Data-driven estimation of semiparametric fractional autoregressive models

Jan Beran and Yuanhua Feng
University of Konstanz

Abstract

In this paper data-driven algorithms for fitting SEMIFAR models (Beran, 1999) are proposed. The algorithms combine the data-driven estimation of the nonparametric trend and maximum likelihood estimation of the parameters. For selecting the bandwidth, the proposal of Beran and Feng (1999) based on the iterative plug-in idea (Gasser et al., 1991) is used. Asymptotic properties of the proposed algorithms are investigated. A large simulation study illustrates the practical performance of the methods.

Key Words: semiparametric models, long-range dependence, fractional ARIMA, antipersistence, nonparametric regression, bandwidth selection.

1 Introduction

The so-called SEMIFAR (semiparametric fractional autoregressive) model, introduced by Beran (1999), provides a unified approach that allows for simultaneous modelling of deterministic trends, stochastic trends and stationary short-memory, long-memory and antipersistent components. Beran (1999) and Beran and Ocker (1999a) investigate the basic properties of this model. The usefulness of SEMIFAR models in practice, especially for analyzing financial time series, is shown in Beran and Ocker (1999a, b). Estimation of the SEMIFAR model requires a data-driven algorithm. Such an algorithm was originally proposed in Beran (1999) and Beran and Ocker (1999a). Beran and Feng (1999) propose a general bandwidth selector for nonparametric regression with short-memory, long-memory and antipersistence.

In this paper, several data-driven algorithms for estimating the SEMIFAR model are proposed using the bandwidth selector in Beran and Feng (1999). Asymptotic
properties of the methods are investigated. 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

\[
\phi(B)(1 - B)^{\delta} \{(1 - B)^m Y_i - g(t_i)\} = \epsilon_i,
\]

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 \( \text{var}(\epsilon_i) = \sigma^2 \). 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

\[
\beta_k(\delta) = (-1)^k \frac{\Gamma(\delta + 1)}{\Gamma(k + 1)\Gamma(\delta - k + 1)}.
\]

Model (1) allows us to analyze stationary \((m = 0)\) or difference-stationary \((m = 1)\) processes with or without deterministic trends, as well as with short-range dependence \((\delta = 0)\), long-range dependence \((\delta > 0)\) and antipersistence \((\delta < 0)\). See Beran (1999) and Beran and Ocker (1999a, b) for detailed remarks on different special cases of model (1).

The paper is organized as follows. Section 2 summarizes the basic estimation methods. Bandwidth selection for estimating \( \hat{g} \) is discussed in section 3. Section 4 proposes the data-driven algorithms for fitting SEMIFAR models and investigates their asymptotic properties. Results of the simulation study are summarized in section 5. Detailed results of this simulation may be found in a discussion paper (Beran and Feng, 2000) as a supplement of the current paper. Section 6 contains some final remarks. Proofs of the results are listed in the appendix.
2 Estimation of the SEMIFAR methods

The estimation of SEMIFAR models consists of two parts: nonparametric estimation of the trend $g$ and estimation of the parameters $m$, $\delta$, $p$ and $\phi_1, \ldots, \phi_p$. In this paper the trend $g$ will be estimated by a kernel method (Hall and Hart, 1990 and Beran, 1999). The parameters will be estimated based on the approximate maximum likelihood approach proposed by Beran (1995).

2.1 Estimation of the trend

Under definition (1) either $Y_i$ ($m = 0$) or the first difference $BY_i = Y_i - Y_{i-1}$ ($m = 1$) is a nonparametric regression model with errors having quite different dependent structures. Denote by $U_i = Y_i$ for $m = 0$ or $U_i = Y_i - Y_{i-1}$ for $m = 1$ (in this case define $U_1 := 0$), and define $X_i = U_i - g(t_i)$. Then we have

$$U_i = g(t_i) + X_i,$$ \hspace{1cm} (4)

where $X_i$ is a stationary fractional autoregressive process. Equation (4) is a nonparametric regression model with a time series error process whose long-term dependence structure depends on the value of $\delta$. The spectral density of $X_i$ in (4) has the form

$$f(\lambda) \sim c_f |\lambda|^{-\alpha} \quad (\text{as } \lambda \to 0)$$ \hspace{1cm} (5)

with $\alpha = 2\delta$, where $c_f$ is the value of the spectral density of the 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. This is sometimes 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 paper are valid for the whole range $\delta \in (-0.5, 0.5)$.
The kernel estimator as proposed by Hall and Hart (1990) and Beran (1999) will be used to estimate the trend $g$. Assume that $m = 0$, then for a given bandwidth $h > 0$ and a second order kernel function $K$, the kernel estimator of $g$ is defined by

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

A similar estimator can be defined for $m = 1$ replacing $Y_i$ by $U_i = Y_i - Y_{i-1}$.

Asymptotic properties of $\hat{g}$ are discussed by Beran (1999). Results for $\delta \geq 0$ may also be found in 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. Define

$$
I(g'') = \int_{-\Delta}^{1-\Delta} [g''(t)]^2 dt
$$

and

$$
I(K) = \int_{-1}^{1} x^2 K(x) dx.
$$

Under the assumptions of Theorem 1 in Beran (1999) we have the following asymptotic formulas for the bias, variance and mean integrated squared error (MISE) of $\hat{g}$.

(i) **Bias:**

$$
E[\hat{g}(t) - g(t)] = h^2 \frac{g''(t)I(K)}{2} + o(h^2)
$$

uniformly in $\Delta < t < 1 - \Delta$;

(ii) **Variance:**

$$
\text{var}(\hat{g}(t)) = \frac{1}{(nh)^{1-2\delta}}[V + o(1)]
$$

uniformly in $\Delta < t < 1 - \Delta$, where $V$ is a constant depending on $c_f$ and the kernel function;

(iii) **MISE:** The mean integrated squared error in $[\Delta, 1 - \Delta]$ is given by

$$
E \left\{ \int_{-\Delta}^{1-\Delta} [\hat{g}(t) - g(t)]^2 dt \right\} = h^4 \frac{I(g'')I^2(K)}{4} + (nh)^{2\delta-1} V(1 - 2\Delta)
$$

$$
+ o(\max(h^4, (nh)^{2\delta-1})).
$$

Formulas for $V$ (with $\delta \in (-0.5, 0.5)$) may be found in Beran and Feng (1999).
2.2 Estimation of the parameters

