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

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

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 ϵ_i $(i = ..., -1, 0, 1, 2, ...)$ 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

$$(1-B)^{\delta} = \sum_{k=0}^{\infty} \beta_k(\delta) B^k \tag{2}$$

with

$$\beta_k(\delta) = (-1)^k \frac{\Gamma(\delta+1)}{\Gamma(k+1)\Gamma(\delta-k+1)}. \tag{3}$$

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, δ, p and ϕ_1, ..., ϕ_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 quit 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

$$U_i = g(t_i) + X_i, \tag{4}$$

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 δ. The spectral density of X_i in (4) has the form

$$f(\lambda) \sim c_f |\lambda|^{-\alpha} \qquad (\text{as } \lambda \to 0) \tag{5}$$

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 non-summable. 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

$$\hat{g}(t; h) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{t - t_i}{h}) Y_i. \tag{6}$$

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

$$I(g^{''}) = \int_{\Delta}^{1-\Delta} [g^{''}(t)]^2 dt \tag{7}$$

and

$$I(K) = \int_{-1}^{1} x^2 K(x) dx. \tag{8}$$

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:*

$$E[\hat{g}(t) - g(t)] = h^2 \frac{g^{''}(t) I(K)}{2} + o(h^2) \tag{9}$$

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

(ii) *Variance:*

$$\text{var}(\hat{g}(t)) = \frac{1}{(nh)^{1-2\delta}} [V + o(1)] \tag{10}$$

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

$$\begin{aligned} E\left\{\int_{\Delta}^{1-\Delta} [\hat{g}(t) - g(t)]^2 dt\right\} &= h^4 \frac{I(g^{''}) I^2(K)}{4} + (nh)^{2\delta-1} V(1 - 2\Delta) \\ &+ o(\max(h^4, (nh)^{2\delta-1})). \end{aligned} \tag{11}$$

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 δ, may be estimated by maximum likelihood (Beran, 1995, 1999). Note that, since m is an integer, m and δ correspond to one parameter $d = m + \delta$ only, through $m = [d + 0.5]$ and $\delta = d - m$, where $[\cdot]$ denotes the integer part. Let $\theta^0 = (\sigma^2_{\epsilon,0}, d^0, \phi^0_1, ..., \phi^0_p)^{\mathrm{T}} = (\sigma^2_{\epsilon,0}, \eta^0)^{\mathrm{T}}$ 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 θ^0, based on the autoregressive representation of the process, is considered in Beran (1995). Beran (1999) extended this idea to estimate θ^0 in the SEMIFAR model with a general nonparametric trend function g. Note that

$$\begin{aligned}\phi(B)(1-B)^{\delta^0}\{(1-B)^{m^0}Y_i - g(t_i)\} &= \sum_{j=0}^{\infty} a_j(\eta^0)B^j\{c_j(\eta^0)Y_i - g(t_i)\} \\ &= \sum_{j=0}^{\infty} a_j(\eta^0)\{c_j(\eta^0)Y_{i-j} - g(t_{i-j})\},\end{aligned}$$

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

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

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 ϵ_i as a function of η. For a chosen value of $\theta = (\sigma^2_\epsilon, m + \delta, \phi_1, ..., \phi_p)^{\mathrm{T}} = (\sigma_\epsilon, \eta)^{\mathrm{T}}$, denote by

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

the (approximate) residuals and by $r_i(\theta) = e_i(\eta)/\sqrt{\theta_1}$ the standardized residuals. Assuming that $\{\epsilon_i(\eta^0)\}$ are independent zero mean normal with variance σ^2_ϵ, an approximate maximum likelihood estimate of θ^0 is obtained by maximizing the approximate log-likelihood

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

with respect to θ and hence by solving the equations

$$\dot{l}(Y_i, ..., Y_n; \theta) = 0, \tag{15}$$

where $\dot{l}$ is the vector of partial derivatives with respect to θ_j $(j = 1, ..., p+2)$. More explicitly, $\hat{\eta}$ is obtained by minimizing

$$S_n(\eta) = \frac{1}{n} \sum_{i=m+2}^{n} e_i^2(\eta) \tag{16}$$

with respect to η and setting

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{i=m+2}^{n} e_i^2(\hat{\eta}). \tag{17}$$

