
Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

# Iterative Plug-In Algorithms for SEMIFAR Models—Definition, Convergence, and Asymptotic Properties

Jan Beran & Yuanhua Feng

# Iterative Plug-In Algorithms for SEMIFAR Models—Definition, Convergence, and Asymptotic Properties

Jan BERAN and Yuanhua FENG

This article proposes data-driven algorithms for fitting SEMIFAR models. The algorithms combine the data-driven estimation of the nonparametric trend and maximum likelihood estimation of the parameters. Convergence and asymptotic properties of the proposed algorithms are investigated. A large simulation study illustrates the practical performance of the methods.

**Key Words:** Antipersistence; Bandwidth selection; Fractional ARIMA; Long-range dependence; Nonparametric regression; Semiparametric models.

## 1. INTRODUCTION

### 1.1 THE SEMIFAR MODEL

The so-called semiparametric fractional autoregressive (SEMIFAR) model, introduced by Beran (1999), provides a unified approach that allows for simultaneous modeling of deterministic trends, stochastic trends and stationary short memory, long memory, and antipersistent components. This article proposes several data-driven algorithms for estimating the SEMIFAR model. Asymptotic properties of the methods are derived. The practical performance is investigated in an extended simulation study.

A SEMIFAR model (Beran 1999) is a Gaussian process $Y_i$ with an existing smallest integer $m \in \{0, 1\}$ such that

$$\phi(B)(1-B)^{\delta}\{(1-B)^m Y_i - g(t_i)\} = \epsilon_i, \tag{1.1}$$

Jan Beran is Professor, and Yuanhua Feng is Senior Research Fellow, Department of Mathematics and Statistics, University of Konstanz, 78457 Konstanz, Germany (E-mail addresses: Jan.Beran@uni-konstanz.de *and* Yuanhua.Feng@uni-konstanz.de).


*Journal of Computational and Graphical Statistics, Volume 11, Number 3, Pages 690–713*
*DOI: 10.1198/106186002420*

where $t_i = (i/n)$, $\delta \in (-0.5, 0.5)$, $g$ is a smooth function on $[0, 1]$, $B$ is the backshift operator, $\phi(x) = 1 - \sum_{j=1}^{p} \phi_j x^j$ is a polynomial with roots outside the unit circle, and $\epsilon_i$ $(i = \ldots, -1, 0, 1, 2, \ldots)$ are iid zero mean normal with var $(\epsilon_i) = \sigma_\epsilon^2$. Where the fractional difference $(1 - B)^\delta$, introduced by Granger and Joyeux (1980) and Hosking (1981), is defined by $(1-B)^\delta = \sum_{k=0}^{\infty} \beta_k(\delta) B^k$ with $\beta_k(\delta) = (-1)^k \Gamma(\delta+1)/[\Gamma(k+1)\Gamma(\delta-k+1)]$. Model (1.1) includes stationary ($m = 0$) and difference-stationary ($m = 1$) processes with or without deterministic trend, and covers the cases of short-range dependence ($\delta = 0$), long-range dependence ($\delta > 0$) and antipersistence ($\delta < 0$). See Beran (1999) and Beran and Ocker (1999, 2001) for detailed remarks on different special cases of model (1.1). The spectral density of $X_i = (1 - B)^m Y_i - g(t_i)$ has the form

$$f(\lambda) \sim c_f |\lambda|^{-\alpha} \qquad (\text{as } \lambda \to 0) \tag{1.2}$$

with $\alpha = 2\delta$, where $c_f$ is the value of the spectral density of an AR($p$) process $Z_i := (1 - B)^\delta X_i$ at the origin. Hence, $X_i$ has long memory if $\delta > 0$. In this case the autocovariances $\gamma(k)$ of $X_i$ are proportional to $k^{2\delta-1}$ (as $k \to \infty$) and hence are nonsummable. If $\delta = 0$, $X_i$ has short-memory and spectral density $f(\lambda)$ converges to a positive constant $c_f$ at the origin with $c_f = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma(k)$. If $\delta < 0$, then the spectral density $f(\lambda)$ of $X_i$ converges to zero at the origin (so-called "antipersistence"). In this case we have $\sum_{k=-\infty}^{\infty} \gamma(k) = 0$. For details on time series with long-memory see Beran (1994) and references therein. All of the discussions in this article are valid for the whole range $\delta \in (-0.5, 0.5)$.

## 1.2 ESTIMATION

The estimation of SEMIFAR models consists of two parts: nonparametric estimation of the trend $g$ and estimation of the parameters $m$, $\delta$, $p$, and $\phi_1, \ldots, \phi_p$. Note, in particular, that the integer differencing parameter $m$ is also estimated from the data. In this article the trend $g$ is estimated by a kernel method (Hall and Hart 1990; Beran 1999) defined by

$$\hat{g}(t; h) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{t - t_i}{h}\right)(1 - B)^{\hat{m}} Y_i, \tag{1.3}$$

where $\hat{m}$ is an estimate of $m$. Throughout the article, $K$ will be a symmetric positive second-order kernel with compact support. The parameter vector $\theta = (\sigma_\epsilon^2, \eta)$ with $\eta_1 = \delta + m$, $\eta_{j+1} = \phi_j$ $(j = 1, \ldots, p)$ are estimated by maximum likelihood, the order $p$ being chosen by minimizing the BIC. Since $\hat{g}$ depends on $\hat{m}$ and the bandwidth $h$ and the optimal bandwidth depends on the unknown value of $\theta$, an iterative procedure is needed that alternates between kernel smoothing and estimation of the parameters.

The algorithms proposed in this article rely on the following asymptotic expressions for the mean integrated squared error (MISE) and the optimal bandwidth (see Beran 1999; for $\delta \geq 0$ also see Hall and Hart 1990): Let $\Delta > 0$ be a small positive constant, which is introduced to avoid the so-called boundary effect of the kernel estimator, and define

$I(g'') = \int_{\Delta}^{1-\Delta}[g''(t)]^2 dt$ and $I(K) = \int_{-1}^{1} x^2 K(x)dx$. Then, as $n \to \infty, h \to 0, nh \to \infty$,

$$\begin{aligned} \text{MISE} &= E\left\{\int_{\Delta}^{1-\Delta} [\hat{g}(t) - g(t)]^2 dt\right\} \\ &= h^4 \frac{I(g'')I^2(K)}{4} + (nh)^{2\delta-1} V(1-2\Delta) \\ &\quad + o(\max(h^4, (nh)^{2\delta-1})), \end{aligned} \tag{1.4}$$

where $V$ is a constant depending on $c_f$ and the kernel function. Formulas for $V$ (with $\delta \in (-0.5, 0.5)$) may be found in Beran and Feng (2002).

For selecting the bandwidth we also need to estimate $g''$. Let $\hat{g}''(t; h_2)$ be a kernel estimator of $g''$ with a kernel $K_2$ and another bandwidth $h_2$, which is different from the bandwidth $h$ for estimating $g$. Throughout this article, $K_2$ will be a symmetric fourth-order kernel for estimating the second derivative with compact support (see, e.g., Gasser and Müller 1984). Assume that $g$ is at lest four times continuously differentiable and that $h_2 \to 0$, $(nh_2)^{1-2\delta} h_2^4 \to \infty$ as $n \to \infty$. Then we have

$$\text{MISE}(\hat{g}'') \doteq O(h_2^4) + O[(nh_2)^{2\delta-1} h^{-4}]. \tag{1.5}$$

## 1.3 AIM OF THIS ARTICLE

Data-driven bandwidth selection is a crucial problem in the practical use of nonparametric regression. Recent proposals for bandwidth selection in nonparametric regression with independent or short-range dependent data may be found, for example, in Müller (1985), Gasser, Kneip, and Köhler (1991), Härdle, Hall, and Marron (1992), Herrmann, Gasser, and Kneip (1992), Fan and Gijbels (1995), Ruppert, Sheather, and Wand (1995) and Heiler and Feng (1998).

Data-driven bandwidth selection in the presence of long memory or nonstationary errors is even more difficult, since spurious stochastic trends may be confounded with deterministic trends. A bandwidth selector for nonparametric regression with long-range dependence based on the iterative plug-in idea (Gasser et al. 1991) is proposed by Ray and Tsay (1997). The contributions of our article are:

1. Two new iterative algorithms are proposed in the context of SEMIFAR models, thus in particular including the possibility of stochastic trends (difference stationarity), antipersistence, and long memory.
2. Convergence of the algorithms is proved.
3. Asymptotic properties of the estimated bandwidths are derived.
4. In contrast to Gasser et al. and Ray and Tsay, our algorithms use an exponential inflation step, which leads to a better relative rate of convergence of the estimated bandwidth.
5. Finite sample behavior is studied in a large simulation study.

## 2. NOTATION AND GENERAL CONSIDERATIONS

The optimal bandwidth, which minimizes the MISE, will be denoted by $h_{\mathrm{M}}$. The so-called asymptotically optimal bandwidth, $h_{\mathrm{A}}$, that minimizes the asymptotic MISE, is given by

$$h_{\mathrm{A}} = C \cdot n^{(2\delta-1)/(5-2\delta)}, \tag{2.1}$$

with

$$C = \left( \frac{(1-2\delta)V(1-2\Delta)}{I(g'')I^2(K)} \right)^{1/(5-2\delta)}. \tag{2.2}$$

Here it is assumed that $I(g'') > 0$. When the uniform kernel is used, the constant $C$ in (2.1) has the explicit form

$$C = \left( \frac{9(1-2\delta)\nu(\delta)(1-2\Delta)c_f}{I(g'')} \right)^{1/(5-2\delta)} \tag{2.3}$$

with $c_f$ as defined before and

$$\nu(\delta) = \frac{2^{2\delta}\Gamma(1-2\delta)\sin(\pi\delta)}{\delta(2\delta+1)} \tag{2.4}$$

for all $-0.5 < \delta < 0.5$ (see Beran 1999).

Plug-in estimators for $h_{\mathrm{M}}$ use formula (2.1), replacing the unknown constants $\delta$, $V$ as well as $I(g'')$ by some consistent estimators. Note that the estimation of $V$ is equivalent to that of $c_f$. The quantities $\delta$ and $V$ may be estimated root-$n-$consistently by maximum likelihood. Hence the key problem is to estimate $I(g'')$. From here on denote $I(g'')$ by $I$ and let it be estimated as follows

$$\hat{I} = n^{-1} \sum_{i=[n\Delta]}^{n-[n\Delta]} \{\hat{g}''(t_i; h_2)\}^2. \tag{2.5}$$

Properties of $\hat{I}$ are investigated by Beran and Feng (2002). Under the assumptions of Equation (1.5) we have

$$E[\hat{I} - I] \doteq h_2^2 \frac{I(K_2)}{12} \int_{\Delta}^{1-\Delta} g''(t)g^{(4)}(t)dt + (nh_2)^{2\delta-1}h_2^{-4}V, \tag{2.6}$$

and

$$\mathrm{var}\,[\hat{I}] \doteq o[(nh_2)^{(4\delta-2)}h_2^{-8}] + O(n^{2\delta-1}). \tag{2.7}$$

The mean squared error (MSE) of $\hat{I}$ is dominated by the squared bias

$$\mathrm{MSE}\{\hat{I}\} \doteq \left\{ h_2^2 \frac{I(K_2)}{12} \int_{\Delta}^{1-\Delta} g''(t)g^{(4)}(t)dt + (nh_2)^{2\delta-1}h_2^{-4}V \right\}^2.$$

The optimal bandwidth for estimating $I$ which minimizes MSE$\{\hat{I}\}$ is

$$h_2^{\text{opt}} = O(n^{(2\delta-1)/(7-2\delta)}).$$

Note that the bandwidth which minimizes MISE$(\hat{g}'')$ is, however,

$$h_2^0 = O(n^{(2\delta-1)/(9-2\delta)}).$$

Following the iterative plug-in idea of Gasser et al. (1991), in the $j$th iteration, $I$ is estimated with a bandwidth $h_{2,j}$, which is obtained from the bandwidth for estimating $g$ in the $j-1$th iteration, $h_{j-1}$ say, with a so-called inflation method. This idea can be adapted to data with different dependence structures (see Herrmann et al. 1992; Ray and Tsay 1997; Beran 1999; Beran and Ocker 2001). An iterative plug-in bandwidth selector is determined by a starting bandwidth $h_0$ and the inflation method with an inflation factor $\alpha$. In general, the process should begin with a very small $h_0$. Gasser et al. (1991) proposed the use of $h_0 = n^{-1}$. For data with long memory, $h_0$ should fulfill the condition $h_0 \to 0$, $nh_0 \to \infty$ as $n \to \infty$, since we have already to estimate $\delta$ and $V$ from the residuals at the first iteration. Hence Ray and Tsay (1997) used an $h_0$, which is selected following Herrmann et al. (1992) by assuming short memory. This article proposes the use of $h_0 = n^{-\beta}$ with $\frac{1}{3} \leq \beta < 1$. Such an $h_0$ satisfies the above condition and it is at the same time small enough. In fact we have $h_0 = o(h_{\text{A}})$ for all $\delta \in (-0.5, 0.5)$. Here we used $h_0 = n^{-5/7}$, which is of order $o(h_{\text{A}}^2)$ for all $\delta \in (-0.5, 0.5)$.