The parameters of the SEMIFAR models, including \( m \) and \( \delta \), may be estimated by maximum likelihood (Beran, 1995, 1999). Note that, since \( m \) is an integer, \( m \) and \( \delta \) correspond to one parameter \( d = m + \delta \) only, through \( m = \lfloor d + 0.5 \rfloor \) and \( \delta = d - m \), where \( \lfloor \cdot \rfloor \) denotes the integer part. Let \( \theta^0 = (\sigma^2_{\epsilon,0}, d^0, \phi_1^0, \ldots, \phi_p^0)^T = (\sigma^2_{\epsilon,0}, \eta^0)^T \) be the true unknown parameter vector in (1) where \( d^0 = m^0 + \delta^0 \), \( -0.5 < \delta^0 < 0.5 \) and \( m^0 \in \{0, 1\} \). For a constant trend function \( g = \mu \), maximum likelihood estimation of \( \theta^0 \), based on the autoregressive representation of the process, is considered in Beran (1995). Beran (1999) extended this idea to estimate \( \theta^0 \) in the SEMIFAR model with a general nonparametric trend function \( g \). Note that

\[
\phi(B)(1 - B)^{d^0} \{ (1 - B)^m Y_i - g(t_i) \} = \sum_{j=0}^{\infty} a_j(\eta^0) B^j \{ c_j(\eta^0) Y_i - g(t_i) \} = \sum_{j=0}^{\infty} a_j(\eta^0) \{ c_j(\eta^0) Y_{i-j} - g(t_{i-j}) \},
\]

where the coefficients \( a_j \) and \( a_j c_j \) are obtained by matching the powers of \( B \). Hence, \( Y_i \) admits an infinite autoregressive representation

\[
\sum_{j=0}^{\infty} a_j(\eta^0) \{ c_j(\eta^0) Y_{i-j} - g(t_{i-j}) \} = \epsilon_i.
\] (12)

Let \( h \) be a bandwidth such that \( h \to 0 \) and \( nh \to \infty \) as \( n \to \infty \), and let \( \hat{g}(t_i) = \hat{g}(t_i; m) \) be the estimated trend function obtained from (4). Consider now \( \epsilon_i \) as a function of \( \eta \). For a chosen value of \( \theta = (\sigma^2_{\epsilon}, m + \delta, \phi_1, \ldots, \phi_p)^T = (\sigma^2_{\epsilon}, \eta)^T \), denote by

\[
e_i(\eta) = \sum_{j=0}^{i-m-2} a_j(\eta) \{ c_j(\eta) Y_{i-j} - \hat{g}(t_{i-j}; m) \}
\] (13)

the (approximate) residuals and by \( r_i(\theta) = \epsilon_i(\eta)/\sqrt{\tau_i} \) the standardized residuals. Assuming that \( \{\epsilon_i(\eta^0)\} \) are independent zero mean normal with variance \( \sigma^2_{\epsilon} \), an approximate maximum likelihood estimate of \( \theta^0 \) is obtained by maximizing the approximate log-likelihood

\[
l(Y_i, \ldots, Y_n; \theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2_{\epsilon} - \frac{1}{2} n^{-1} \sum_{i=m+2}^{n} r_i^2
\] (14)

with respect to \( \theta \) and hence by solving the equations

\[
\hat{l}(Y_i, \ldots, Y_n; \theta) = 0,
\] (15)
where \( \hat{l} \) is the vector of partial derivatives with respect to \( \theta_j \) \((j = 1, ..., p + 2)\). More explicitly, \( \hat{\eta} \) is obtained by minimizing

\[
S_n(\eta) = \frac{1}{n} \sum_{i=m+2}^{n} e_i^2(\eta)
\]

with respect to \( \eta \) and setting

\[
\hat{\sigma}_\varepsilon^2 = \frac{1}{n} \sum_{i=m+2}^{n} e_i^2(\hat{\eta}).
\]

For the case where \( g \) is known to be constant, it follows from Beran (1995) that, if the constant \( g = \mu \) is estimated consistently, then (as \( n \to \infty \)) \( \hat{\theta} \) converges in probability to \( \theta^0 \), and \( \sqrt{n}(\hat{\theta} - \theta^0) \) converges in distribution to a normal random variable with zero mean vector and covariance matrix equal to the inverse Fisher-Information matrix. Here, both, the fractional differencing parameter \( \delta \) and the integer differencing parameter \( m \) are estimated from the data. Also, the asymptotic covariance matrix does not depend on \( m \). This result also holds for SEMIFAR models. If \( g \) is estimated consistently, then \( \sqrt{n}(\hat{\theta} - \theta^0) \) converges in distribution to a normal random variable with zero mean vector and covariance matrix

\[
\Sigma = 2D^{-1},
\]

where

\[
D_{ij} = (2\pi)^{-1} \left\{ \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f(x) \frac{\partial}{\partial \theta_j} \log f(x) dx \right\} |_{\theta = \theta_\varepsilon^{0}}
\]

with \( \theta_\varepsilon^{0} = (\sigma_{\varepsilon,0}^2, \theta^0, \phi_1^0, ..., \phi_p^0)^T \) (see Theorem 2 in Beran, 1999). This result can be extended to the case where the innovations \( \varepsilon_i \) are not normal and satisfy suitable moment conditions.

These results are given under the assumption that the order \( p = p_0 \) of the autoregressive polynomial in (1) is known. This cannot be assumed in practice. Thus, \( p_0 \) should be selected by applying a suitable model choice criterion. In this paper \( p_0 \) will be selected by BIC (Bayesian information criterion) (Schwarz, 1978, Akaike, 1979). Consistency properties of the BIC were shown in Beran et al. (1998) for FARIMA (fractional autoregressive integrated moving average) models without trend. For an extension to SEMIFAR models see Beran (1999). Note that in Algorithms B and C described in section 4, \( m^0 \) will also be selected by BIC to reduce computing time.
3 Bandwidth selection

Data-driven bandwidth selection is a crucial problem in the practical use of non-parametric regression. Recent proposals for bandwidth selection in nonparametric regression with independent or short-range dependent data may be found e.g. in Müller (1985), Gasser et al. (1991), Härdle et al. (1992), Herrmann et al. (1992), Fan and Gijbels (1995), Ruppert et al. (1995) and Heiler and Feng (1998).

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). Beran (1999), Beran and Ocker (1999a) and Beran and Feng (1999) proposed a bandwidth selector for data with several dependence structures (long-memory, short-memory and antipersistence) using a variant of the iterative plug-in approach. A special case of the proposal in Beran and Feng (1999) with \( k = 2 \) and \( l = 4 \) will be discussed here in detail.