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 θ^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 δ 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

$$\Sigma = 2D^{-1}, \tag{18}$$

where

$$D_{ij} = (2\pi)^{-1} \left\{ \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f(x) \frac{\partial}{\partial \theta_j} \log f(x) dx \right\} |_{\theta=\theta_*^0} \tag{19}$$

with $\theta_*^0 = (\sigma_{\epsilon,0}^2, \delta^0, \phi_1^0, ..., \phi_p^0)^{\mathrm{T}}$ (see Theorem 2 in Beran, 1999). This result can be extended to the case where the innovations ϵ_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 nonparametric 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

$$h_{\mathrm{A}} = C \cdot n^{(2\delta-1)/(5-2\delta)} \tag{20}$$

with

$$C = \left(\frac{(1-2\delta)V(1-2\Delta)}{I(g'')I^2(K)} \right)^{1/(5-2\delta)}. \tag{21}$$

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

$$C = \left(\frac{9(1-2\delta)\nu(\delta)(1-2\Delta)c_f}{I(g'')} \right)^{1/(5-2\delta)} \tag{22}$$

with c_f as defined before and

$$\nu(\delta) = \frac{2^{2\delta}\Gamma(1-2\delta)\sin(\pi\delta)}{\delta(2\delta+1)} \tag{23}$$

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

Plug-in estimators for h_{M} use formula (20), replacing the unknown constants δ, 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, δ 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

$$\hat{I}(g'') = n^{-1} \sum_{i=[n\Delta]}^{n-[n\Delta]} \{\hat{g}''(t_i; h_2)\}^2. \tag{24}$$

Properties of $\hat{I}(g'')$ are investigated by Beran and Feng (1999). Under the assumption of Proposition 1 in Beran and Feng (1999) we have

$$E[\hat{I}(g'') - I(g'')] \doteq h_2^2 \frac{I(K_2)}{12} \int_{\Delta}^{1-\Delta} g''(t) g^{(4)}(t) dt + (nh_2)^{2\delta-1} h_2^{-4} V \tag{25}$$

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

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

The optimal bandwidth for estimating $I(g'')$ which minimizes the MSE is $h_2^o = O(n^{(2\delta-1)/(7-2\delta)})$.

Following the iterative plug-in idea of Gasser et al. (1991), in the jth iteration, $I(g'')$ is estimated with a bandwidth $h_{2,j}$, which is obtained from the bandwidth for estimating g in the j-1th 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 α. 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 δ 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 $h_0 = o(h_{\rm A})$ for all $\delta \in (-0.5, 0.5)$. Here we used $h_0 = n^{-5/7}$, which is of oder $o(h_{\rm A}^2)$ 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 α. The iterative plug-in algorithm is motivated by fixed point search (see Lemma 1 in the appendix). So α should be chosen in a way that $c \cdot h_{\rm A} n^\alpha = h_2^{\rm o}$ by the MIM, or $c \cdot (h_{\rm A})^\alpha = h_2^{\rm o}$ 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 $\alpha_o = (5 - 2\delta)/(7 - 2\delta)$ 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 α, namely the naive one $\alpha_{\rm n}$ that optimizes $\hat{g}''$ itself and the variance optimal one $\alpha_{\rm v}$ for which the square of second term in (25) is of the order $O(n^{2\delta-1})$. The required bandwidths to estimate $\hat{g}''$ in these two cases are $h_2^{\rm n} = O(n^{(2\delta-1)/(9-2\delta)})$ and $h_2^{\rm v} = O(n^{(2\delta-1)/(2(5-2\delta))})$, respectively. For the MIM we have $\alpha_{\rm n} = (4 - 8\delta)/[(5 - 2\delta)(9 - 2\delta)]$ and $\alpha_{\rm v} = (1 - 2\delta)/(10 - 4\delta)$. They are $\alpha_{\rm n} = (5 - 2\delta)/(9 - 2\delta)$ and $\alpha_{\rm v} = \frac{1}{2}$ for the EIM. The rate of convergence of $\hat{h}$ with $\alpha_{\rm n}$ lies between the two with $\alpha_{\rm o}$ and $\alpha_{\rm v}$. Ray and Tsay (1997) used the MIM with $\alpha_{\rm v}$, while the EIM with $\alpha_{\rm n}$ 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 α 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 $m = 0$.

i) Start with the bandwidth $h_0 = n^{-\beta}$ with $\frac{1}{3} \le \beta < 1$ and set $j = 1$.

