(x)dx$. Observing $\lim\limits_{\delta\to 0}\Gamma(1-2\delta) = \lim\limits_{\delta\to 0}\frac{\sin(\pi\delta)}{\pi\delta} = 1$, we have, following (A.1) and (A.3),

$$\begin{aligned}\lim_{\delta\to 0^+} V(\delta) &= \lim_{\delta\to 0^+}\delta\int_{-1}^{1} K_{(\nu,k)}(y) \\ &\quad \times \left\{\int_{-1}^{1} K^b(y-x)|x-y|^{2\delta-1}dx + K_{(\nu,k)}(y)\int_{-1}^{1}|x-y|^{2\delta-1}dx\right\}dy \\ &= \frac{1}{2}\lim_{\delta\to 0^+}\int_{-1}^{1} K^2_{(\nu,k)}(y)[(1+y)^{2\delta} + (1-y)^{2\delta}]dy. \qquad \text{(A.4)}\end{aligned}$$

The last equation in (A.4) is due to that $\int_{-1}^{1} K_{(\nu,k)}(y) \int_{-1}^{1} K^b(y-x)|x-y|^{2\delta-1}dxdy$ is bounded for all δ. Furthermore, it is easy to show that

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

and

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

Following (A.4) we have $\lim_{\delta\to 0^+} V(\delta) = \int_{-1}^{1} K^2_{(\nu,k)}(y)dy$.

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

Proof of Lemma 2.

i) Set $z = y - x$, we have

$$\begin{aligned} \int_{-1}^{1} y^l \int_{-1}^{y} x^m (x-y)^{2\delta-1}dxdy &= -\int_{-1}^{1} y^l \int_{y+1}^{0} (y-z)^m z^{2\delta-1}dzdy \\ &= \int_{-1}^{1} y^l \left\{ \sum_{i=0}^{m} (-1)^i \binom{m}{i} y^{m-i} \int_0^{1+y} z^{2\delta+i-1}dz \right\} dy \\ &= \sum_{i=0}^{m} (-1)^i \binom{m}{i} \frac{1}{2\delta+i} \left\{ \int_{-1}^{1} y^{l+m-i}(1+y)^{2\delta+i}dy \right\} \end{aligned} \quad (A.5)$$

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

$$\begin{aligned} \int_{-1}^{1} y^l \int_{y}^{1} x^m (x-y)^{2\delta-1}dxdy &= \int_{-1}^{1} y^l \int_{0}^{1-y} (y+z)^m z^{2\delta-1}dzdy \\ &= \int_{-1}^{1} y^l \left\{ \sum_{i=0}^{m} \binom{m}{i} y^{m-i} \int_0^{1-y} z^{2\delta+i-1}dz \right\} dy \\ &= \sum_{i=0}^{m} \binom{m}{i} \frac{1}{2\delta+i} \left\{ \int_{-1}^{1} y^{l+m-i}(1-y)^{2\delta+i}dy \right\}. \end{aligned} \quad (A.6)$$

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

$$\begin{aligned} \int_{-1}^{1} y^{l+m-i}(1-y)^{2\delta+i}dy &= -\int_{1}^{-1} (-z)^{l+m-i}(1+z)^{2\delta+i}dz \\ &= (-1)^i \int_{-1}^{1} z^{l+m-i}(1+z)^{2\delta+i}dz. \end{aligned} \quad (A.7)$$

The proof of Lemma 2 is finished by inserting (A.7) into (A.6). ◇

Proof of Theorem 2.

i) The proof of (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 (A.6). Set $z = 1 - y$, we have

$$\begin{aligned} \int_{-1}^{1} y^{l+m-i}(1-y)^{2\delta+i}dy &= -\int_{2}^{0}(1-z)^{l+m-i}z^{2\delta+i}dz \\ &= \left\{\sum_{j=0}^{l+m-i}(-1)^j\binom{l+m-i}{j}\int_0^2 z^{2\delta+i+j}dz\right\} \\ &= \sum_{j=0}^{l+m-i}(-1)^j\binom{l+m-i}{j}\frac{2^{2\delta+i+j+1}}{2\delta+i+j+1}. \end{aligned} \tag{A.8}$$