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

\[
h_{A} = C \cdot n^{(2\delta - 1)/(5 - 2\delta)}
\]

with

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

Here it is assumed that \( I(g'') > 0 \). When the uniform kernel is used, the constant \( C \) in (20) has the explicit form

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

with \( c_f \) as defined before and

\[
\nu(\delta) = \frac{2^{2\delta} \Gamma(1 - 2\delta) \sin(\pi \delta)}{\delta(2\delta + 1)}
\]

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

Plug-in estimators for \( h_{M} \) use formula (20), 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 \( c_f \). Following section 2.2, both, \( \delta \) and \( V \) may be estimated root
n consistently. Hence the key problem is to estimate \( I(g'') \). This will be discussed
in the following. Let \( \hat{g}''(t; h_2) \) be a kernel estimator for \( g'' \) with a kernel \( K_2 \) of order
4 (see e.g. Gasser and Müller, 1984) and a bandwidth \( h_2 \), which is different from
the bandwidth \( h \) for estimating \( g \). And let \( I(g'') \) be estimated as follows
\[
\hat{I}(g'') = n^{-1} \sum_{i = [n\Delta]}^{n-1} \{ \hat{g}''(t_i; h_2) \}^2. \tag{24}
\]

Properties of \( \hat{I}(g'') \) are investigated by Beran and Feng (1999). Under the assump-
tion of Proposition 1 in Beran and Feng (1999) we have
\[
E[\hat{I}(g'') - I(g'')] \doteq h_2 \frac{I(K_2)}{12} \int_{\Delta}^{\Delta} g''(t)g^{(4)}(t)dt + (nh_2)^{2\delta-1}h_2^{-4}V \tag{25}
\]
and
\[
\text{var}[\hat{I}(g'')] \doteq o[(nh_2)^{(4\delta-2)}h_2^{-8}] + O(n^{2\delta-1}). \tag{26}
\]

The mean squared error (MSE) of \( \hat{I}(g'') \) is dominated by the squared bias
\[
\text{MSE}\{\hat{I}(g'')\} \doteq \left( h_2 \frac{I(K_2)}{12} \int_{\Delta}^{\Delta} g''(t)g^{(4)}(t)dt + (nh_2)^{2\delta-1}h_2^{-4}V \right)^2.
\]
The optimal bandwidth for estimating \( I(g'') \) which minimizes the MSE is \( h_2^o = O(n^{(2\delta-1)/(\delta-2)}) \).

Following the iterative plug-in idea of Gasser et al. (1991), in the \( j \)th iteration,
\( I(g'') \) 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 and Beran and Ocker, 1999a). 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 \( h_0 \rightarrow 0, nh_0 \rightarrow \infty \) as \( n \rightarrow \infty \), 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. In this paper we propose 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 $h_0 = o(h_A)$ for all $\delta \in (-0.5, 0.5)$. Here we used $h_0 = n^{-5/7}$, which is of order $o(h_A^2)$ for all $\delta \in (-0.5, 0.5)$.

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 (1999a) propose to use 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 (see Lemma 1 in the appendix). So $\alpha$ should be chosen in a way that $c \cdot h_A n^\alpha = h_2^\alpha$ by the MIM, or $c \cdot (h_A)^\alpha = h_2^\alpha$ by the EIM, respectively. The optimal choice for the MIM is $\alpha = (2 - 4\delta)/[(5 - 2\delta)(7 - 2\delta)]$ (see Herrmann and Gasser, 1994 for the case with $\delta = 0$). For the EIM $\alpha_0 = (5 - 2\delta)/(7 - 2\delta)$ should be used. 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 the naive one $\alpha_n$ that optimizes $\mathcal{g}''$ itself and the variance optimal one $\alpha_v$ for which the square of second term in (25) is of the order $O(n^{2\delta - 1})$. The required bandwidths to estimate $\mathcal{g}''$ in these two cases are $h_2^\alpha = O(n^{(2\delta - 1)/(9 - 2\delta)})$ and $h_2^\alpha = O(n^{(2\delta - 1)/(2(5 - 2\delta))})$, respectively. For the MIM we have $\alpha_n = (4 - 8\delta)/[(5 - 2\delta)(9 - 2\delta)]$ and $\alpha_v = (1 - 2\delta)/(10 - 4\delta)$. They are $\alpha_n = (5 - 2\delta)/(9 - 2\delta)$ and $\alpha_v = \frac{1}{2}$ for the EIM. The rate of convergence of $\hat{h}$ with $\alpha_n$ lies between the two with $\alpha_0$ and $\alpha_v$. Ray and Tsay (1997) used the MIM with $\alpha_v$, while the EIM with $\alpha_n$ was used by Beran (1999) and Beran and Ocker (1999a) (see Algorithm A in the next section).

Denote by $j^0$ the number of iterations required for obtaining a satisfactory bandwidth selector. $j^0$ can be calculated following the idea in Gasser et al. (1991) and Herrmann and Gasser (1994), if $h_0$, the inflation method and $\alpha$ are given. See Beran and Feng (1999) for detailed discussion. We propose the following bandwidth selector for the kernel estimator $\hat{g}$ with independent data, long-memory data or antipersistent data. Here it is assumed that $m = 0$. 

9
i) Start with the bandwidth \( h_0 = n^{-\beta} \) with \( \frac{1}{2} \leq \beta < 1 \) and set \( j = 1 \).

ii) Estimate \( g \) using \( h_{j-1} \) and let \( \tilde{X}_i = Y_i - \hat{g}(t_i) \). Estimate \( \delta \) and \( V \) from \( \tilde{X}_i \) with the method proposed in section 2.2.

iii) Set \( h_{2j} = (h_{j-1})^\alpha \) with \( \frac{1}{2} \leq \alpha < 1 \) and improve \( h_{j-1} \) by

\[
h_j = \left( \frac{1 - 2\hat{\delta}}{\beta^2} \frac{(1 - 2\Delta)\hat{V}}{I(g^\#(t; h_{2j}))} \right)^{1/(5-2\delta)} \cdot n^{(2\delta-1)/(5-2\delta)}.
\]

(27)

vi) Increase \( j \) by 1 and repeat steps ii) and iii) until convergence is reached or until a given number of iterations has been done.