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

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

$$h_j = \left(\frac{1 - 2\hat{\delta}}{\beta^2} \frac{(1 - 2\Delta)\hat{V}}{\hat{I}(g''(t; h_{2,j}))} \right)^{1/(5-2\hat{\delta})} \cdot n^{(2\hat{\delta}-1)/(5-2\hat{\delta})}. \qquad (27)$$

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 α). 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 $h_{\mathrm{A}} - h_{\mathrm{M}} = o_p(\hat{I}(g'') - I(g''))$, 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 $\alpha = \alpha_{\mathrm{v}} = \frac{1}{2}$

$$\hat{h} = h_{\mathrm{M}} \left\{ 1 + O(n^{(2\delta-1)/(5-2\delta)}) + O_p(n^{(2\delta-1)/2}) + O_p(n^{-1/2}) \right\}. \qquad (28)$$

ii) For $\alpha = \alpha_{\mathrm{n}} = (5 - 2\delta)/(9 - 2\delta)$

$$\hat{h} = h_{\mathrm{M}} \left\{ 1 + O_p(n^{2(2\delta-1)/(9-2\delta)}) \right\}. \qquad (29)$$

iii) For $\alpha = \alpha_{\mathrm{o}} = (5 - 2\delta)/(7 - 2\delta)$

$$\hat{h} = h_{\mathrm{M}} \left\{ 1 + O_p(n^{2(2\delta-1)/(7-2\delta)}) \right\}. \qquad (30)$$

Proof of these results will be omitted to save place. If $\alpha = \alpha_{\mathrm{o}}$ 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).

4 Data-driven algorithms

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 Δ_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 $\tilde{e}_i(d) = \sum_{j=0}^{i-1} \beta_j(\delta)\hat{X}_{i-j}$, where the coefficients β_j are defined by (3).

Step 3d: Estimate the autoregressive parameters $\phi_1, ..., \phi_p$ from $\tilde{e}_i(d)$ and obtain the estimates $\hat{\sigma}_\epsilon^2 = \hat{\sigma}_\epsilon^2(d; j)$ 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 $\hat{\sigma}_\epsilon^2$ equal to $n^{-1} \sum \tilde{e}_i^2(d)$ and $\hat{c}_f$ equal to $\hat{\sigma}_\epsilon^2/(2\pi)$.

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

$$h_j = \left(\frac{1 - 2\delta}{\beta^2} \frac{(1 - 2\Delta)\hat{V}}{\hat{I}(g''(t; h_{2,j}))} \right)^{1/(5-2\delta)} \cdot n^{(2\delta-1)/(5-2\delta)}. \quad (31)$$

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 $\hat{\sigma}_\epsilon^2(d)$, as a function of d.

Step 4: Define $\hat{d}$ to be the value of d for which $\hat{\sigma}_\epsilon^2(d)$ is minimal. This together with the corresponding estimates of the AR parameters, yields an information criterion, e.g. $\text{BIC}(p) = n \log \hat{\sigma}_\epsilon^2(p) + p \log n$, 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 θ^0 and g.

Here Δ_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 δ 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)

$$\hat{h} = h_{\mathrm{M}}\{1 + O_p(n^{2(2\delta^0-1)/(9-2\delta^0)})\}, \tag{32}$$

c) and

$$\hat{g}(t) = g(t)\{1 + O_p(n^{2(2\delta^0-1)/(5-2\delta^0)})\} \tag{33}$$

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

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:

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.

Step 1b: For each $p = 0, 1, ..., L$, 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.

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

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

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

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 estimation of all parameters, especially $\hat{m}^0$.

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 *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 we have $\hat{p}_0 \stackrel{\text{P}}{\to} p_0$ 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 in the first difference, $\tilde{X}_i = X_i - X_{i-1}$, follows an ARMA($p$, 1), i.e. an AR($\infty$) model. By selecting m^0 just one time and by putting $L = \hat{p}_0$ 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*

$$\hat{h} = h_{\mathrm{M}}\{1 + O_p(n^{2(2\delta^0-1)/(7-2\delta^0)})\}, \tag{34}$$

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:

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

for $t \in [0,1]$ (see Figures 1f through 3f). The range of these trends is kept the same. These trends are chosen as "*orthogonal*" as possible so that the practical performance of the proposed algorithms in different cases may be found. The case without trend ($g_0 :\equiv 0$) 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 $h_{\rm A}$ calculated from (20). Note that $h_{\rm A}$ 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 $\phi_1^0 = 0.7$. 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_i 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 δ not on ϕ_1^0. However, the finite sample performance of $\hat{h}$ depends strongly on both parameters. In general, the larger ϕ_1^0 and/or δ 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 $\phi_1^0 = 0.7$ 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). $\hat{h}$ performs quite quite the same way for $m^0 = 0$ and $m^0 = 1$. Figures 1 through 3 show the estimated kernel densities of $\log(\hat{h}/h_{\mathrm{A}})$ from the 200 replications for each case with $m^0 = 0$, where densities for the same ϕ_1^0 with different δ'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* δ *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 $h_{j-1} = O(n^{-\alpha_\delta + d_\delta})$ *with* $0 < d_\delta < \alpha_\delta$,

iii) $h_{j-1} = o(h_j)$, if $h_{j-1} = O(n^{-\alpha_\delta - d_\delta})$ *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 δ. For given δ^0, define $\delta_f = \max\{(4\delta^0 - 1)/2, -0.5\}$. It is clear that $\delta_f < \delta^0$. Let $\tilde{\alpha} = (1-2\delta)/(9-2\delta)$. For $\delta_f < \delta < 0.5$, we have $h_{2,1} = h_0^{(5-2\delta)/(9-2\delta)} = O(n^{-\tilde{\alpha}})$ with $0 < \tilde{\alpha} < (1-2\delta^0)/(5-2\delta^0)$. In this case $\hat{I}$ 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, $\hat{I}$ is a constant rather than a consistent estimate. It can be shown that the required α_δ 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, δ_α is monotone increasing in δ^0 and monotone decreasing in δ 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 $h_j = h_{\mathrm{M}}(1 + o(1))$, for j large enough. Now, if $h_{\mathrm{M}} = o(h_{j-1})$, h_{j-1} will be deflated. If $h_{j-1} = o(h_{\mathrm{M}})$, h_{j-1} will be inflated. This procedure will be iterative carried out until $\hat{h} = h_{\mathrm{M}}(1 + o(1))$ 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 δ^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 $\hat{h} = h_{\mathrm{M}}(1 + o(1))$, 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 δ, 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

Akaike, H. (1979), "A Bayesian extension of the minimum AIC procedure of autoregressive model fitting," *Biometrika,* 26, 237-242.

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

Beran, J. (1995), "Maximum likelihood of estimation of the differencing parameter for invertible short- and long-memory autoregressive integrated moving average models," *J. Roy. Statist. Soc. Ser. B*, 57, 659–672.

Beran, J. (1999), "SEMIFAR models – A semiparametric framework for modelling trends, long range dependence and nonstationarity," Discussion paper No. 99/16, Center of Finance and Econometrics, University of Konstanz.

Beran, J., 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. (1999), "Locally polynomial fitting with long-range dependent errors," Preprint, University of Konstanz.

Beran, J. and Feng, Y. (2000), "Supplement to 'Data-driven estimation of semiparametric fractional autoregressive models' – Detailed simulation results," Preprint, University of Konstanz.

Beran, J. and Ocker, D. (1999a) "Volatility of stock market indices - An analysis based on SEMIFAR models," Discussion paper No. 99/14, Center of Finance and Econometrics, University of Konstanz.

Beran, J. and Ocker, D. (1999b), "SEMIFAR forecasts, with applications to foreign exchange rates", *J. Statistical Planning and Inference*, 80, 137–153.

Fan, J. and Gijbels, I. (1995), "Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation," J. Roy. Statist. Soc. Ser. B, 57, 371–394.

Gasser, T., Kneip, A. and Köhler, W. (1991), "A flexible and fast method for automatic smoothing," *J. Amer. Statist. Assoc.*, 86, 643–652.

Gasser, T. and Müller, H.G. (1984), "Estimating regression functions and their derivatives by the kernel method," Scand. J. Statist., 11, 171–185.

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

Härdle, W., Hall, P. and Marron, J.S. (1992), "Regression smoothing parameters that are not far from their optimum," *J. Amer. Statist. Assoc.*, 87, 227–233.

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

Heiler, S. and Feng, Y. (1998), "A root n bandwidth selector for nonparametric regression," *J. Nonparametric Statist.*, 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", *Statist. Decisions*, Supp. 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," *J. Amer. Statist. Assoc.* 90, 1257–1270.