## 3. THE EIM-BANDWIDTH SELECTOR—DEFINITION AND ASYMPTOTIC PROPERTIES

### 3.1 THE MIM AND THE EIM APPROACHES

There are different ways to obtain $h_{2,j}$ from $h_{j-1}$. In Gasser et al. (1991), Herrmann et al. (1992), and Ray and Tsay (1997) the formula $h_{2,j} = c \cdot h_{j-1} n^{\alpha}$ is used. This is called multiplicative inflation method (MIM). Beran (1999) and Beran and Ocker (2001) suggested use of the formula $h_{2,j} = c \cdot (h_{j-1})^{\alpha}$. We call this exponential inflation method (EIM). For each inflation method one has also to choose the inflation factor $\alpha$. The iterative plug-in algorithm is motivated by fixed point search. So $\alpha$ and $c$ should be chosen in a way that $c \cdot h_{\text{A}} n^{\alpha} = h_2^{\text{opt}}$ by the MIM, or $c \cdot (h_{\text{A}})^{\alpha} = h_2^{\text{opt}}$ by the EIM, respectively. The optimal choice for $\alpha$ is $\alpha_{\text{opt}} = (2-4\delta)/[(5-2\delta)(7-2\delta)]$ for the MIM [see Herrmann and Gasser (1994) for the case with $\delta = 0$] and

$$\alpha_{\text{opt}} = (5-2\delta)/(7-2\delta)$$

for the EIM. The choice of $c$ does not affect the rate of convergence of $\hat{h}$. We will simply put $c = 1$.

There are two other reasonable choices of $\alpha$, namely (for the EIM)

$$\alpha_0 = (5 - 2\delta)/(9 - 2\delta) \qquad \text{and} \qquad \alpha_{\text{var}} = \frac{1}{2}.$$

The choice of $\alpha_0$ is motivated by the fact that $(h_{\text{A}})^{\alpha_0} = O(h_2^0)$ which minimizes the MISE$(\hat{g}'')$ in (1.5). The choice of $\alpha_{\text{var}}$ is motivated by balancing the terms that are due to the variance. More specifically, using $\alpha_{\text{var}}$ the square of the second term on the right hand side of (2.6) is of the same order $O(n^{2\delta-1})$ as the second term in (2.7).

Since $\alpha_{\text{opt}}$ leads to a bandwidth $h_2$ that minimizes the MSE$(\hat{I})$, the rate of convergence of the resulting $\hat{h}$ is optimal. Using $\alpha_0$, the rate of convergence of $\hat{h}$ lies between the two rates for $\alpha_{\text{opt}}$ and $\alpha_{\text{var}}$. Note that the $O(n^{2\delta-1})$ term in (2.7) provides a lower bound for the error in $\hat{I}$. Hence choosing $\alpha$ smaller than $\alpha_{\text{var}}$ does not improve the order of the variance of $\hat{I}$ anymore, whereas it increases the order of the bias. Thus, $\alpha_{\text{var}} = \frac{1}{2}$ is the smallest reasonable choice of $\alpha$ for EIM. In the procedure defined in Section 3.2 we use the EIM with $\alpha \geq \frac{1}{2}$. In Ray and Tsay (1997) the MIM with an $\alpha = (1 - 2\delta)/(10 - 4\delta)$ is used. Note that their choice of $\alpha$ is $\alpha_{\text{var}}$ for MIM.

## 3.2 DEFINITION OF THE EIM BANDWIDTH SELECTOR

Given $m$ (or a consistent estimate of $m$), we propose the following bandwidth selector (after taking the $m$th difference of the data):

1. Start with the bandwidth $h_0 = n^{-\beta}$ with $\frac{1}{3} \leq \beta < 1$ and set $j = 1$.
2. Estimate $g$ using $h_{j-1}$ and let $\hat{X}_i = Y_i - \hat{g}(t_i)$. Estimate $\delta$ and $V$ from $\hat{X}_i$ by maximum likelihood.
3. Set $h_{2,j} = (h_{j-1})^{\alpha}$ with $\frac{1}{2} \leq \alpha < 1$ and improve $h_{j-1}$ by

$$h_j = \left( \frac{1 - 2\hat{\delta}}{I^2(K)} \frac{(1 - 2\Delta)\hat{V}}{\hat{I}(g''(t; h_{2,j}))} \right)^{1/(5-2\hat{\delta})} \cdot n^{(2\hat{\delta}-1)/(5-2\hat{\delta})}. \tag{3.1}$$

4. Increase $j$ by 1 and repeat Steps 2 and 3 until convergence is reached or until a given number of iterations has been done. And set $\hat{h} = h_j$.

The required number of iterations for obtaining a satisfactory bandwidth selector depends on $\delta$, $h_0$, the inflation method, and $\alpha$. See Herrmann and Gasser (1994), Ray and Tsay, (1997) and Beran and Feng (2002) for the idea to calculate it. In particular, the MIM requires considering more iterations than the EIM (see Beran and Feng 2002).

## 3.3 ASYMPTOTIC PROPERTIES

In this article, $\hat{V}$ and $\hat{\delta}$ will be assumed to be both $\sqrt{n}$-consistent. $\sqrt{n}$-consistent estimates of $\hat{V}$ and $\hat{\delta}$ may be obtained, for example, following the maximum likelihood approach proposed by Beran (1999). The difference between the asymptotically optimal bandwidth $h_{\text{A}}$ and the optimal bandwidth $h_{\text{M}}$ provides a natural lower bound for the rate of convergence of a plug-in bandwidth selector. Results on $h_{\text{A}} - h_{\text{M}}$ in the case of nonparametric regression with short memory (i.e., $\delta = 0$) may be found, for example, in Gasser et al. (1991), Herrmann et al. (1992), and Herrmann and Gasser (1994). In the following, unified results will be given for all $\delta \in (-0.5, 0.5)$ under regularity conditions, which extends the results in Proposition 1 in Herrmann and Gasser (1994), where only the case of iid errors was considered.

**Proposition 1.** *Assume that $g$ is at least four times continuously differentiable and that $h \to 0$, $nh \to \infty$ as $n \to \infty$. Then we have, for all $\delta \in (-0.5, 0.5)$,*

$$(h_{\mathrm{A}} - h_{\mathrm{M}})/h_{\mathrm{M}} \doteq O(h_{\mathrm{M}}^2). \tag{3.2}$$

The proof of Proposition 1 is given in the Appendix.

Observe that $\hat{I} - I$ is at least of the order $O_p\{(h_2^{\mathrm{opt}})^2\}$ and $h_{\mathrm{M}} = o(h_2^{\mathrm{opt}})$ so that $(h_{\mathrm{A}} - h_{\mathrm{M}})/h_{\mathrm{M}} = o_p(\hat{I} - I)$ for all $\delta \in (-0.5, 0.5)$. This implies that the rate of convergence of $\hat{h}$ depends only on the error in $\hat{I}$ (see the Appendix). The following lemma gives the rates of convergence of $\hat{h}$ under different choices of $\alpha$.

**Lemma 1.** *Assume that $g$ is at least four times continuously differentiable and that $\hat{\delta}$ and $\hat{V}$ are both $\sqrt{n}$-consistent. Then we have*

*1. For $\alpha = \alpha_{\mathrm{var}} = \frac{1}{2}$*

$$\hat{h} = h_{\mathrm{M}} \left\{1 + O(n^{(2\delta-1)/(5-2\delta)}) + O_p(n^{(2\delta-1)/2}) + O_p(n^{-1/2})\right\}. \tag{3.3}$$

*2. For $\alpha = \alpha_0 = (5 - 2\delta)/(9 - 2\delta)$*

$$\hat{h} = h_{\mathrm{M}} \left\{1 + O(n^{2(2\delta-1)/(9-2\delta)}) + O_p(n^{4(2\delta-1)/(9-2\delta)}) + O_p(n^{-1/2})\right\}. \tag{3.4}$$

*3. For $\alpha = \alpha_{\mathrm{opt}} = (5 - 2\delta)/(7 - 2\delta)$*

$$\hat{h} = h_{\mathrm{M}} \left\{1 + O(n^{2(2\delta-1)/(7-2\delta)}) + O_p(n^{2(2\delta-1)/(7-2\delta)}) + O_p(n^{-1/2})\right\}. \tag{3.5}$$

The proof of Lemma 1 is given in the Appendix. The following remarks clarify the results of this Lemma.

*Remark 1.* The $O_p(n^{-1/2})$ term in (3.3) to (3.5) is due to the error in $\hat{V}$ and $\hat{\delta}$.

*Remark 2.* The proposal in Ray and Tsay (1997) has the same asymptotic behavior as given in (3.3). The variance term $O(n^{(2\delta-1)/2})$ is in comparison with the $O(n^{(2\delta-1)/(5-2\delta)})$ term, which is due to the bias, asymptotically negligible.

*Remark 3.* Note that the order of the variance term in (3.3) reaches the lower bound $O_p(n^{\delta-\frac{1}{2}})$, whereas the corresponding term in (3.4) and (3.5) are of a larger order. On the other hand the bias term in (3.3) is of a larger order than in (3.4) and (3.5). The overall rate of convergence is faster in (3.4) and (3.5) but (3.3) is more stable in the sense of having the smallest variance. The best overall rate is achieved in (3.5).

*Remark 4.* For $\delta = 0$, the rate in (3.5) is $n^{-2/7}$. This is the same as obtained by Ruppert et al. (1995) for iid data. For positive $\delta$, the rate is slower than $n^{-2/7}$ and for negative $\delta$

faster than $n^{-2/7}$. Moreover, it tends to (but is always slower than) $n^{-1/2}$ as $\delta \to -0.5$. The $O_p(n^{-1/2})$ term in (3.3) through (3.5) is asymptotically negligible.

*Remark 5.* Equations (3.4) and (3.5) give the rates of convergence for the bandwidth selectors used in Algorithms A and B proposed in the next section, respectively.