The rate of convergence of \( \hat{h} \) depends on the inflation method (and \( \alpha \)). It also depends on the difference between \( h_A \) and \( h_M \). Results on the latter may be found e.g. in Gasser et al. (1991), Herrmann and Gasser (1994) and Ray and Tsay (1997). In this paper we will simply assume that \( h_A - h_M = o_p(\hat{I}(g^\#) - I(g^\#)) \), i.e. the difference between \( h_A \) and \( h_M \) is negligible. (For iid data, it can be shown that this relationship holds for kernel estimator, if \( g \) is at least fourth continuously differentiable.) Under this condition and conditions as given in Proposition 1 in Beran and Feng (1999), we have

i) For \( \alpha = \alpha_v = \frac{1}{2} \)

\[
\hat{h} = h_M \left\{ 1 + O(n^{(2\delta-1)/(5-2\delta)}) + O_p(n^{(2\delta-1)/2}) + O_p(n^{-1/2}) \right\}.
\]

(28)

ii) For \( \alpha = \alpha_n = (5 - 2\delta)/(9 - 2\delta) \)

\[
\hat{h} = h_M \left\{ 1 + O_p(n^{2(2\delta-1)/(9-2\delta)}) \right\}.
\]

(29)

iii) For \( \alpha = \alpha_o = (5 - 2\delta)/(7 - 2\delta) \)

\[
\hat{h} = h_M \left\{ 1 + O_p(n^{2(2\delta-1)/(7-2\delta)}) \right\}.
\]

(30)

Proof of these results will be omitted to save place. If \( \alpha = \alpha_o \) is used, then the rate of convergence of \( \hat{h} \) is \( n^{2(2\delta-1)/(7-2\delta)} \). It is \( n^{-2/7} \) for iid data and is the same as for the proposal in Ruppert et al. (1995).
This section deals with data-driven algorithms for estimating the SEMIFAR models. The symbols for the true unknown parameters as introduced in section 2.2 will be used. The original data-driven algorithm (Beran, 1999 and Beran and Ocker, 1999a) is an adaptation of Beran (1995) by replacing \( \hat{\mu} \) by the kernel estimator \( \hat{g} \). This algorithm 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. The original algorithm (with some minor changes) is defined as follows (see Beran and Ocker, 1999a):

**Algorithm A:**

Step 1: Define \( L = \) maximal order of \( \phi(B) \) that will be tried, and a sufficiently fine grid \( G \in (-0.5, 1.5) \setminus \{0.5\} \). Then, for each \( p \in \{0, 1, ..., L\} \), 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 \( \hat{\epsilon}_i(d) = \sum_{j=0}^{L} \beta_j(\delta)\hat{X}_{i-j} \), where the coefficients \( \beta_j \) are defined by (3).

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

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

\[
h_j = \left( \frac{1 - 2\delta}{\beta^2 I(g''(t; h_{2,j}))} \right)^{1/(5-2\delta)} \cdot n^{(2\delta-1)/(5-2\delta)}.
\] (31)
Step 3f: Increase $j$ by one and repeat steps 3b to 3e four times. This yields for each $d \in G$ separately, the ultimate value of $\hat{\sigma}^2_r(d)$, as a function of $d$.

Step 4: Define $\hat{d}$ to be the value of $d$ for which $\hat{\sigma}^2_r(d)$ is minimal. This together with the corresponding estimates of the AR parameters, yields an information criterion, e.g. $\text{BIC}(p) = n \log \hat{\sigma}^2_r(p) + p \log n$, 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 $g$.

Here $\Delta_0$ is used so that the starting bandwidth is not too large. We propose the use of $\Delta_0 = 2\Delta = 0.2$. This means that, at the first iteration, at most 20% observations are used for estimating $g$ at each point and $t_i \in [\Delta, 1-\Delta]$ are all interior points. Note that by this algorithm we have trial values of $\delta$ and $m$ beforehand. The proposed number of iterations at step 3 is due to the following fact. If $\delta = \delta^0$, then $h_0$ is of the optimal order so that $h_1$ is already consistent. In the second iteration the affect of $h_0$ will be clearly reduced. The other two iterations are proposed to improve the finite sample property of $\hat{h}$. If $\delta \neq \delta^0$, the selected bandwidth in any iteration would in general not be optimal. In this case more iterations are not necessary. Lemma 1 in the appendix shows insight into AlgA.

The estimated parameters, the selected bandwidth $\hat{h}$ as well as the estimated trend $\hat{g}(t), t \in [0,1]$, by Algorithm A (AlgA) are all consistent.

**Theorem 1.** Let the assumptions of Theorem 3 in Beran (1999) and Proposition 1 in Beran and Feng (1999) hold. Then we have

a) the results for $\hat{\theta}$ as given in theorem 2 in Beran (1999) hold,

b) $\hat{h} = h_M \{1 + O_p(n^{2(2\delta^0-1)/(9-2\delta^0)})\}$,

c) and $\hat{g}(t) = g(t) \{1 + O_p(n^{2(2\delta^0-1)/(5-2\delta^0)})\}$

for $t \in [\Delta, 1-\Delta]$. 

12
The rate of convergence of the selected bandwidth given in (32) follows from (29). A sketched proof of Theorem 1 is given in the appendix. The computing time of AlgA is very long, especially when the grid is fine, since the iterative procedure has to be carried out for each trial value \(d \in G\). In the following we will propose an Algorithm B (AlgB), which is much faster than AlgA, where all parameters, except for \(p\) and \(m\), are estimated from the residuals by means of the S-PLUS function `arima.fracdiff`.

The steps of AlgB are defined as follows:

**Algorithm B:**

**Step 1:** To obtain a bandwidth for selecting \(m\):

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

2. **Step 1b:** For each \(p = 0, 1, \ldots, L\), where \(L\) is as defined in AlgA, estimate a FARIMA model from the residuals using the S-PLUS function `arima.fracdiff`, where the order of the MA component is put to be zero.

3. **Step 1c:** Select the best AR order \(p\) following the BIC. Now we obtain estimates of all parameters except for \(m^0\).

4. **Step 1d:** Calculate the bandwidth \(h_1\) following the procedure in section 3 with \(\alpha = (5 - 2\hat{\delta})/(7 - 2\hat{\delta})\).

5. **Step 1e:** Put \(L = \hat{p}_0\).

**Step 2:** Estimate \(m^0\):

1. **Step 2a:** Carry out steps 1a to 1c with \(h_1\) for \(m = 0\) and \(m = 1\) separately.

2. **Step 2b:** Select the best pair of \(m\) and \(p\) following the BIC. Now we obtain an estimation of all parameters, especially \(\hat{m}^0\).

3. **Step 2c:** Put \(m = \hat{m}^0\).

**Step 3:** Further iterations: Carry out further iterations with \(L\) defined in step 1e, \(m = \hat{m}^0\) 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 \( \text{arima.fracdiff} \) 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 \( \hat{m}_0 \) selected at the first iteration we have \( \hat{m}_0 \xrightarrow{p} p_0 \) in probability, if \( m^0 = 1 \). If \( m^0 = 0 \), then \( \hat{m}_0 \) tends to the maximal order \( L \) in probability, since now the error process in the first difference, \( X_i = X_i - X_{i-1} \), follows an ARMA\((p, 1)\), i.e. an AR\((\infty)\) model. By selecting \( m^0 \) just one time and by putting \( L = \hat{m}_0 \) at the end of step 1 much computing time will be saved. We have

**Theorem 2.** Under the assumptions of Theorem 1 the same results as given in Theorem 1 hold for the estimates obtained by AlgB, except for that here

\[
\hat{h} = h_M \{ 1 + O_p(n^{2(\theta^0-1)/(7-2\theta^0)}) \},
\]

which follows from (30).