Schwarz, G. (1978), "Estimating the dimension of a model", *Ann. Statist.* 6, 461-464.

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

		g_1		g_2		g_3		g_0	
d^0	ϕ_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

Table 2: Frequencies in 200 replications when m^0 or p_0 is correctly selected (for simulation using AlgB with $n = 500$ and $m^0 = 1$).

		g_1		g_2		g_3		g_0	
d^0	ϕ_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

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	ϕ_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

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	ϕ_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

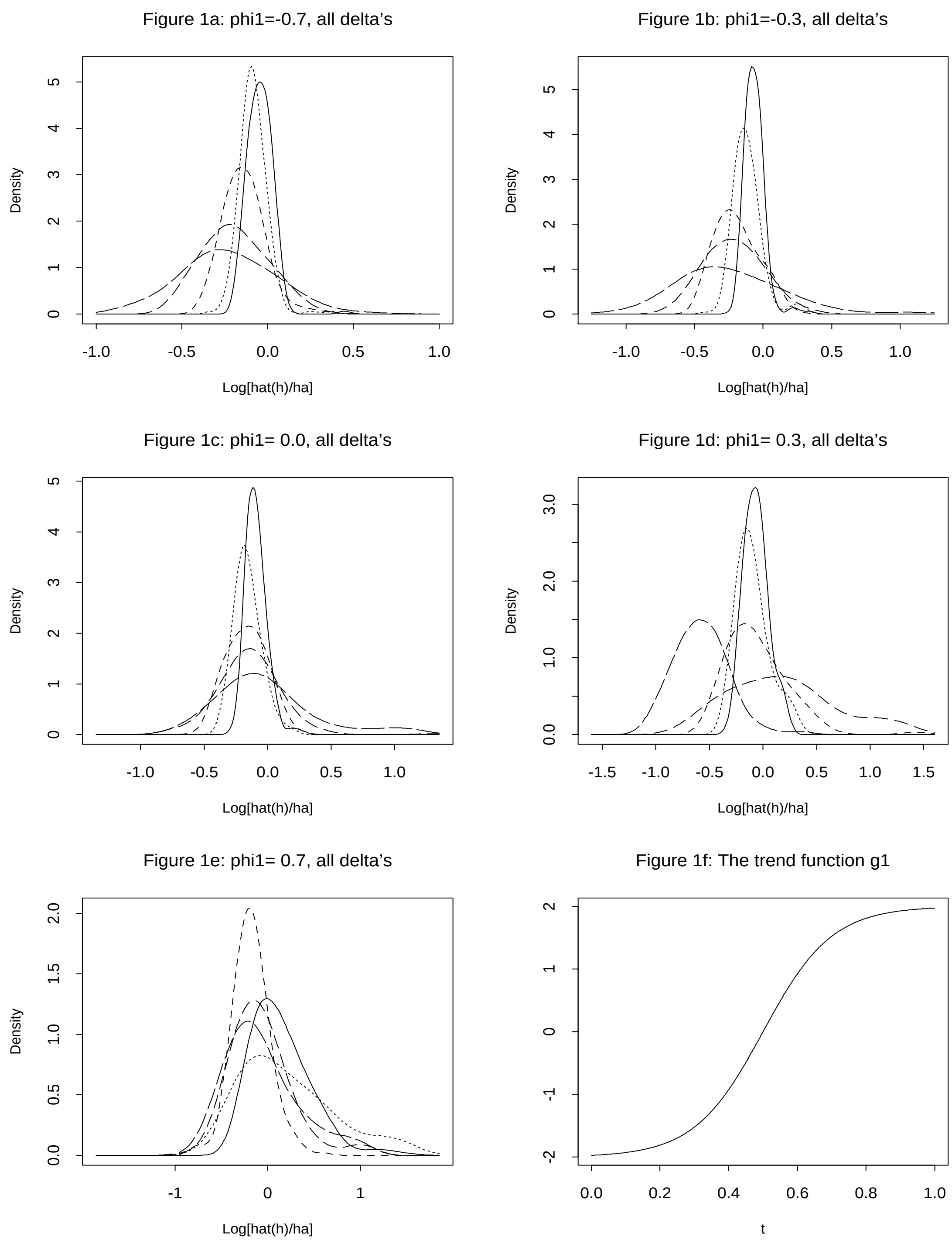

Figure 1: Kernel densities of $\log(\hat{h}/h_{\mathrm{A}})$ selected by AlgB for g_1 with $m^0 = 0$, $n = 500$. Lines in Figures 1a trough 1e are for $\phi_1^0 = -0.7$ to $\phi_1^0 = 0.7$ with all δ^0's — solid line: $\delta^0 = -0.4$, points: $\delta^0 = -0.2$, short dashes: $\delta^0 = 0$, middle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