## 4. DATA-DRIVEN ALGORITHMS

This section deals with data-driven algorithms for estimating the SEMIFAR models. Convergence and asymptotic properties of three algorithms are obtained. The first algorithm (Algorithm A, or AlgA ) relies on a full search with respect to $d$, and was originally proposed (in a slightly different version) in Beran (1999) and Beran and Ocker (2001). No theoretical results were derived there, however. Algorithms B and C are two new algorithms that run much faster than Algorithm A, since no full search is needed. In the following the true unknown parameter vector will be denoted by $\theta^0 = (\sigma^2_{\epsilon,0}, \delta^0 + m^0, \phi^0_1, \ldots, \phi^0_{p^0})$.

### 4.1 ALGORITHM A

AlgA is an adaptation of the procedure given by Beran (1995), replacing $\hat{\mu}$ by the kernel estimator $\hat{g}$. It makes use of the fact that $d$ is the only additional parameter, besides the autoregressive parameters, so that a systematic search with respect to $d$ can be made. Let $\Delta_0$ be a small positive number and $\alpha = \alpha_0$. Then the procedure is defined as follows:

**Algorithm A**:

Step 1: Define $L$ = maximal order of $\phi(B)$ that will be tried, and a sufficiently fine grid $G \in (-0.5, 1.5)\backslash\{0.5\}$. Then, for each $p \in \{0, 1, \ldots, L\}$, carry out Steps 2 through 4.

Step 2: For each $d \in G$, set $m = [d + 0.5]$, $\delta = d - m$, and $U_i(m) = (1 - B)^m Y_i$, and carry out Step 3.

Step 3: Carry out the following iteration:

Step 3a: Let $h_0 = \Delta_0 \min(n^{(2\delta-1)/(5-2\delta)}, 0.5)$ and set $j = 1$.

Step 3b: Calculate $\hat{g}(t_i; m)$ using the bandwidth $h_{j-1}$. Set $\hat{X}_i = U_i(m) - \hat{g}(t_i; m)$.

Step 3c: Set $\tilde{e}_i(d) = \sum_{j=0}^{i-1} \beta_j(\delta)\hat{X}_{i-j}$, where the coefficients $\beta_j$ are defined before.

Step 3d: Estimate the autoregressive parameters $\phi_1, \ldots, \phi_p$ from $\tilde{e}_i(d)$ and obtain the estimates $\hat{\sigma}^2_\epsilon = \hat{\sigma}^2_\epsilon(d; j)$ and $\hat{c}_f = \hat{c}_f(j)$. Estimation of the parameters can be done, for instance, by using the S-Plus function `ar.burg` or `arima.mle`. If $p = 0$, set $\hat{\sigma}^2_\epsilon$ equal to $n^{-1}\sum \tilde{e}^2_i(d)$ and $\hat{c}_f$ equal to $\hat{\sigma}^2_\epsilon/(2\pi)$.

Step 3e: Set $h_{2,j} = (h_{j-1})^\alpha$ with $\alpha = \alpha_0 = (5 - 2\delta)/(9 - 2\delta)$, improve $h_{j-1}$ by

$$h_j = \left(\frac{1-2\delta}{I^2(K)} \frac{(1-2\Delta)\hat{V}}{\hat{I}(g''(t; h_{2,j}))}\right)^{1/(5-2\delta)} \cdot n^{(2\delta-1)/(5-2\delta)}. \tag{4.1}$$

Step 3f: Increase $j$ by one and repeat Steps 3b to 3e until convergence is reached or until a given number of iterations has been done. This yields for each $d \in G$ separately, the ultimate value of $\hat{\sigma}^2_\epsilon(d)$, as a function of $d$.

Step 4: Define $\hat{d}$ to be the value of $d$ for which $\hat{\sigma}^2_\epsilon(d)$ is minimal. This together with the corresponding estimates of the AR parameters, yields an information criterion, for example, $\text{BIC}(p) = n \log \hat{\sigma}^2_\epsilon(p) + p \log n$, as a function of $p$ and the corresponding values of $\hat{\theta}$ and $\hat{g}$ for the given order $p$.

Step 5: Select the order $p$ that minimizes $\text{BIC}(p)$. This yields the final estimates of $\theta^0$ and $g$.

Note that a "small" constant $\Delta_0$ is used in order that the starting bandwidth is not too large. In our implementation we used $\Delta = 0.1$ and $\Delta_0 = 0.2$. This means that, in the first iteration, at most 20% of the observations are used for estimating $g$ at each point and $t_i \in [\Delta, 1-\Delta]$ are all interior points. Note that in Step 3, the (trial) values of $\delta$ and $m$ are fixed. Note also that, if $\delta = \delta^0$, then $h_0$ in Step 3 is of the optimal order so that $h_1$ is already consistent. In the second iteration the effect of $h_0$ will be clearly reduced. Further iterations improve the finite sample properties of $\hat{h}$. If $\delta \neq \delta^0$, the selected bandwidth in any iteration is not optimal in general.

The following lemma is needed to prove i) in Theorem 1. Denote by $\hat{h}(\delta)$ the bandwidth selected by AlgA for a given trial value $\delta$ in the case of $m = m^0$. Lemma 2 shows that $\hat{h}(\delta) \to 0$, $n\hat{h}(\delta) \to \infty$ in probability as $n \to \infty$.

**Lemma 2.** *Assume that the trial value of $m$ (in AlgA) is equal to $m^0$ and that the conditions of Lemma 1 hold. Then for each trial value $\delta$ there exists an order $\alpha_\delta$ such that $(1-2\delta)/(5-2\delta) \le \alpha_\delta < \frac{5}{9}$ and $\hat{h}(\delta) = O_p(n^{-\alpha_\delta})$.*

The proof of Lemma 2 is given in the appendix.

Under the conditions of Lemma 1 the following results for AlgA hold:

**Theorem 1.** *Let $\hat{h}$, $\hat{g}$ and $\hat{\theta}$ be obtained by AlgA with $L \ge p^0$, and let $0 < \Delta < 1/2$. Then*

1. *$\sqrt{n}(\hat{\theta} - \theta^o)$ converges in distribution to a zero mean normal random variable with a covariance matrix equal to the inverse Fisher information matrix;*
2. $$\hat{h} = h_{\text{M}}\{1 + O_p(n^{2(2\delta^0-1)/(9-2\delta^0)})\}. \tag{4.2}$$

The rate of convergence of the selected bandwidth given in (4.2) follows from (3.4). A sketched proof of Theorem 1 is given in the Appendix.

## 4.2 ALGORITHM B

The computing time for AlgA is very long, since the iterative procedure has to be carried out for each trial value $d \in G$. In the following we propose a much faster Algorithm B (AlgB), where all parameters, except for $p$ and $m$, are estimated directly from the residuals by maximizing the likelihood function. In our implementation, the S-Plus function `arima.fracdiff` was used.

The steps of AlgB are defined as follows:

**Algorithm B**:

Step 1: Obtain a bandwidth for estimating $m$:

Step 1a: Set $m = 1$. Calculate $U_i(m)$. Estimate $g$ from $U_i(m)$ with the starting bandwidth $h_0 = n^{-1/3}$. Calculate the residuals.

Step 1b: For each $p = 0, 1, \ldots, L$, where $L$ is as defined in AlgA, fit a FARIMA model to the residuals using the S-Plus function `arima.fracdiff`, where the order of the MA component is set equal to zero.

Step 1c: Select the best AR order $p$ following the BIC. Now we obtain estimates of all parameters except $m^0$.

Step 1d: Calculate the bandwidth $h_1$ following the procedure in Section 3 with $\alpha = \hat{\alpha}_{\text{opt}} = (5 - 2\hat{\delta})/(7 - 2\hat{\delta})$.

Step 1e: Set $L' = \hat{p}_0$.

Step 2: Estimate $m^0$:

Step 2a: Carry out Steps 1a to 1c with $h_1$ for $m = 0$ and $m = 1$ separately.

Step 2b: Select the best pair of $m$ and $p$ following the BIC. Now we obtain an estimate of all parameters, including $m^0$.

Step 2c: Set $m = \hat{m}^0$.

Step 3: Further iterations: Carry out further iterations with $L$ being replaced by $L'$, $m = \hat{m}^0$ and a new starting bandwidth $h_2 := n^{-5/7}$ until convergence is reached or a given number of iterations has been done.

Here $m = 1$ is used at the first iteration in order that the input of the S-Plus function `arima.fracdiff` is stationary. $m^0$ is selected at the second iteration. Afterwards, $\hat{m}^0$ is used. The estimate $\hat{m}^0$ is consistent, since $h_1 \to 0$, $nh_1 \to \infty$ as $n \to \infty$. For $\hat{p}_0$ selected at the first iteration in Step 1 we have $\hat{p}_0 \to p_0$ in probability, if $m^0 = 1$. If $m^0 = 0$, then $\hat{p}_0$ tends to the maximal order $L$ in probability, since now the error process is an ARMA($p$, 1), that is, an AR($\infty$) model. By selecting $m^0$ just one time and by setting $L' = \hat{p}_0$ at the end of step 1 much computing time can be saved. For AlgB we have

**Theorem 2.** *Let $\hat{h}$, $\hat{g}$ and $\hat{\theta}$ be obtained by AlgB with $L \geq p^0$, and let $0 < \Delta < 1/2$. Then the same results as given in Theorem 1 hold, except that now*

$$\hat{h} = h_{\mathrm{M}}\{1 + O_p(n^{2(2\delta^0-1)/(7-2\delta^0)})\}, \tag{4.3}$$

The proof of Theorem 2 is straightforward and is hence omitted. The rate of convergence of $\hat{h}$ obtained by AlgB is given by (3.5), which is faster than for AlgA, since $2(2\delta-1)/(9-2\delta) > 2(2\delta-1)/(7-2\delta)$ for all $\delta \in (-0.5, 0.5)$.

## 4.3 ALGORITHM C

The iteration at Step 1 is carried out so that $h_1$ adapts automatically to the structure of $g$ and the variation in the data. However, this starting bandwidth is quite large, which will sometimes result in $\hat{m}^0 = 0$ in the case when $m^0 = 1$. This motivates us to propose the following algorithm by using a smaller $h_0$ at the beginning and carrying out more iterations at Step 1:

**Algorithm C**:

Let $h_0 = n^{-1/3}$ at step 1 of AlgB be replaced by $h_0 = n^{-5/7}$. Carry out similarly the iteration six times using $m = 1$. The bandwidth $h_6$ is then used at Step 2 to select $m^0$. Carry out step 3 as in AlgB with $h_7$ selected at Step 2, if $\hat{m}^0 = 1$, or with $h_7 = n^{-5/7}$ otherwise.

The basic idea behind Algorithm C (AlgC) is as follows. If $m^0 = 1$, then $h_6$ obtained at the end of Step 1 is already a good estimate of $h_{\mathrm{M}}$. The estimation of $m$ using $h_6$ will have high accuracy. In the case $m^0 = 0$, $h_6$ will be a bandwidth adapted to the structure of $g$ and the variation in the data. So that it can be used for selecting $m^0$. The computing time of AlgC is slightly longer than for AlgB. It is clear that the estimates obtained by these two algorithms have the same asymptotic properties.

# 5. SIMULATION

## 5.1 DESCRIPTION OF THE SIMULATION STUDY

To show the practical performance of the data-driven SEMIFAR models, a large simulation has been done. The following three trend functions are used:

$$\begin{aligned} g_1(t) &= 2\tanh(5(t-0.5)), \\ g_2(t) &= 4\sin^2((t-0.5)\pi), \quad \text{and} \\ g_3(t) &= 2\sin(5(t-0.5)\pi) \end{aligned}$$

for $t \in [0, 1]$ (see Figures 1f through 3f). The range of these trends is kept the same. These trends are chosen as different as possible so that the practical performance of the proposed algorithms in different cases may be studied. The case without trend ($g_0 :\equiv 0$) is also included as a comparison.