The proof of Theorem 2 is straightforward and is hence omitted.

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 a little large, which will sometimes result in \( \hat{m}_0 = 0 \) in the case when \( m^0 = 1 \) (see Beran and Feng, 2000). 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 by AlgB be replaced by \( h_0 = n^{-5/7} \). Carry out similarly the iteration 6 times with the assumption \( m = 1 \). 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 \( h_M \). 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:

\[ g_1(t) = 2 \tan(5(t - 0.5)), \]
\[ g_2(t) = 4 \sin^2((t - 0.5)\pi) \quad \text{and} \]
\[ g_3(t) = 2 \sin(5(t - 0.5)\pi) \]

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

50 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 \( \text{var} (X_i) = 1 \) in all cases. 200 replications were done for each parameter combination with two sample sizes \( n = 500 \) and \( n = 1000 \). The simulations were carried out using AlgB and AlgC, separately. The maximal iterative number was equal to 20. Simulation using AlgA has not been done due to long computing time.

5.2 Summary of results

A detailed analysis of the simulation results is given in a preprint (Beran and Feng, 2000) as a supplement of the current paper, where more detailed description on this simulation may also be found. In the following only a brief summary on the simulation with \( n = 500 \) using AlgB will be given. Tables 1 and 2 give frequencies in 200 replications, when \( m^0 \) or \( p_0 \) is correctly selected, 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 $h_A$ calculated from (20). Note that $h_A$ is the same for a pair of cases with the same parameters except for $m^0$. 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 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 much 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 $\phi^0_1 = 0.7$. And, $\hat{m}^0$ for $g_0$ with $m^0 = 1$ is not satisfactory. This means that now it is difficult to decide, if $Y_t$ is stationary or not. For this case AlgC works clearly better than AlgB (see Beran and Feng, 200).

The order $p_0$ is more difficult to select than $m^0$. There are mainly two reasons for this. Firstly, different autoregressive models may have quite similar finite sample performance. Secondly, 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 whole correct. Note that model (b) in Beran et al. (1998) is the same as the case without trend used in this paper. 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 of the 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 $\phi^0_1$. However, the finite sample performance of $\hat{h}$ depends strongly on both parameters. In general, the larger $\phi^0_1$ and/or $\delta$ is the larger the variation in $\hat{h}$. The performance of $\hat{h}$ also depends on the trend function. The selection of the bandwidth by $g_1$ is more difficult than that for $g_2$ or $g_3$. Estimation of $m^0$ and $p_0$ 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 one (see the case with $\delta^0 = -0.2$ and $\phi^0_1 = 0.7$ 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). \( \hat{h} \) 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 put together. The same results for cases with \( m^0 = 1 \) are shown in Figures 4 to 6.

6 Final remarks

In this paper it is shown that the data-driven SEMIFAR models work well for simultaneous modelling of trend, short-memory as well as long-memory. By checking the detailed simulation results in Beran and Feng (2000) we can find: 1. In general, AlgB works better for \( m^0 = 0 \), while AlgC works better for \( m^0 = 1 \). This becomes more clear by checking the results for the cases \( g_3 \) with \( m^0 = 0 \) and \( g_0 \) with \( m^0 = 1 \). 2. The difference between AlgB and AlgC depends on the trend. For \( g_1 \) and \( g_2 \), their performance is quite similar. 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 over estimated, the short-memory parameter will often be under estimated, and vice versa (see Beran and Feng, 2000).

Acknowledgements

This paper 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.

Appendix: Proofs

The following Lemma will be needed for the proof of Theorem 1. It provides a deeper understanding for the process of AlgA in the case with when \( m = m^0 \).
**Lemma 1.** Assume that the trial value of \( m \) (in AlgA) is equal to \( m^0 \). And assume that the other conditions of Theorem 1 hold. Then for each trial value \( \delta \) there exists an order \((1 - 2\delta)/(5 - 2\delta) \leq \alpha_\delta < \frac{5}{9}\) such that

i) \( h_j = O(h_{j-1}) \), if \( h_{j-1} = O(n^{-\alpha_\delta}) \),

ii) \( h_j = o(h_{j-1}) \), if \( h_{j-1} = O(n^{-\alpha_\delta + d}) \) with \( 0 < d_\delta < \alpha_\delta \),

iii) \( h_{j-1} = o(h_j) \), if \( h_{j-1} = O(n^{-\alpha_\delta - d}) \) with \( 0 < d_\delta < 1 - \alpha_\delta \).

**Proof of Lemma 1:**

i) In the following we will call a bandwidth \( h_f(\delta) = O(n^{-\alpha_\delta}) \) a stable bandwidth for the iterative plug-in procedure with the trial value \( \delta \). For given \( \delta^0 \), define \( \delta_f = \max\{(4\delta^0 - 1)/2, -0.5\} \). It is clear that \( \delta_f < \delta^0 \). Let \( \bar{\alpha} = (1 - 2\delta)/(9 - 2\delta) \). For \( \delta_f < \delta < 0.5 \), we have \( h_{2,1} = h_0^{(5-2\delta)/(9-2\delta)} = O(n^{-\bar{\alpha}}) \) with \( 0 < \bar{\alpha} < (1-2\delta)/(5-2\delta) \).

In this case \( \hat{h} \) is consistent. Now, we have \( h_1 = O(h_0) \) and \( h_j = h_{j-1}(1 + o(1)) \) for \( j = 2, \ldots \). In this case \( \alpha_\delta = (1 - 2\delta)/(5 - 2\delta) \).

The case \( \delta \leq \delta_f \) can only occur if \( \delta_f > -0.5 \) (i.e. \( \delta^0 > 0 \)). Thus suppose that \( \delta_f > -0.5 \). Then we can also obtain that \( \alpha_\delta = (1 - 2\delta)/(5 - 2\delta) \) for \( \delta = \delta_f \). But now, \( \hat{h} \) is a constant rather than a consistent estimate. It can be shown that the required \( \alpha_\delta \) is \( \alpha_\delta = 2(\delta^0 - \delta)(9 - 2\delta)/\{(5 - 2\delta)(4 + 2(\delta^0 - \delta))\} \) for \(-0.5 < \delta < \delta_f \). 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, \( \delta_\alpha \) is monotone increasing in \( \delta^0 \) and monotone decreasing in \( \delta \) with the upper bound \( \frac{5}{9} \).

ii) and iii) can be shown by straightforward calculations using the results in Proposition 1 in Beran and Feng (1999).