Combining (A.6) and (A.8) we obtain

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

Following Lemma 2 we have

$$\begin{aligned} T_{l,m} &= 2\int_{-1}^{1} y^l \int_y^1 x^m(x-y)^{2\delta-1}dxdy \\ &= 2\sum_{i=0}^{m}\binom{m}{i}\frac{1}{2\delta+i}\sum_{j=0}^{l+m-i}(-1)^j\binom{l+m-i}{j}\frac{2^{2\delta+i+j+1}}{2\delta+i+j+1}. \end{aligned} \tag{A.9}$$

Theorem 2 is proved. ◇

Proof of Corollary 2. Following (12) we have

$$V(\delta) = \frac{9}{4\pi}\Gamma(1-2\delta)\sin(\pi\delta)T_{1,1}. \tag{A.10}$$

Furthermore, following (13)

$$\begin{aligned} T_{1,1} &= 2\sum_{i=0}^{1}\frac{1}{2\delta+i}\sum_{j=0}^{2-i}(-1)^j\binom{2-i}{j}\frac{2^{2\delta+j+i+1}}{2\delta+j+i+1} \\ &= \frac{1}{\delta}\left[\frac{2^{2\delta+1}}{2\delta+1} - 2\frac{2^{2\delta+2}}{2\delta+2} + \frac{2^{2\delta+3}}{2\delta+3}\right] \\ &\quad + \frac{2}{2\delta+1}\left[\frac{2^{2\delta+2}}{2\delta+2} - \frac{2^{2\delta+3}}{2\delta+3}\right]. \end{aligned} \tag{A.11}$$

Straightforward calculation leads to

$$T_{1,1} = \frac{(1-2\delta)2^{2\delta+1}}{\delta(2\delta+1)(2\delta+3)} \tag{A.12}$$

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

References

Altman, N.S. (1990). Kernel smoothing with correlated errors. *J. Am. Statist. Assoc.*, 85, 749–759.

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman & Hall, New York.

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.

Beran, J. and Feng, Y. (2002a). Local polynomial fitting with long-memory, short-memory and antipersistent errors. *Annals Inst. Statist. Math.*, 54, 291–311.

Beran, J. and Feng, Y. (2002b). Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties. *J. Comput. and Graph. Statist.*, 11, 690–713.

Beran, J. and Feng, Y. (2002c). SEMIFAR models - A semiparametric framework for modelling trends, long-range dependence and nonstationarity. *Comput. Statist. and Data Anal.*, 40, 393–419.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications.* Chapman & Hall, London.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *J. Time Ser. Anal.*, 1, 15-30.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, New York.

Hall, P. and Hart, J.D. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.*, 36, 339–351.

Hart, J.D. (1991). Kernel regression estimation with time series errors. *J. R. Statist.Soc. Ser. B*, 53, 173–188.

Hastie, T. and Loader, C. (1993). Local regression: Automatic kernel carpentry (with discussion). *Statist. Sci.*, 8, 120–143.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, 68, 165-176.

Müller, H.G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.*, 82, 231–238.

Müller, H.G. (1988). *Nonparametric Analysis of Longitudinal Data.* Springer-Verlag, Berlin.

Müller, H.G. (1991). On the Boundary Kernel Methods for Non-parametric Curve Estimation Near the Endpoints. *Scand. J. Statist.*, 20, 313–328.

Müller, H.G. and Wang, J.L. (1994). Hazard Rate Estimation Under Random Censoring with Varying Kernels and Bandwidths. *Biometrics*, 50, 61–76.

Ray, B.K. and Tsay, R.S. (1997). Bandwidth selection for kernel regression with long-range dependence. *Biometrika*, 84, 791–802.

Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, 22, 1346–1370.