Fifty parameter combinations with $m^0 \in \{0, 1\}$, $\delta^0 \in \{-0.4, -0.2, 0, 0.2, 0.4\}$, $\phi_1^0 \in \{-0.7, -0.3, 0, 0.3, 0.7\}$ were selected for the simulation. Here we have $p_0 = 0$ for $\phi_1^0 = 0$ and $p_0 = 1$ otherwise. The error process is standardized so that $\mathrm{var}(X_i) = 1$ in all cases. Two hundred replications were done for each parameter combination with two sample sizes $n = 500$ and $n = 1{,}000$. The simulations were carried out using AlgB and AlgC, separately. The maximal number of iterations was set equal to 20. Simulation using AlgA has not been done due to long computing time.

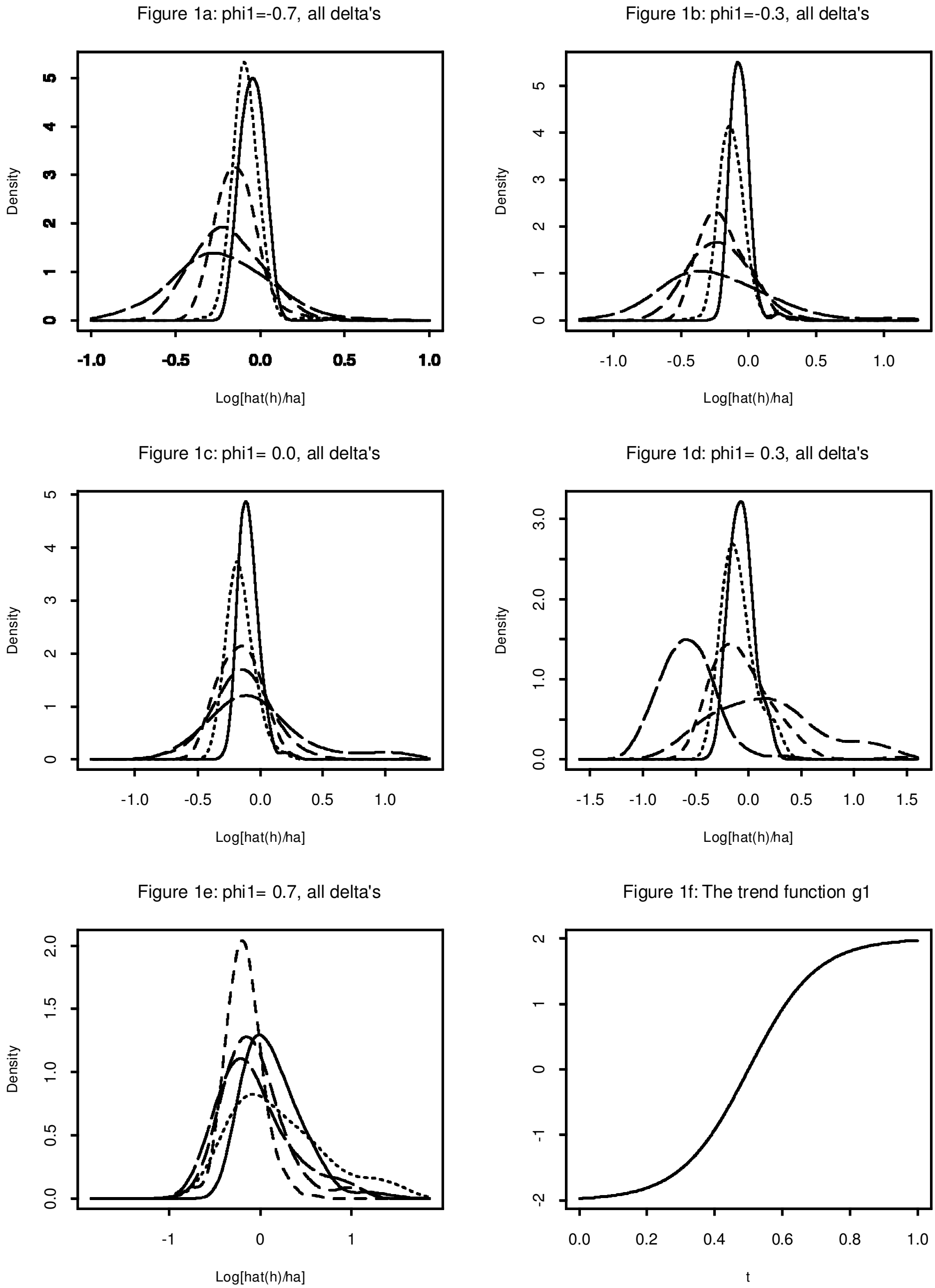


*Figure 1. Kernel densities of* $\log(\hat{h}/h_A)$ *selected by AlgB for* $g_1$ *with* $m^0 = 0$, $n = 500$*. Lines in Figures 1a through 1e are for* $\phi_1^0 = -0.7$ *to* $\phi_1^0 = 0.7$ *with all* $\delta^0$*'s (solid line:* $\delta^0 = -0.4$*; points:* $\delta^0 = -0.2$*; short dashes:* $\delta^0 = 0$*; middle dashes:* $\delta^0 = 0.2$ *and long dashes:* $\delta^0 = 0.4$*). The trend function* $g_1$ *is shown in Figure 1f.*

Table 1. Frequencies (in 200 replications) of Estimating $m^0$ and $p_0$ Correctly (for simulation using AlgB with $n = 500$ and $m^0 = 0$).

| | | $g_1$ | | $g_2$ | | $g_3$ | | $g_0$ | |
|---|---|---|---|---|---|---|---|---|---|
| $d^0$ | $\phi_1^0$ | $m^0$ | $p_0$ | $m^0$ | $p_0$ | $m^0$ | $p_0$ | $m^0$ | $p_0$ |
| −0.4 | −0.7 | 200 | 194 | 200 | 184 | 200 | 173 | 200 | 192 |
| −0.4 | −0.3 | 200 | 194 | 200 | 188 | 200 | 187 | 200 | 190 |
| −0.4 | 0 | 200 | 197 | 200 | 199 | 200 | 195 | 200 | 193 |
| −0.4 | 0.3 | 200 | 170 | 200 | 141 | 200 | 117 | 200 | 183 |
| −0.4 | 0.7 | 200 | 101 | 200 | 101 | 200 | 33 | 200 | 119 |
| −0.2 | −0.7 | 200 | 190 | 200 | 196 | 200 | 195 | 200 | 149 |
| −0.2 | −0.3 | 200 | 160 | 200 | 181 | 200 | 181 | 200 | 113 |
| −0.2 | 0 | 200 | 179 | 200 | 187 | 200 | 198 | 200 | 182 |
| −0.2 | 0.3 | 200 | 185 | 200 | 175 | 200 | 175 | 200 | 183 |
| −0.2 | 0.7 | 102 | 19 | 110 | 14 | 112 | 21 | 110 | 23 |
| 0 | −0.7 | 200 | 159 | 200 | 180 | 200 | 162 | 200 | 132 |
| 0 | −0.3 | 200 | 111 | 200 | 120 | 200 | 81 | 200 | 115 |
| 0 | 0 | 200 | 169 | 200 | 186 | 200 | 179 | 200 | 176 |
| 0 | 0.3 | 200 | 155 | 200 | 138 | 200 | 86 | 200 | 157 |
| 0 | 0.7 | 192 | 191 | 182 | 180 | 158 | 153 | 185 | 180 |
| 0.2 | −0.7 | 200 | 166 | 200 | 172 | 200 | 94 | 200 | 175 |
| 0.2 | −0.3 | 200 | 131 | 200 | 129 | 200 | 75 | 200 | 139 |
| 0.2 | 0 | 200 | 172 | 200 | 180 | 200 | 167 | 200 | 179 |
| 0.2 | 0.3 | 158 | 19 | 159 | 22 | 153 | 9 | 161 | 19 |
| 0.2 | 0.7 | 197 | 195 | 199 | 198 | 187 | 186 | 199 | 198 |
| 0.4 | −0.7 | 196 | 195 | 196 | 196 | 200 | 183 | 196 | 190 |
| 0.4 | −0.3 | 185 | 148 | 193 | 127 | 200 | 52 | 191 | 137 |
| 0.4 | 0 | 196 | 199 | 197 | 198 | 199 | 198 | 195 | 198 |
| 0.4 | 0.3 | 150 | 150 | 152 | 151 | 56 | 49 | 152 | 150 |
| 0.4 | 0.7 | 187 | 199 | 184 | 195 | 186 | 188 | 185 | 196 |

## 5.2 SUMMARY OF RESULTS

In the following a brief summary of the simulations with $n = 500$ using AlgB will be given. The results for $n = 1{,}000$ and for AlgC are similar and are hence omitted to save place. Detailed simulation results were reported by Beran and Feng (2000) as supplement of the current article. Tables 1 and 2 give frequencies (in 200 replications) of correctly estimating $m^0$ and $p_0$, respectively, for $m^0 = 0$ and $m^0 = 1$ separately. Here the results for $g_0$ are also given, since $\hat{m}^0$ and $\hat{p}_0$ are still root $n$ consistent for the case without trend. Tables 3 and 4 give the mean and standard deviation of $\hat{h}$ for $m^0 = 0$ and $m^0 = 1$, separately, together with $h_\mathrm{A}$ calculated from (2.1). Note that $h_\mathrm{A}$ does not depend on $m^0$. These results are only given for $g_1$ through $g_3$, since $\hat{h}$ is not consistent for $g_0$.

The short-memory component of the SEMIFAR model depends on the selection of $m^0$ and $p_0$. The selection of $m^0$ plays a more important role than that of $p_0$, since it determines, whether the first difference should be used in the further calculation. From Tables 1 and 2 we see that $m^0$ is easy to select. In most cases, $\hat{m}^0$ is always (or almost always) correct. Estimation of $m^0$ appears difficult for $m^0 = 0$ with $\delta = -0.2$ and $\phi_1^0 = 0.7$ and for $g_0$ with $m^0 = 1$. This means that in these cases it is difficult to decide, if $Y_i$ is stationary or not. In

Table 2. Frequencies (in 200 replications) of Estimating $m^0$ and $p_0$ Correctly (for simulation using AlgB with $n = 500$ and $m^0 = 1$).

| | | $g_1$ | | $g_2$ | | $g_3$ | | $g_0$ | |
|---|---|---|---|---|---|---|---|---|---|
| $d^0$ | $\phi_1^0$ | $m^0$ | $p_0$ | $m^0$ | $p_0$ | $m^0$ | $p_0$ | $m^0$ | $p_0$ |
| 0.6 | −0.7 | 200 | 193 | 200 | 187 | 200 | 200 | 165 | 190 |
| 0.6 | −0.3 | 200 | 195 | 200 | 196 | 200 | 192 | 91 | 135 |
| 0.6 | 0 | 200 | 199 | 200 | 198 | 200 | 192 | 191 | 194 |
| 0.6 | 0.3 | 200 | 59 | 200 | 6 | 220 | 110 | 15 | 187 |
| 0.6 | 0.7 | 200 | 188 | 200 | 179 | 200 | 9 | 183 | 191 |
| 0.8 | −0.7 | 200 | 199 | 200 | 194 | 200 | 200 | 187 | 187 |
| 0.8 | −0.3 | 199 | 163 | 200 | 186 | 200 | 186 | 50 | 11 |
| 0.8 | 0 | 200 | 197 | 200 | 200 | 200 | 196 | 187 | 187 |
| 0.8 | 0.3 | 197 | 160 | 200 | 34 | 200 | 33 | 38 | 191 |
| 0.8 | 0.7 | 199 | 189 | 200 | 194 | 200 | 81 | 158 | 158 |
| 1 | −0.7 | 200 | 196 | 200 | 192 | 200 | 200 | 175 | 170 |
| 1 | −0.3 | 200 | 129 | 200 | 135 | 200 | 96 | 45 | 25 |
| 1 | 0 | 200 | 193 | 200 | 199 | 200 | 169 | 178 | 176 |
| 1 | 0.3 | 199 | 167 | 200 | 149 | 200 | 7 | 172 | 185 |
| 1 | 0.7 | 200 | 171 | 200 | 197 | 199 | 141 | 132 | 131 |
| 1.2 | −0.7 | 200 | 180 | 200 | 196 | 200 | 200 | 171 | 157 |
| 1.2 | −0.3 | 200 | 123 | 200 | 107 | 200 | 39 | 80 | 55 |
| 1.2 | 0 | 200 | 185 | 200 | 198 | 200 | 200 | 182 | 176 |
| 1.2 | 0.3 | 200 | 184 | 200 | 182 | 200 | 42 | 190 | 188 |
| 1.2 | 0.7 | 200 | 156 | 200 | 167 | 200 | 191 | 102 | 96 |
| 1.4 | −0.7 | 200 | 158 | 200 | 190 | 200 | 200 | 176 | 133 |
| 1.4 | −0.3 | 200 | 108 | 200 | 109 | 200 | 33 | 146 | 106 |
| 1.4 | 0 | 200 | 178 | 200 | 187 | 200 | 200 | 180 | 155 |
| 1.4 | 0.3 | 200 | 190 | 200 | 195 | 200 | 9 | 179 | 172 |
| 1.4 | 0.7 | 200 | 140 | 200 | 138 | 200 | 185 | 136 | 87 |

these cases AlgC turned out to perform better than AlgB (see Beran and Feng, 2000).