**Remark.** Note in particular that, for \( \delta = \delta^0 \), \( \alpha_{\delta^0} = (1 - 2\delta^0)/(5 - 2\delta^0) \). In this case, i) of Lemma 1 may be written as \( h_j = h_M(1 + o(1)) \), for \( j \) large enough. Now, if \( h_M = o(h_{j-1}) \), \( h_{j-1} \) will be deflated. If \( h_{j-1} = o(h_M) \), \( h_{j-1} \) will be inflated. This procedure will be iterative carried out until \( \hat{h} = h_M(1 + o(1)) \) is reached. This is the key point behind the iterative plug-in bandwidth selection rule. It is true for any iterative plug-in bandwidth selector with known \( \delta^0 \) or a consistent estimate.
of it (see Herrmann and Gasser, 1994 for a detailed analysis in the case of iid data). This shows that \( \hat{h} \) selected by any iterative plug-in method has the property 
\( \hat{h} = h_M(1 + o(1)) \), which does not depend on \( h_0 \) and the inflation method, although the rate of convergence of \( \hat{h} \) does.

**A sketched proof of Theorem 1:**

a). Note that, for each \( \delta \), the bandwidth selected at the end of step 3 of AlgA is 
\( \hat{h}(\delta) = h_4 \). Following the proof of Theorem 2 in Beran (1999) it is enough to show that

i) for \( m = m^0 \), \( h_4 \to 0 \), \( nh_4 \to \infty \), and

ii) for \( m \neq m^0 \), \( nh_4 \to \infty \)

as \( n \to \infty \). For \( m \neq m^0 \), the condition \( h_4 \to 0 \) as \( n \to \infty \) is unnecessary, although it can be shown that it holds.

Condition i) follows immediately from Lemma 1.

ii). In the case \( m^0 = 1 \) with \( m = 0 \) we have \( \hat{I} = O(n^2) \) and hence, for each \( j \),

\[ h_j \geq O(n^{-2/(5-2\delta)}n^{(2\delta-1)/(5-2\delta)}) = O(n^{(2\delta-3)/(5-2\delta)}) \].

We have \( nh_4 \to \infty \). In the case \( m^0 = 0 \) with \( m = 1 \), it may be shown that \( \hat{I} \) will be asymptotically dominated by the bias part of order \( h_{2,j}^2 \). Hence, asymptotically, \( h_{j-1} \) will always be enlarged, i.e. \( h_{j-1} = o(h_j) \). The required condition holds. Further proof of part a) follows from the proof of Theorem 2 in Beran (1999).

The proof of part b) is similar to that of Theorem 1 in Beran and Feng (1999). Part c) can be obtained following straightforward calculation by inserting the optimal bandwidth in (9) and (10). The proof of Theorem 1 is finished.

**REFERENCES**


Table 1: Frequencies in 200 replications when $m^0$ or $p_0$ is correctly selected (for simulation using AlgB with $n = 500$ and $m^0 = 0$).

<table>
<thead>
<tr>
<th>$d^0$</th>
<th>$\phi_{10}^0$</th>
<th>$g_1$</th>
<th>$g_2$</th>
<th>$g_3$</th>
<th>$g_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m^0$</td>
<td>$p_0$</td>
<td>$m^0$</td>
<td>$p_0$</td>
<td>$m^0$</td>
</tr>
<tr>
<td>-0.4</td>
<td>-0.7</td>
<td>200</td>
<td>194</td>
<td>200</td>
<td>184</td>
</tr>
<tr>
<td>-0.4</td>
<td>-0.3</td>
<td>200</td>
<td>194</td>
<td>200</td>
<td>188</td>
</tr>
<tr>
<td>-0.4</td>
<td>0</td>
<td>200</td>
<td>197</td>
<td>200</td>
<td>199</td>
</tr>
<tr>
<td>-0.4</td>
<td>0.3</td>
<td>200</td>
<td>170</td>
<td>200</td>
<td>141</td>
</tr>
<tr>
<td>-0.4</td>
<td>0.7</td>
<td>200</td>
<td>101</td>
<td>200</td>
<td>101</td>
</tr>
<tr>
<td>-0.2</td>
<td>-0.7</td>
<td>200</td>
<td>190</td>
<td>200</td>
<td>196</td>
</tr>
<tr>
<td>-0.2</td>
<td>-0.3</td>
<td>200</td>
<td>160</td>
<td>200</td>
<td>181</td>
</tr>
<tr>
<td>-0.2</td>
<td>0</td>
<td>200</td>
<td>179</td>
<td>200</td>
<td>187</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.3</td>
<td>200</td>
<td>185</td>
<td>200</td>
<td>175</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.7</td>
<td>102</td>
<td>19</td>
<td>110</td>
<td>14</td>
</tr>
<tr>
<td>0</td>
<td>-0.7</td>
<td>200</td>
<td>159</td>
<td>200</td>
<td>180</td>
</tr>
<tr>
<td>0</td>
<td>-0.3</td>
<td>200</td>
<td>111</td>
<td>200</td>
<td>120</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>200</td>
<td>169</td>
<td>200</td>
<td>186</td>
</tr>
<tr>
<td>0</td>
<td>0.3</td>
<td>200</td>
<td>155</td>
<td>200</td>
<td>138</td>
</tr>
<tr>
<td>0</td>
<td>0.7</td>
<td>192</td>
<td>191</td>
<td>182</td>
<td>180</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.7</td>
<td>200</td>
<td>166</td>
<td>200</td>
<td>172</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.3</td>
<td>200</td>
<td>131</td>
<td>200</td>
<td>129</td>
</tr>
<tr>
<td>0.2</td>
<td>0</td>
<td>200</td>
<td>172</td>
<td>200</td>
<td>180</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>158</td>
<td>19</td>
<td>159</td>
<td>22</td>
</tr>
<tr>
<td>0.2</td>
<td>0.7</td>
<td>197</td>
<td>195</td>
<td>199</td>
<td>198</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.7</td>
<td>196</td>
<td>195</td>
<td>196</td>
<td>196</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.3</td>
<td>185</td>
<td>148</td>
<td>193</td>
<td>127</td>
</tr>
<tr>
<td>0.4</td>
<td>0</td>
<td>196</td>
<td>199</td>
<td>197</td>
<td>198</td>
</tr>
<tr>
<td>0.4</td>
<td>0.3</td>
<td>150</td>
<td>150</td>
<td>152</td>
<td>151</td>
</tr>
<tr>
<td>0.4</td>
<td>0.7</td>
<td>187</td>
<td>199</td>
<td>184</td>
<td>195</td>
</tr>
</tbody>
</table>
Table 2: Frequencies in 200 replications when $m^0$ or $p_0$ is correctly selected (for simulation using AlgB with $n = 500$ and $m^0 = 1$).