Dr. Yuanhua Feng
Department of Mathematics and Statistics
University of Konstanz
D-78457 Konstanz, Germany

Email: yuanhua.feng@uni-konstanz.de
Tel. +49-7531-88-3763
Fax. +49-7531-88-2407

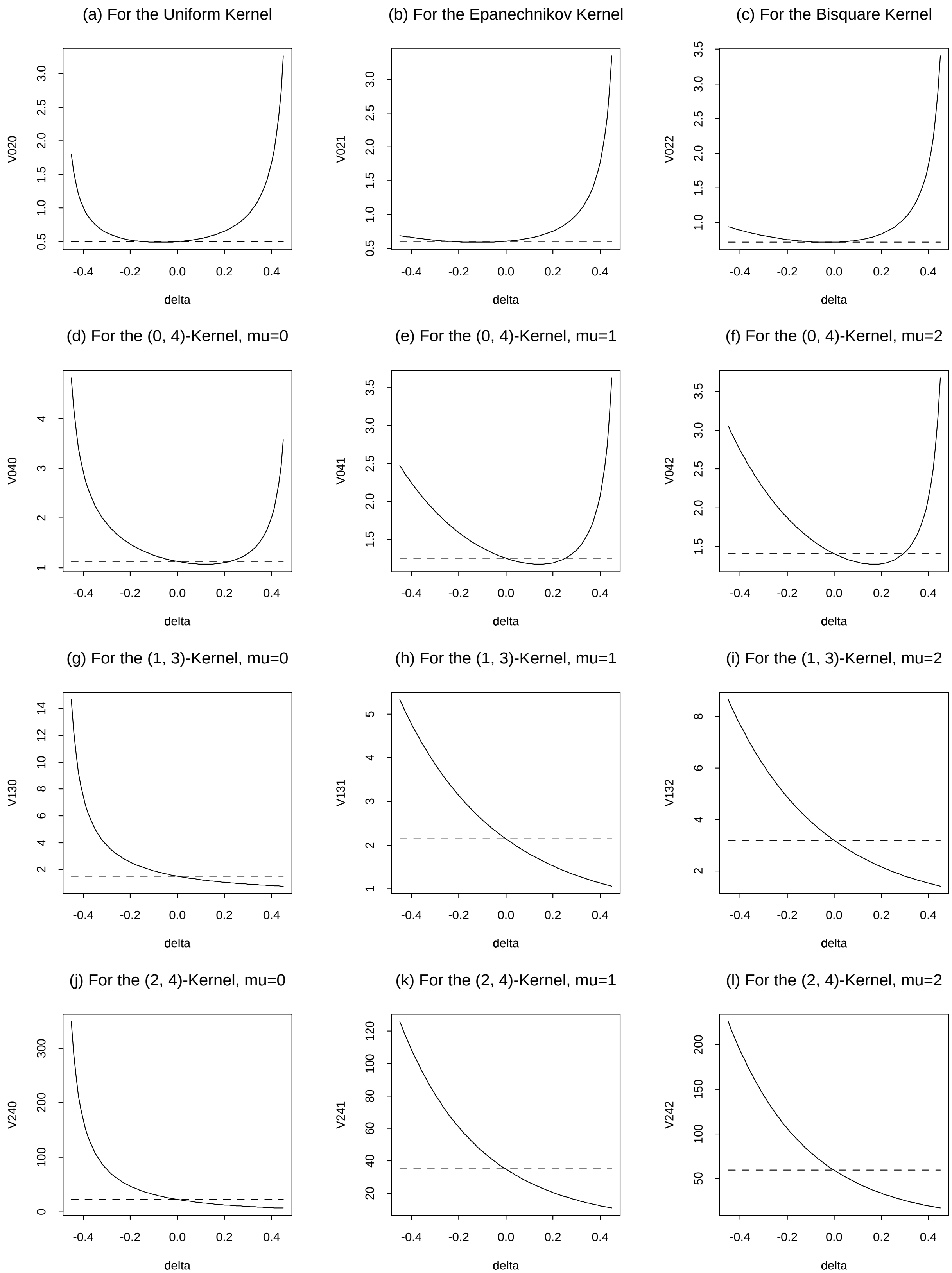

Figure 1: Kernel dependent functions $V(\delta)$ on $[-0.45, 0.45]$ (solid curves), where the dashed lines show the corresponding values $V(0) = R(K_{(\nu,k)})$.