The order $p_0$ is more difficult to select than $m^0$. There are mainly two reasons for this. First, different autoregressive models may have quite similar finite sample paths. Second, in some cases, it is difficult to separate autocorrelation from a complex trend like $g_3$, when $n$ is not large enough. Hence, $\hat{p}_0$ works worst for $g_3$. The rate of correctly estimated $p_0$ may be very low, even when $\hat{m}^0$ is correct. Note that model (b) in Beran, Bhansali, and Ocker (1998) is the same as the case without trend used in this article. Comparing the results here and those in Table 1 in Beran et al. (1998), we can find that the rate of correctly estimated $p_0$ is similar. In our case, however, estimation of $p_0$ is more difficult, because knowledge of a constant trend is not assumed.

Results in Tables 3 and 4 show that the proposed bandwidth selector works well in all cases, although $m^0$ and $p_0$ have also to be estimated simultaneously. The rate of convergence of $\hat{h}$ depends only on $\delta$ not on $\phi_1^0$. However, the finite sample performance of $\hat{h}$ strongly depends on both parameters. In general, the larger $\phi_1^0$ and/or $\delta$ the larger is the variation in $\hat{h}$. The performance of $\hat{h}$ also depends on the trend function. The selection of the bandwidth in the case of trend function $g_1$ is more difficult than for $g_2$ or $g_3$. Estimation of $m^0$ and

Table 3. Mean and Standard deviation of $\hat{h}$ (using AlgB with $n = 500$, $m^0 = 0$).

| | | $g_1$ | | | $g_2$ | | | $g_3$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $d^0$ | $\phi_1^0$ | $h_A$ | Mean | SD | $h_A$ | Mean | SD | $h_A$ | Mean | SD |
| −0.4 | −0.7 | 0.053 | 0.050 | 0.0039 | 0.039 | 0.040 | 0.0015 | 0.021 | 0.021 | 0.0009 |
| −0.4 | −0.3 | 0.065 | 0.061 | 0.0048 | 0.048 | 0.051 | 0.0013 | 0.026 | 0.027 | 0.0008 |
| −0.4 | 0 | 0.075 | 0.068 | 0.0059 | 0.055 | 0.058 | 0.0017 | 0.029 | 0.031 | 0.0007 |
| −0.4 | 0.3 | 0.086 | 0.081 | 0.0094 | 0.063 | 0.066 | 0.0036 | 0.034 | 0.035 | 0.0014 |
| −0.4 | 0.7 | 0.114 | 0.139 | 0.0563 | 0.084 | 0.106 | 0.0214 | 0.045 | 0.055 | 0.0063 |
| −0.2 | −0.7 | 0.059 | 0.054 | 0.0046 | 0.043 | 0.044 | 0.0018 | 0.022 | 0.022 | 0.0009 |
| −0.2 | −0.3 | 0.074 | 0.066 | 0.0074 | 0.053 | 0.055 | 0.0023 | 0.027 | 0.028 | 0.0011 |
| −0.2 | 0 | 0.084 | 0.072 | 0.0080 | 0.061 | 0.062 | 0.0038 | 0.031 | 0.032 | 0.0010 |
| −0.2 | 0.3 | 0.097 | 0.089 | 0.0145 | 0.070 | 0.073 | 0.0064 | 0.035 | 0.039 | 0.0023 |
| −0.2 | 0.7 | 0.125 | 0.176 | 0.1076 | 0.090 | 0.131 | 0.0365 | 0.046 | 0.082 | 0.0210 |
| 0 | −0.7 | 0.075 | 0.066 | 0.0083 | 0.053 | 0.054 | 0.0037 | 0.025 | 0.025 | 0.0013 |
| 0 | −0.3 | 0.094 | 0.079 | 0.0126 | 0.066 | 0.065 | 0.0076 | 0.032 | 0.032 | 0.0016 |
| 0 | 0 | 0.106 | 0.091 | 0.0144 | 0.075 | 0.076 | 0.0089 | 0.036 | 0.038 | 0.0029 |
| 0 | 0.3 | 0.120 | 0.120 | 0.0493 | 0.084 | 0.095 | 0.0208 | 0.041 | 0.050 | 0.0080 |
| 0 | 0.7 | 0.150 | 0.128 | 0.0267 | 0.106 | 0.105 | 0.0147 | 0.051 | 0.061 | 0.0113 |
| 0.2 | −0.7 | 0.102 | 0.086 | 0.0164 | 0.069 | 0.068 | 0.0089 | 0.031 | 0.036 | 0.0339 |
| 0.2 | −0.3 | 0.126 | 0.104 | 0.0226 | 0.086 | 0.083 | 0.0139 | 0.039 | 0.042 | 0.0258 |
| 0.2 | 0 | 0.140 | 0.125 | 0.0385 | 0.095 | 0.096 | 0.0137 | 0.043 | 0.047 | 0.0056 |
| 0.2 | 0.3 | 0.154 | 0.208 | 0.1157 | 0.105 | 0.134 | 0.0311 | 0.047 | 0.074 | 0.0184 |
| 0.2 | 0.7 | 0.180 | 0.179 | 0.0757 | 0.123 | 0.125 | 0.0193 | 0.055 | 0.065 | 0.0101 |
| 0.4 | −0.7 | 0.141 | 0.118 | 0.0423 | 0.093 | 0.090 | 0.0141 | 0.039 | 0.066 | 0.0942 |
| 0.4 | −0.3 | 0.164 | 0.139 | 0.0666 | 0.107 | 0.100 | 0.0222 | 0.045 | 0.066 | 0.1029 |
| 0.4 | 0 | 0.173 | 0.185 | 0.0913 | 0.114 | 0.122 | 0.0200 | 0.048 | 0.057 | 0.0453 |
| 0.4 | 0.3 | 0.181 | 0.105 | 0.0293 | 0.119 | 0.092 | 0.0157 | 0.050 | 0.069 | 0.0341 |
| 0.4 | 0.7 | 0.193 | 0.197 | 0.0923 | 0.126 | 0.133 | 0.0250 | 0.053 | 0.064 | 0.0140 |

$p_0$ also affects the accuracy of $\hat{h}$. For instance, if $m^0 = 0$ and $\hat{m}^0 = 1$, $\hat{h}$ is clearly larger than the optimal bandwidth (see the case with $\delta^0 = -0.2$ and $\phi_1^0 = 0.7$ in Table 3). In the case $m^0 = 1$ with $\hat{m}^0 = 0$, $\hat{h}$ is practically zero, when there is a trend in the data (see Beran and Feng 2000). Note also that $\hat{h}$ performs quite quite the same way for $m^0 = 0$ and $m^0 = 1$. Figures 1 through 3 show the estimated kernel densities of $\log(\hat{h}/h_A)$ from the 200 replications for each case with $m^0 = 0$, where densities for the same $\phi_1^0$ with different $\delta$'s are overlaid in the same plot. The same results for cases with $m^0 = 1$ are shown in Figures 4–6.

## 6. FINAL REMARKS

In this article three data-driven algorithms for fitting SEMIFAR models are investigated. The asymptotic behavior of the algorithms and the selected bandwidths is derived. In particular, it is shown theoretically that the new algorithms, AlgB and AlgC, perform better than AlgA (see Theorems 1 and 2). Simulations confirm the good performance of AlgB and AlgC and the potential usefulness of the SEMIFAR models for estimating and distinguishing deterministic trends and a stochastic component. The detailed simulation results in Beran

Table 4. Mean and Standard Deviation of $\hat{h}$ (using AlgB with $n = 500$, $m^0 = 1$).

| | | $g_1$ | | | $g_2$ | | | $g_3$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $d^0$ | $\phi_1^0$ | $h_A$ | Mean | SD | $h_A$ | Mean | SD | $h_A$ | Mean | SD |
| 0.6 | −0.7 | 0.053 | 0.050 | 0.0036 | 0.039 | 0.041 | 0.0013 | 0.021 | 0.021 | 0.0008 |
| 0.6 | −0.3 | 0.065 | 0.062 | 0.0060 | 0.048 | 0.051 | 0.0013 | 0.026 | 0.027 | 0.0007 |
| 0.6 | 0 | 0.075 | 0.068 | 0.0057 | 0.055 | 0.058 | 0.0019 | 0.029 | 0.031 | 0.0005 |
| 0.6 | 0.3 | 0.086 | 0.079 | 0.0106 | 0.063 | 0.065 | 0.0045 | 0.034 | 0.035 | 0.0013 |
| 0.6 | 0.7 | 0.114 | 0.108 | 0.0221 | 0.084 | 0.092 | 0.0159 | 0.045 | 0.055 | 0.0054 |
| 0.8 | −0.7 | 0.059 | 0.053 | 0.0042 | 0.043 | 0.044 | 0.0019 | 0.022 | 0.023 | 0.0009 |
| 0.8 | −0.3 | 0.074 | 0.065 | 0.0066 | 0.053 | 0.055 | 0.0023 | 0.027 | 0.029 | 0.0009 |
| 0.8 | 0 | 0.084 | 0.070 | 0.0073 | 0.061 | 0.061 | 0.0033 | 0.031 | 0.032 | 0.0008 |
| 0.8 | 0.3 | 0.097 | 0.091 | 0.0139 | 0.070 | 0.077 | 0.0078 | 0.035 | 0.040 | 0.0025 |
| 0.8 | 0.7 | 0.125 | 0.109 | 0.0215 | 0.090 | 0.095 | 0.0142 | 0.046 | 0.064 | 0.0140 |
| 1 | −0.7 | 0.075 | 0.061 | 0.0064 | 0.053 | 0.052 | 0.0029 | 0.025 | 0.025 | 0.0010 |
| 1 | −0.3 | 0.094 | 0.073 | 0.0112 | 0.066 | 0.062 | 0.0060 | 0.032 | 0.031 | 0.0012 |
| 1 | 0 | 0.106 | 0.084 | 0.0124 | 0.075 | 0.073 | 0.0069 | 0.036 | 0.037 | 0.0021 |
| 1 | 0.3 | 0.120 | 0.110 | 0.0268 | 0.084 | 0.092 | 0.0186 | 0.041 | 0.051 | 0.0047 |
| 1 | 0.7 | 0.150 | 0.136 | 0.0355 | 0.106 | 0.107 | 0.0162 | 0.051 | 0.063 | 0.0102 |
| 1.2 | −0.7 | 0.102 | 0.082 | 0.0167 | 0.069 | 0.066 | 0.0075 | 0.031 | 0.029 | 0.0020 |
| 1.2 | −0.3 | 0.126 | 0.100 | 0.0470 | 0.086 | 0.076 | 0.0121 | 0.039 | 0.035 | 0.0031 |
| 1.2 | 0 | 0.140 | 0.123 | 0.0323 | 0.095 | 0.094 | 0.0121 | 0.043 | 0.045 | 0.0036 |
| 1.2 | 0.3 | 0.154 | 0.124 | 0.0547 | 0.105 | 0.099 | 0.0238 | 0.047 | 0.063 | 0.0107 |
| 1.2 | 0.7 | 0.180 | 0.193 | 0.0883 | 0.123 | 0.140 | 0.0633 | 0.055 | 0.063 | 0.0067 |
| 1.4 | −0.7 | 0.141 | 0.133 | 0.0707 | 0.093 | 0.088 | 0.0140 | 0.039 | 0.038 | 0.0041 |
| 1.4 | −0.3 | 0.164 | 0.150 | 0.0820 | 0.107 | 0.108 | 0.0636 | 0.045 | 0.039 | 0.0058 |
| 1.4 | 0 | 0.173 | 0.196 | 0.1036 | 0.114 | 0.124 | 0.0508 | 0.048 | 0.051 | 0.0063 |
| 1.4 | 0.3 | 0.181 | 0.120 | 0.0562 | 0.119 | 0.096 | 0.0439 | 0.050 | 0.062 | 0.0054 |
| 1.4 | 0.7 | 0.193 | 0.230 | 0.1311 | 0.126 | 0.155 | 0.0949 | 0.053 | 0.059 | 0.0095 |