<table>
<thead>
<tr>
<th>$d^0$</th>
<th>$\phi_1^0$</th>
<th>$g_1$</th>
<th>$g_2$</th>
<th>$g_3$</th>
<th>$g_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m^0$</td>
<td>$p_0$</td>
<td>$m^0$</td>
<td>$p_0$</td>
<td>$m^0$</td>
</tr>
<tr>
<td>0.6</td>
<td>-0.7</td>
<td>200</td>
<td>193</td>
<td>200</td>
<td>187</td>
</tr>
<tr>
<td>0.6</td>
<td>-0.3</td>
<td>200</td>
<td>199</td>
<td>196</td>
<td>192</td>
</tr>
<tr>
<td>0.6</td>
<td>0</td>
<td>200</td>
<td>199</td>
<td>198</td>
<td>192</td>
</tr>
<tr>
<td>0.6</td>
<td>0.3</td>
<td>200</td>
<td>59</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7</td>
<td>200</td>
<td>188</td>
<td>200</td>
<td>179</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.7</td>
<td>200</td>
<td>199</td>
<td>200</td>
<td>194</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.3</td>
<td>199</td>
<td>163</td>
<td>200</td>
<td>186</td>
</tr>
<tr>
<td>0.8</td>
<td>0</td>
<td>200</td>
<td>197</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>0.8</td>
<td>0.3</td>
<td>197</td>
<td>160</td>
<td>200</td>
<td>34</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7</td>
<td>199</td>
<td>189</td>
<td>200</td>
<td>194</td>
</tr>
<tr>
<td>1</td>
<td>-0.7</td>
<td>200</td>
<td>196</td>
<td>200</td>
<td>192</td>
</tr>
<tr>
<td>1</td>
<td>-0.3</td>
<td>200</td>
<td>129</td>
<td>200</td>
<td>135</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>200</td>
<td>193</td>
<td>200</td>
<td>199</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>199</td>
<td>167</td>
<td>200</td>
<td>149</td>
</tr>
<tr>
<td>1</td>
<td>0.7</td>
<td>200</td>
<td>171</td>
<td>200</td>
<td>197</td>
</tr>
<tr>
<td>1.2</td>
<td>-0.7</td>
<td>200</td>
<td>180</td>
<td>200</td>
<td>196</td>
</tr>
<tr>
<td>1.2</td>
<td>-0.3</td>
<td>200</td>
<td>123</td>
<td>200</td>
<td>107</td>
</tr>
<tr>
<td>1.2</td>
<td>0</td>
<td>200</td>
<td>185</td>
<td>200</td>
<td>198</td>
</tr>
<tr>
<td>1.2</td>
<td>0.3</td>
<td>200</td>
<td>184</td>
<td>200</td>
<td>182</td>
</tr>
<tr>
<td>1.2</td>
<td>0.7</td>
<td>200</td>
<td>156</td>
<td>200</td>
<td>167</td>
</tr>
<tr>
<td>1.4</td>
<td>-0.7</td>
<td>200</td>
<td>158</td>
<td>200</td>
<td>190</td>
</tr>
<tr>
<td>1.4</td>
<td>-0.3</td>
<td>200</td>
<td>108</td>
<td>200</td>
<td>109</td>
</tr>
<tr>
<td>1.4</td>
<td>0</td>
<td>200</td>
<td>178</td>
<td>200</td>
<td>187</td>
</tr>
<tr>
<td>1.4</td>
<td>0.3</td>
<td>200</td>
<td>190</td>
<td>200</td>
<td>195</td>
</tr>
<tr>
<td>1.4</td>
<td>0.7</td>
<td>200</td>
<td>140</td>
<td>200</td>
<td>138</td>
</tr>
</tbody>
</table>
Table 3: Mean and standard deviation of \( \hat{h} \) (using AlgB with \( n = 500, m^0 = 0 \)).

<table>
<thead>
<tr>
<th>( d^0 )</th>
<th>( \phi_1^0 )</th>
<th>( g_1 )</th>
<th>( g_2 )</th>
<th>( g_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( h_A ) Mean SD</td>
<td>( h_A ) Mean SD</td>
<td>( h_A ) Mean SD</td>
</tr>
<tr>
<td>-0.4</td>
<td>-0.7</td>
<td>0.053 0.050 0.0039</td>
<td>0.039 0.040 0.0015</td>
<td>0.021 0.021 0.0009</td>
</tr>
<tr>
<td>-0.4</td>
<td>-0.3</td>
<td>0.065 0.061 0.0048</td>
<td>0.048 0.051 0.0013</td>
<td>0.026 0.027 0.0008</td>
</tr>
<tr>
<td>-0.4</td>
<td>0</td>
<td>0.075 0.068 0.0059</td>
<td>0.055 0.058 0.0017</td>
<td>0.029 0.031 0.0007</td>
</tr>
<tr>
<td>-0.4</td>
<td>0.3</td>
<td>0.086 0.081 0.0094</td>
<td>0.063 0.066 0.0036</td>
<td>0.034 0.035 0.0014</td>
</tr>
<tr>
<td>-0.4</td>
<td>0.7</td>
<td>0.114 0.139 0.0563</td>
<td>0.084 0.106 0.0214</td>
<td>0.045 0.055 0.0063</td>
</tr>
<tr>
<td>-0.2</td>
<td>-0.7</td>
<td>0.059 0.054 0.0046</td>
<td>0.043 0.044 0.0018</td>
<td>0.022 0.022 0.0009</td>
</tr>
<tr>
<td>-0.2</td>
<td>-0.3</td>
<td>0.074 0.066 0.0074</td>
<td>0.053 0.055 0.0023</td>
<td>0.027 0.028 0.0011</td>
</tr>
<tr>
<td>-0.2</td>
<td>0</td>
<td>0.084 0.072 0.0080</td>
<td>0.061 0.062 0.0038</td>
<td>0.031 0.032 0.0010</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.3</td>
<td>0.097 0.089 0.0145</td>
<td>0.070 0.073 0.0064</td>
<td>0.035 0.039 0.0023</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.7</td>
<td>0.125 0.176 0.1076</td>
<td>0.090 0.131 0.0365</td>
<td>0.046 0.082 0.0210</td>
</tr>
<tr>
<td>0</td>
<td>-0.7</td>
<td>0.075 0.066 0.0083</td>
<td>0.053 0.054 0.0037</td>
<td>0.025 0.025 0.0013</td>
</tr>
<tr>
<td>0</td>
<td>-0.3</td>
<td>0.094 0.079 0.0126</td>
<td>0.066 0.065 0.0076</td>
<td>0.032 0.032 0.0016</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.106 0.091 0.0144</td>
<td>0.075 0.076 0.0089</td>
<td>0.036 0.038 0.0029</td>
</tr>
<tr>
<td>0</td>
<td>0.3</td>
<td>0.120 0.120 0.0493</td>
<td>0.084 0.095 0.0208</td>
<td>0.041 0.050 0.0080</td>
</tr>
<tr>
<td>0</td>
<td>0.7</td>
<td>0.150 0.128 0.0267</td>
<td>0.106 0.105 0.0147</td>
<td>0.051 0.061 0.0113</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.7</td>
<td>0.102 0.086 0.0164</td>
<td>0.069 0.068 0.0089</td>
<td>0.031 0.036 0.0339</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.3</td>
<td>0.126 0.104 0.0226</td>
<td>0.086 0.083 0.0139</td>
<td>0.039 0.042 0.0258</td>
</tr>
<tr>
<td>0.2</td>
<td>0</td>
<td>0.140 0.125 0.0385</td>
<td>0.095 0.096 0.0137</td>
<td>0.043 0.047 0.0056</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>0.154 0.208 0.1157</td>
<td>0.105 0.134 0.0311</td>
<td>0.047 0.074 0.0184</td>
</tr>
<tr>
<td>0.2</td>
<td>0.7</td>
<td>0.180 0.179 0.0757</td>
<td>0.123 0.125 0.0193</td>
<td>0.055 0.065 0.0101</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.7</td>
<td>0.141 0.118 0.0423</td>
<td>0.093 0.090 0.0141</td>
<td>0.039 0.066 0.0942</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.3</td>
<td>0.164 0.139 0.0666</td>
<td>0.107 0.100 0.0222</td>
<td>0.045 0.066 0.1029</td>
</tr>
<tr>
<td>0.4</td>
<td>0</td>
<td>0.173 0.185 0.0913</td>
<td>0.114 0.122 0.0200</td>
<td>0.048 0.057 0.0453</td>
</tr>
<tr>
<td>0.4</td>
<td>0.3</td>
<td>0.181 0.105 0.0293</td>
<td>0.119 0.092 0.0157</td>
<td>0.050 0.069 0.0341</td>
</tr>
<tr>
<td>0.4</td>
<td>0.7</td>
<td>0.193 0.197 0.0923</td>
<td>0.126 0.133 0.0250</td>
<td>0.053 0.064 0.0140</td>
</tr>
</tbody>
</table>
Table 4: Mean and standard deviation of $\hat{h}$ (using AlgB with $n = 500, m^0 = 1$).