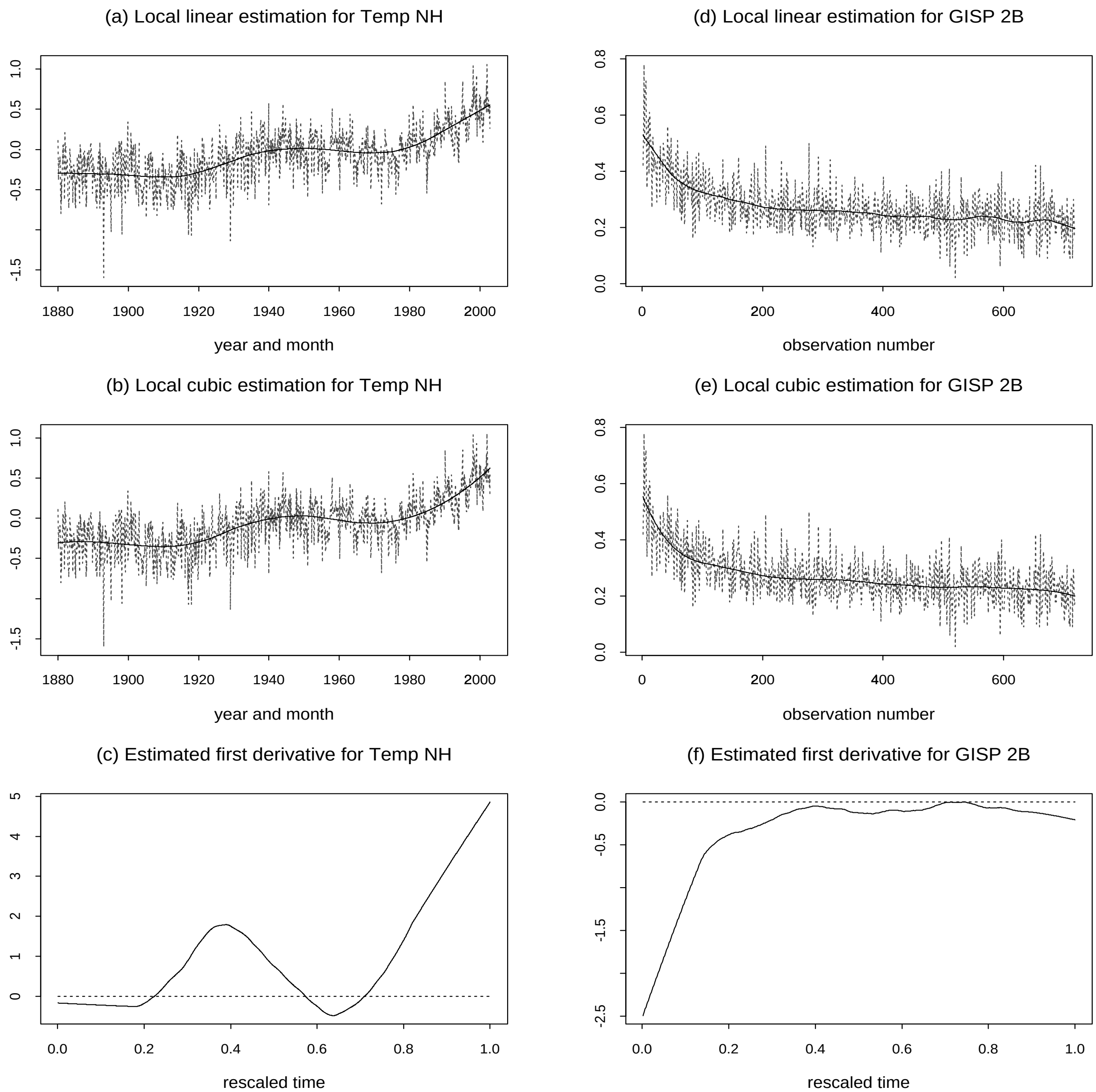

Figure 2: Data-driven local linear (upper) and local cubic (middle) estimates of g, and local quadratic estimate of g' (below) for the Temp NH (left) and GISP 2B (right) series.

Table 1: 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