and Feng (2000) also indicated that in general, AlgB works better for $m^0 = 0$, while AlgC works better for $m^0 = 1$. The difference between AlgB and AlgC however also depends on the trend. For $g_1$ and $g_2$, their performance is almost the same. The simulation results also show that the estimates of the short- and long-memory parameters depend on each other. When the long-memory parameter is overestimated, the short-memory parameter will often be underestimated, and vice versa.

## APPENDIX: PROOFS

***Proof of Proposition 1:*** Denote the MISE by $M$. A well-known approximation of $M$ is the sum of the first two terms on the right-hand side of (1.4) under the assumption that $g$ is at least twice continuously differentiable. Denote this sum by $M_2$, we have $M_2 = O(h^4) + O\{(nh)^{2\delta-1}\}$. Now, $h_A$ is the minimizer of $M_2$. Note that $M_2$ is obtained based on the second order approximation of the bias $B_2 = O(h^2) + o(h^2)$. If $g^{(4)}$ is continuous then the bias may be approximated by $B_4 = O(h^2) + O(h^4) + o(h^4)$. We obtain a more accurate approximation of $M$, namely, $M_4 = O(h^4) + O(h^6) + O\{(nh)^{2\delta-1}\}$, due to the fact that

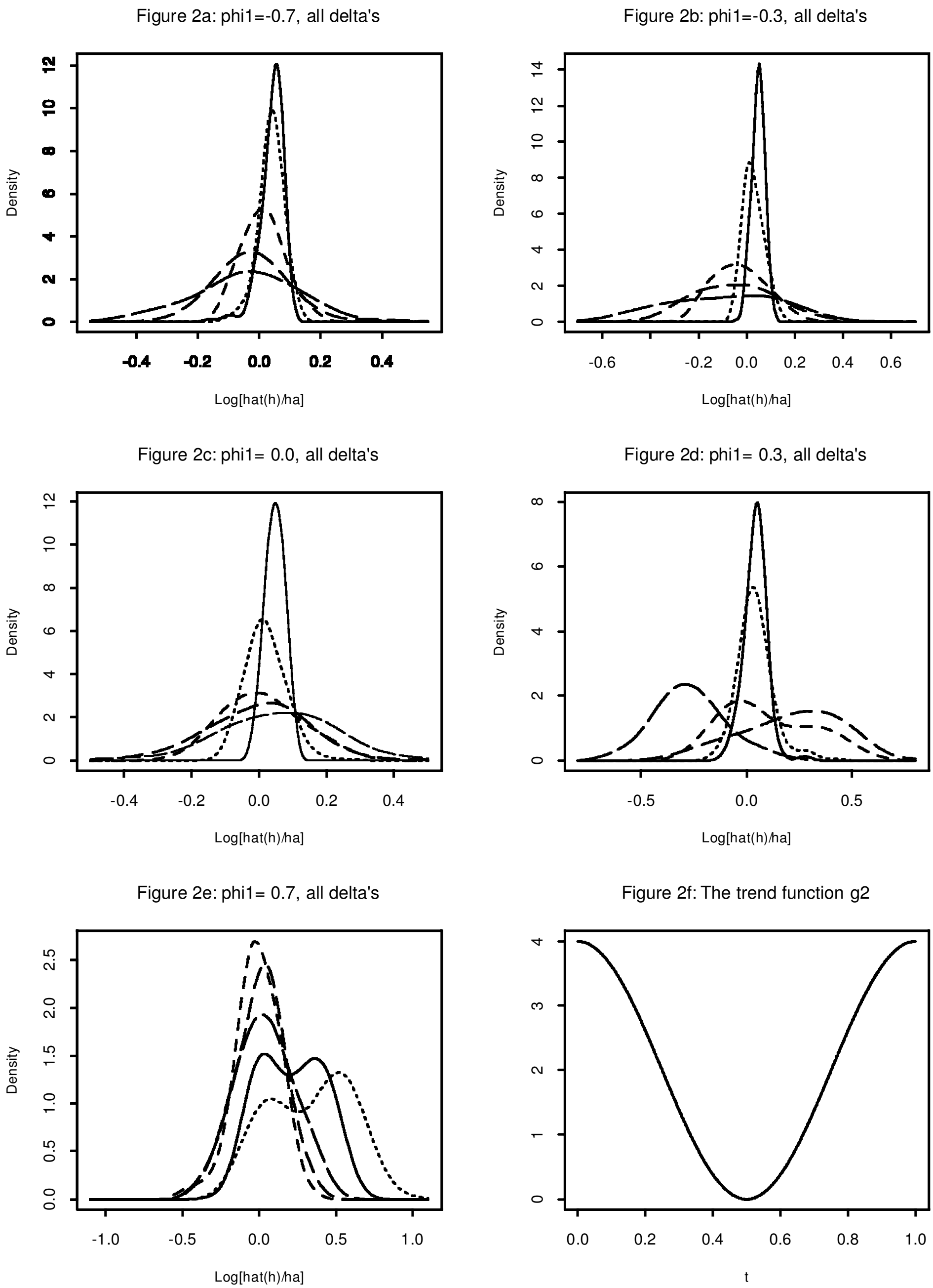


*Figure 2.* *The same results as given in Figure 1 but for the trend function $g_2$.*

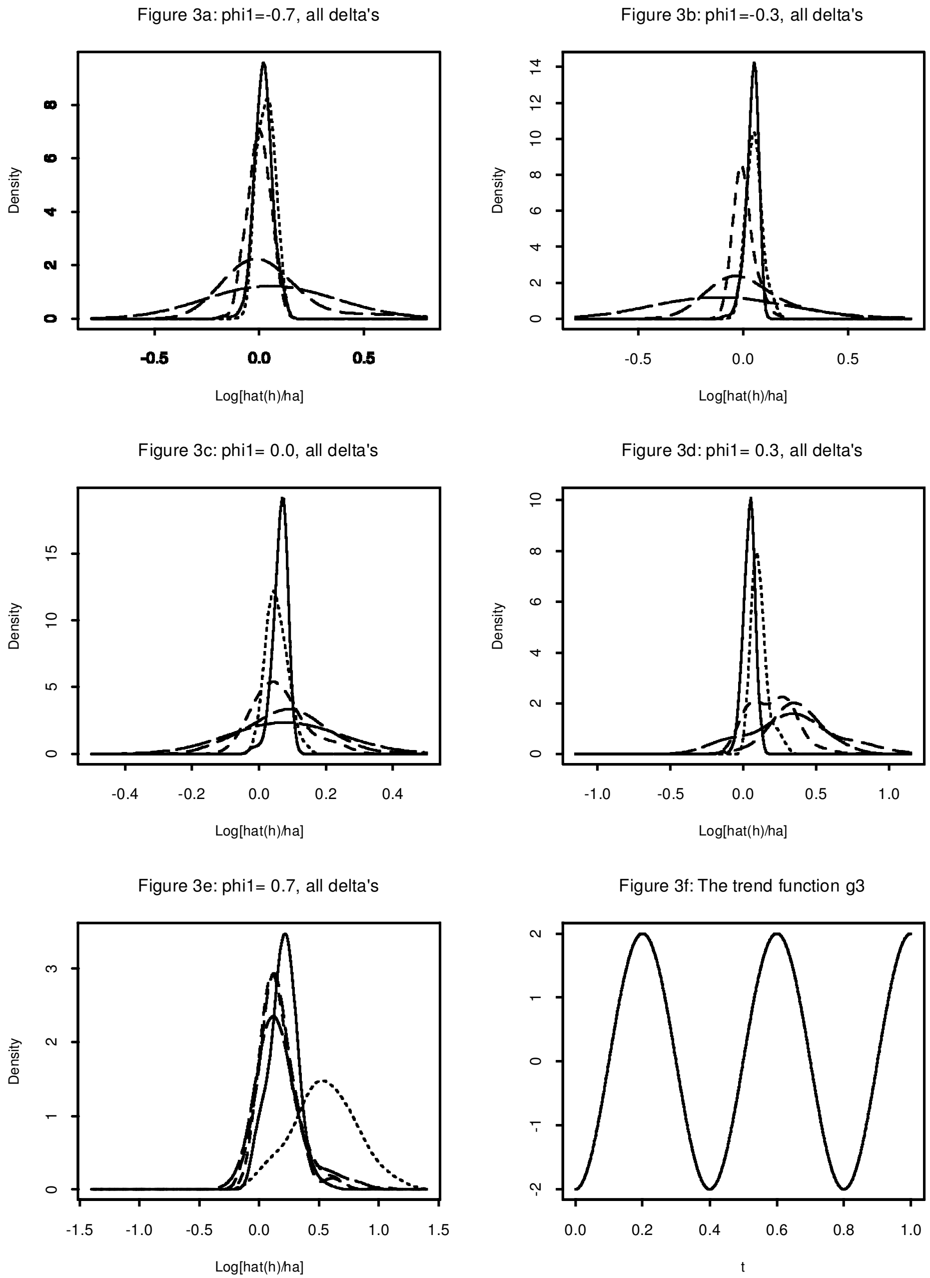


*Figure 3.* *The same results as given in Figure 1 but for the trend function $g_3$.*