<table>
<thead>
<tr>
<th>$d^0$</th>
<th>$\phi_1^0$</th>
<th>$g_1$</th>
<th>$g_2$</th>
<th>$g_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h_A$</td>
<td>Mean</td>
<td>SD</td>
<td>$h_A$</td>
</tr>
<tr>
<td>0.6</td>
<td>0.053</td>
<td>0.050</td>
<td>0.0036</td>
<td>0.039</td>
</tr>
<tr>
<td>0.6</td>
<td>0.065</td>
<td>0.062</td>
<td>0.0060</td>
<td>0.048</td>
</tr>
<tr>
<td>0.6</td>
<td>0.075</td>
<td>0.068</td>
<td>0.0057</td>
<td>0.055</td>
</tr>
<tr>
<td>0.6</td>
<td>0.086</td>
<td>0.079</td>
<td>0.0106</td>
<td>0.063</td>
</tr>
<tr>
<td>0.6</td>
<td>0.114</td>
<td>0.108</td>
<td>0.0221</td>
<td>0.084</td>
</tr>
<tr>
<td>0.8</td>
<td>0.059</td>
<td>0.053</td>
<td>0.0042</td>
<td>0.043</td>
</tr>
<tr>
<td>0.8</td>
<td>0.074</td>
<td>0.065</td>
<td>0.0066</td>
<td>0.053</td>
</tr>
<tr>
<td>0.8</td>
<td>0.084</td>
<td>0.070</td>
<td>0.0073</td>
<td>0.061</td>
</tr>
<tr>
<td>0.8</td>
<td>0.097</td>
<td>0.091</td>
<td>0.0139</td>
<td>0.070</td>
</tr>
<tr>
<td>0.8</td>
<td>0.125</td>
<td>0.109</td>
<td>0.0215</td>
<td>0.090</td>
</tr>
<tr>
<td>1</td>
<td>0.075</td>
<td>0.061</td>
<td>0.0064</td>
<td>0.053</td>
</tr>
<tr>
<td>1</td>
<td>0.094</td>
<td>0.073</td>
<td>0.0112</td>
<td>0.066</td>
</tr>
<tr>
<td>1</td>
<td>0.106</td>
<td>0.084</td>
<td>0.0124</td>
<td>0.075</td>
</tr>
<tr>
<td>1</td>
<td>0.120</td>
<td>0.110</td>
<td>0.0268</td>
<td>0.084</td>
</tr>
<tr>
<td>1</td>
<td>0.150</td>
<td>0.136</td>
<td>0.0355</td>
<td>0.106</td>
</tr>
<tr>
<td>1.2</td>
<td>0.102</td>
<td>0.082</td>
<td>0.0167</td>
<td>0.069</td>
</tr>
<tr>
<td>1.2</td>
<td>0.126</td>
<td>0.100</td>
<td>0.0470</td>
<td>0.086</td>
</tr>
<tr>
<td>1.2</td>
<td>0.140</td>
<td>0.123</td>
<td>0.0323</td>
<td>0.095</td>
</tr>
<tr>
<td>1.2</td>
<td>0.154</td>
<td>0.124</td>
<td>0.0547</td>
<td>0.105</td>
</tr>
<tr>
<td>1.2</td>
<td>0.180</td>
<td>0.193</td>
<td>0.0883</td>
<td>0.123</td>
</tr>
<tr>
<td>1.4</td>
<td>0.141</td>
<td>0.133</td>
<td>0.0707</td>
<td>0.093</td>
</tr>
<tr>
<td>1.4</td>
<td>0.164</td>
<td>0.150</td>
<td>0.0820</td>
<td>0.107</td>
</tr>
<tr>
<td>1.4</td>
<td>0.173</td>
<td>0.196</td>
<td>0.1036</td>
<td>0.114</td>
</tr>
<tr>
<td>1.4</td>
<td>0.181</td>
<td>0.120</td>
<td>0.0562</td>
<td>0.119</td>
</tr>
<tr>
<td>1.4</td>
<td>0.193</td>
<td>0.230</td>
<td>0.1311</td>
<td>0.126</td>
</tr>
</tbody>
</table>
Figure 1: Kernel densities of $\log(\hat{h}/h)$ 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.
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 5a: $\phi_1 = -0.7$, all delta's

Figure 5b: $\phi_1 = -0.3$, all delta's

Figure 5c: $\phi_1 = 0.0$, all delta's

Figure 5d: $\phi_1 = 0.3$, all delta's

Figure 5e: $\phi_1 = 0.7$, all delta's
Figure 6: The same results as given in Figures 3a through 3e but for $m^0 = 1$. 