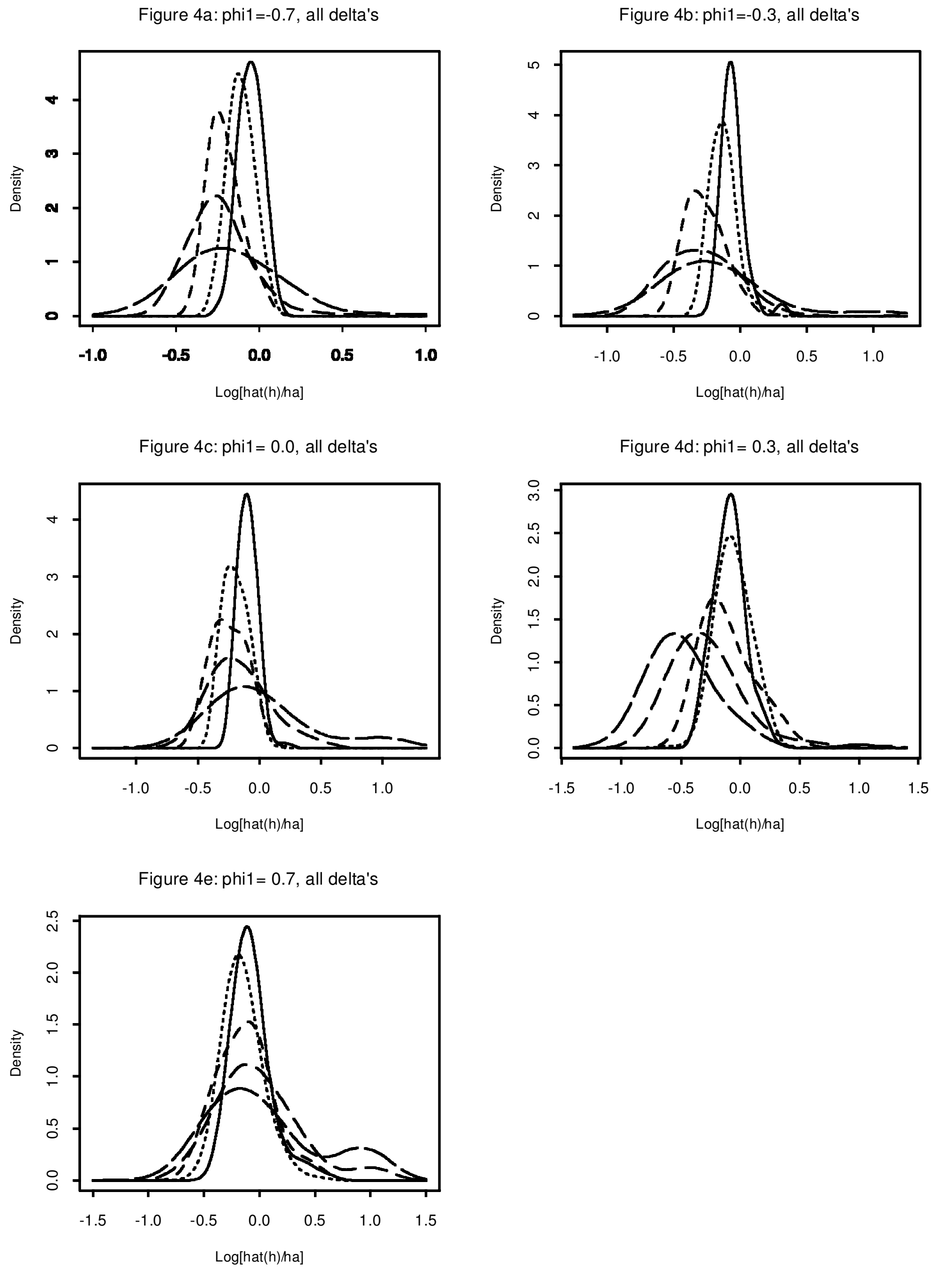


*Figure 4. The same results as given in Figures 1a through 1e but for $m^0 = 1$.*

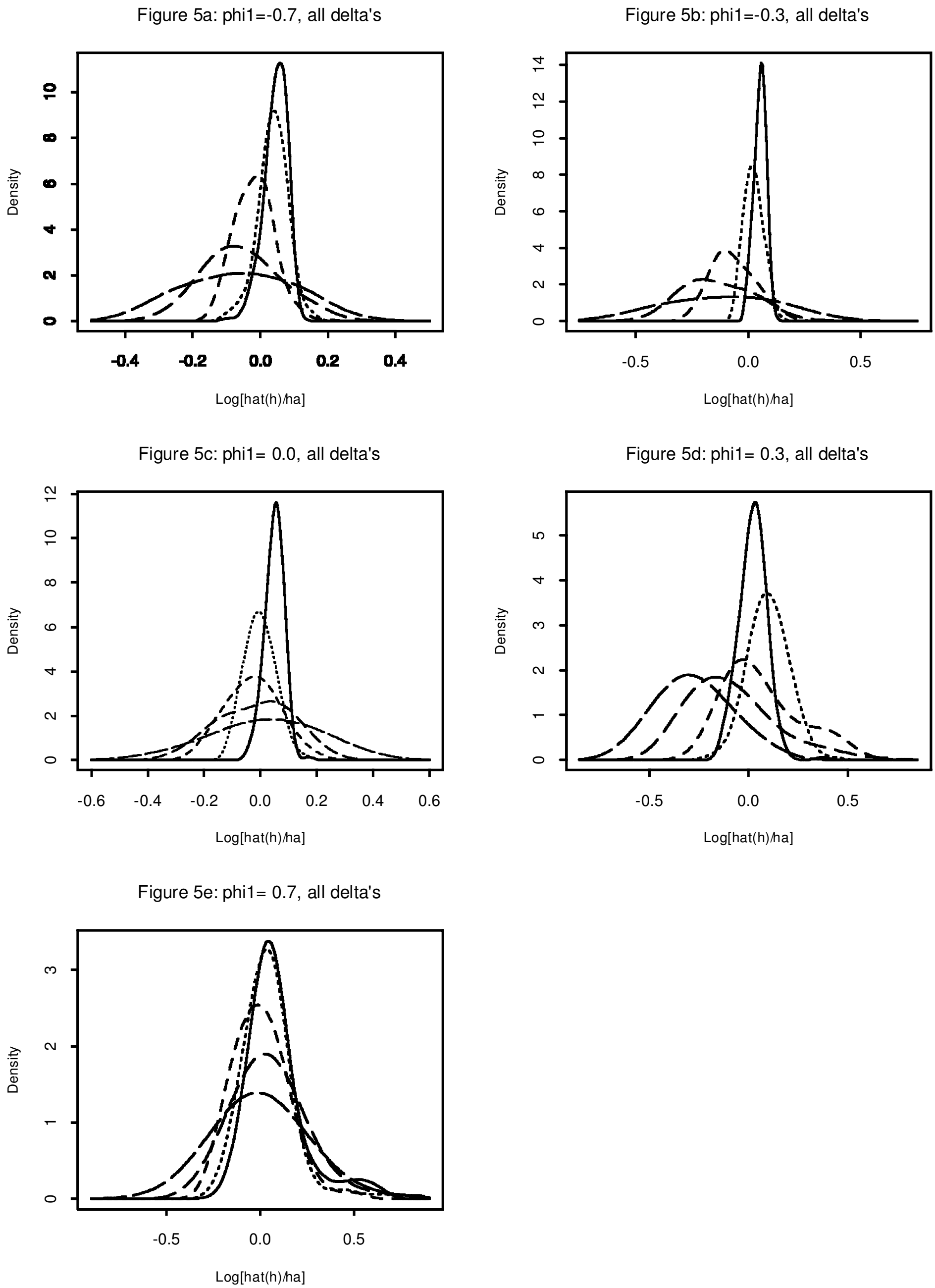


*Figure 5.* *The same results as given in Figures 2a through 2e but for $m^0 = 1$.*

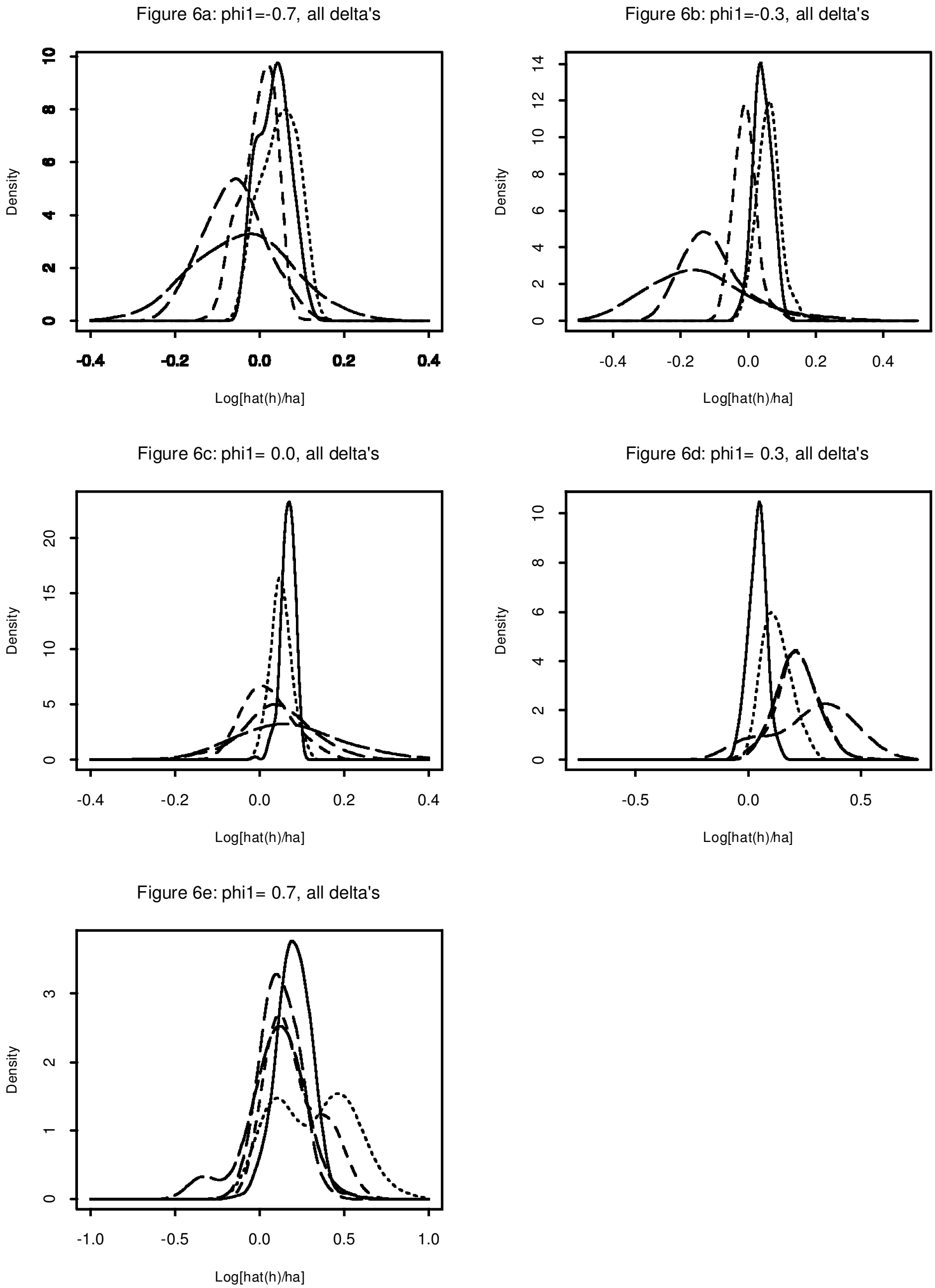


*Figure 6. The same results as given in Figures 3a through 3e but for $m^0 = 1$.*

the second term in the variance part is asymptotically negligible in the neighborhood of $h_{\mathrm{M}}$. Hence we have

$$M(h_{\mathrm{M}}) - M_2(h_{\mathrm{M}}) \doteq M_4(h_{\mathrm{M}}) - M_2(h_{\mathrm{M}}) \doteq O(h^6).$$

Following the idea in Härdle et al. (1992) and Feng (1999), it can be shown that

$$\begin{aligned} h_{\mathrm{A}} - h_{\mathrm{M}} &\doteq -[M_2'(h_{\mathrm{M}}) - M'(h_{\mathrm{M}})]/M''(h_{\mathrm{M}}) \\ &\doteq -[M_2'(h_{\mathrm{M}}) - M_4'(h_{\mathrm{M}})]/M''(h_{\mathrm{M}}). \end{aligned}$$

Note that $M''(h_{\mathrm{M}}) \doteq O(h_{\mathrm{M}}^2)$, we have $h_{\mathrm{A}} - h_{\mathrm{M}} \doteq O(h_{\mathrm{M}}^3)$ and $(h_{\mathrm{A}} - h_{\mathrm{M}})/h_{\mathrm{M}} \doteq O(h_{\mathrm{M}}^2)$. □

***Proof of Lemma 1:*** In the following proof the unknown parameters $\delta$ and $V$ instead of their estimates will be used. The error caused by doing this is quantified by the $O_p(n^{-1/2})$ term given in (3.3)–(3.5). By the iterative plug-in algorithm a bandwidth of order $O_p(h_{\mathrm{M}})$ plays the key role. Note that $\hat{g}''$ (and hence $\hat{I}$) using a bandwidth of order $O_p(h_{\mathrm{M}})$ will be of order $O_p(1)$ but not consistent (see (1.5)). Assuming that $h_0 = o(h_{\mathrm{M}})$ (e.g., $O(n^{-5/7})$) is used at the beginning as proposed in this article, an iterative plug-in bandwidth selection procedure may be divided into the following three steps according to the relationship between $h_{2,j}$ and $h_{\mathrm{M}}$.

Step 1. When $h_{2,j} = o_p(h_{\mathrm{M}})$, the bias of $\hat{I}$ is of the order $O_p(n^{\delta-1/2}h_{2,j}^{\delta-5/2}) \to \infty$ as $n \to \infty$ (see (1.5)). In this case one obtains $h_j = O_p(h_{2,j})$ (see (3.1)). This implies that $h_{j-1} = o(h_j)$, i.e. the bandwidth is inflated of a larger order. This step will be iteratively carried out till Step 2 or Step 3 is reached.

Step 2. When $h_{2,j} = O_p(h_{\mathrm{M}})$, $\hat{I} = O_p(1)$. Now, $h_j = O_p(h_{\mathrm{M}})$. But, in general, the constant is not consistent. Now we also have $h_{j-1} = o(h_j)$. Step 2 consists of only one iteration, if it occurs.

Step 3. When $h_{\mathrm{M}} = o_p(h_{2,j})$, $\hat{I}$ is consistent and then $h_j$ will be a consistent estimate of $h_{\mathrm{A}}$ (and $h_{\mathrm{M}}$).

After Step 3 we have $h_{2,j}^{\mathrm{var}} = O_p(n^{(2\delta-1)/(10-4\delta)}) = O_p(h_2^{\mathrm{var}})$, $h_{2,j}^0 = O_p(n^{(2\delta-1)/(9-2\delta)}) = O_p(h_2^0)$ and $h_{2,j}^{\mathrm{opt}} = O_p(n^{(2\delta-1)/(7-2\delta)}) = O_p(h_2^{\mathrm{opt}})$ for $\alpha = \alpha_{\mathrm{var}}$, $\alpha_0$ and $\alpha_{\mathrm{opt}}$, respectively. The corresponding rates of convergence in these three cases may be calculated as follows: Taylor expansion gives

$$\hat{h} - h_{\mathrm{A}} \doteq -\frac{1}{5-2\delta} h_{\mathrm{A}} I^{-1}(\hat{I} - I).$$

It follows that

$$(\hat{h} - h_{\mathrm{A}})/h_{\mathrm{A}} \doteq -\frac{1}{5-2\delta} I^{-1}(\hat{I} - I).$$

It is obvious that $(\hat{h} - h_{\mathrm{M}})/h_{\mathrm{M}} \doteq (\hat{h} - h_{\mathrm{M}})/h_{\mathrm{A}}$. Observe that $(h_{\mathrm{A}} - h_{\mathrm{M}})/h_{\mathrm{M}} = o_p(\hat{I} - I)$ we have $(\hat{h} - h_{\mathrm{M}})/h_{\mathrm{A}} \doteq (\hat{h} - h_{\mathrm{A}})/h_{\mathrm{A}} \doteq -\frac{1}{5-2\delta} I^{-1}(\hat{I} - I)$. Inserting $h_2^{\mathrm{var}}$, $h_2^0$ and $h_2^{\mathrm{opt}}$

into the right-hand side of (2.6), and observing that the second term is negligible for Case 2 and the two terms are of the same order for Case 3, we obtain the rates of convergence in the three cases, respectively. □

***Proof of Lemma 2:*** In the following we will call $\alpha_\delta$ a stable order. Define $\delta_1 = \max\{2\delta^0 - 1/2, -0.5\}$. It is clear that $\delta_1 < \delta^0$. For a trial value $\delta_1 < \delta < 0.5$, we have $h_{2,1} = h_0^{(5-2\delta)/(9-2\delta)} = O(n^{-\tilde{\alpha}})$ with $0 < \tilde{\alpha} < (1-2\delta^0)/(5-2\delta^0)$. This means that $\hat{I}(h_{2,1})$ is consistent and $h_1 = O(h_0)$. For $j \geq 2$ we have $h_j = h_{j-1}(1 + o_p(1))$. In this case $\hat{h}(\delta) = O_p(n^{-\alpha_\delta})$ with the stable order $\alpha_\delta = (1-2\delta)/(5-2\delta)$ and a convergent constant part, whose limit depending on $\delta$.

The case $\delta \leq \delta_1$ can only occur if $\delta_1 > -0.5$ (i.e., $\delta^0 > 0$). Thus, suppose that $\delta_1 > -0.5$. In the case $\delta = \delta_1$ we also have $\alpha_\delta = (1-2\delta)/(5-2\delta)$. But now, $\hat{I} = O_p(1)$ is not consistent. This results in $\hat{h}(\delta) = O_p(n^{-\alpha_\delta})$ with an divergent constant part.

For $-0.5 < \delta < \delta_1$ it can be shown that the stable order is $\alpha_\delta = 2(\delta^0 - \delta)(9 - 2\delta)/\{(5-2\delta)(4 + 2(\delta^0 - \delta))\}$. Now the constant part is also divergent. In this case $\alpha_\delta > (1-2\delta)/(5-2\delta)$, i.e. the stable bandwidth is now of a smaller order than $n^{(2\delta-1)/(5-2\delta)}$. Now, $\alpha_\delta$ is monotonically increasing in $\delta^0$ and monotonically decreasing in $\delta$ with the upper bound $\frac{5}{9}$. □

***A sketched proof of Theorem 1:***

*Part 1:* To show the asymptotic normality of $\sqrt{n}(\hat{\theta} - \theta^o)$ we have at first to show that $\hat{h}(\delta)$, the bandwidth selected at the end of AlgA for each trial value, satisfies

a) for $m = m^0$, $\hat{h}(\delta) \to 0$, $n\hat{h}(\delta) \to \infty$, and

b) for $m \neq m^0$, $n\hat{h}(\delta) \to \infty$

in probability as $n \to \infty$. For $m \neq m^0$, the condition $h(\delta) \to 0$ in probability as $n \to \infty$ is unnecessary, although it can be shown that it holds. The validity of a) and b) can be seen as follows:

Condition a) follows directly from Lemma 2.

For condition b) we have two cases: In the case $m^0 = 1$ with $m = 0$ we have $\hat{I} = O_p(n^2)$ and hence, for each $j$, $h_j \geq O_p(n^{-2/(5-2\delta)}n^{(2\delta-1)/(5-2\delta)}) = O_p(n^{(2\delta-3)/(5-2\delta)})$. In the case $m^0 = 0$ with $m = 1$, $\hat{I}$ will be asymptotically dominated by the second term on the right-hand side of (2.6), which is of order $O[(nh_{2,j})^{2\delta-1}h_{2,j}^{-4}]$. Hence we have $h_j = O_p(h_{2,j})$ in any iteration. In both cases we have $n\hat{h}(\delta) \to \infty$.

Further proof of part i) follows from the proof of Theorem 2 in Beran (1999).

*Part 2:* Since $\hat{m}$ is consistent, we only need to consider the rate of convergence of $\hat{h}$ in the case when $\hat{m} = m^0$. Now, the rate of convergence of $\hat{h}$ by AlgA follows from Case ii) in Lemma 1. □

## ACKNOWLEDGMENTS

This article was supported in part by the Center of Finance and Econometrics at the University of Konstanz, Germany, and by an NSF (SBIR, phase 2) grant to MathSoft, Inc.

*[Received May 2000. Revised 2001.]*

# REFERENCES

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

——— (1995), "Maximum Likelihood of Estimation of the Differencing Parameter for Invertible Short- and Long-Memory Autoregressive Integrated Moving Average Models," *Journal of the Royal Statistical Society*, Ser. B, 57, 659–672.

——— (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., Bhansali, R. J., and Ocker, D. (1998), "On Unified Model Selection for Stationary and Nonstationary Short- and Long-Memory Autoregressive Processes," *Biometrika*, 85, 921–934.

Beran, J., and Feng, Y. (2000), "Supplement to `Data-Driven Estimation of Semiparametric Fractional Autoregressive Models'—Detailed Simulation Results," Preprint, University of Konstanz.

——— (2002), "Locally Polynomial Fitting With Long-Memory, Short-Memory and Antipersistent Errors," *Annals of the Institute of Statistical Mathematics*, 54, 291–311.

Beran, J., and Ocker, D. (1999), "SEMIFAR Forecasts, With Applications to Foreign Exchange Rates", *Journal of Statistical Planning and Inference*, 80, 137–153.

——— (2001) "Volatility of Stock Market Indices—An Analysis Based on SEMIFAR Models," *Journal of Business and Economic Statistics*, 19, 103–116.

Fan, J., and Gijbels, I. (1995), "Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation," *Journal of the Royal Statistical Society*, Ser. B, 57, 371–394.

Feng, Y. (1999), *Kernel and Locally Weighted Regression With Application to Time Series Decomposition*, Berlin: Verlag.

Gasser, T., Kneip, A., and Köhler, W. (1991), "A Flexible and Fast Method for Automatic Smoothing," *Journal of the American Statistical Association*, 86, 643–652.

Gasser, T., and Müller, H. G. (1984), "Estimating Regression Functions and Their Derivatives by the Kernel Method," *Scandinavian Journal of Statistics*, 11, 171–185.

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.

Härdle, W., Hall, P., and Marron, J. S. (1992), "Regression Smoothing Parameters That Are Not Far From Their Optimum," *Journal of the American Statistical Association*, 87, 227–233.

Hall, P., and Hart, J. D. (1990), "Nonparametric Regression with Long-Range Dependence," *Stochastic Processes and Applications*, 36, 339–351.

Heiler, S., and Feng, Y. (1998), "A Root $n$ Bandwidth Selector for Nonparametric Regression," *Journal of Nonparametric Statistics*, 9, 1–21.

Herrmann, E., and Gasser, T. (1994), "Iterative Plug-In Algorithm for Bandwidth Selection in Kernel Regression Estimation," preprint, Darmstadt Institute of Technology and University of Zürich.

Herrmann, E., Gasser, T., and Kneip, A. (1992), "Choice of Bandwidth for Kernel Regression When Residuals are Correlated," *Biometrika*, 79, 783–795.

Hosking, J. R. M. (1981), "Fractional Differencing," *Biometrika*, 68, 165–176.

Müller, H. G. (1985), "Empirical Bandwidth Choice for Nonparametric Kernel Regression by Means of Pilot Estimators," *Statistical Decisions*, Supplemental Issue, 2, 193–206.

Ray, B. K., and Tsay, R. S. (1997), "Bandwidth Selection for Kernel Regression With Long-Range Dependence," *Biometrika*, 84, 791–802.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, 90, 1257–1270.