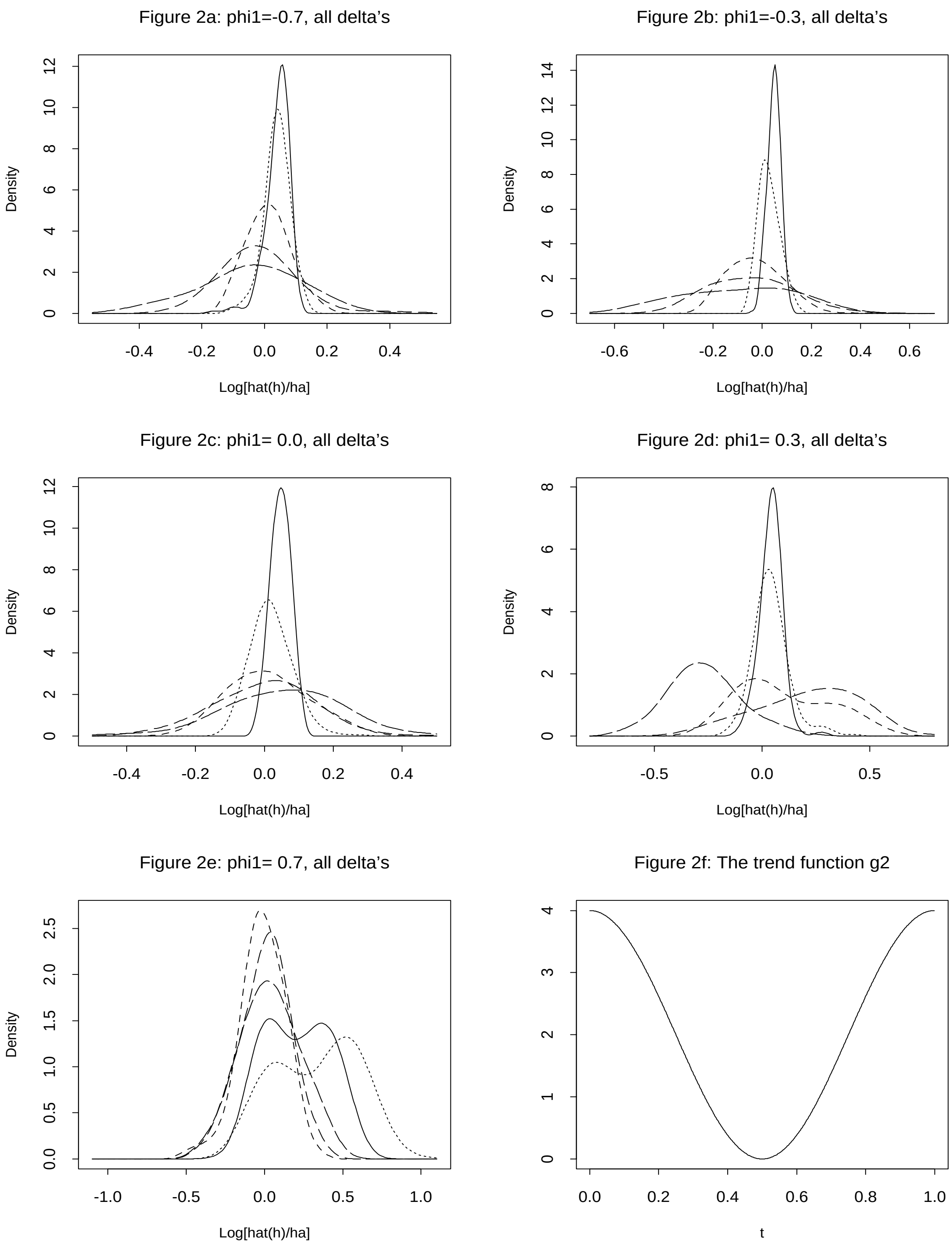

Figure 2: The same results as given in Figure 1 but for the trend function g_2.

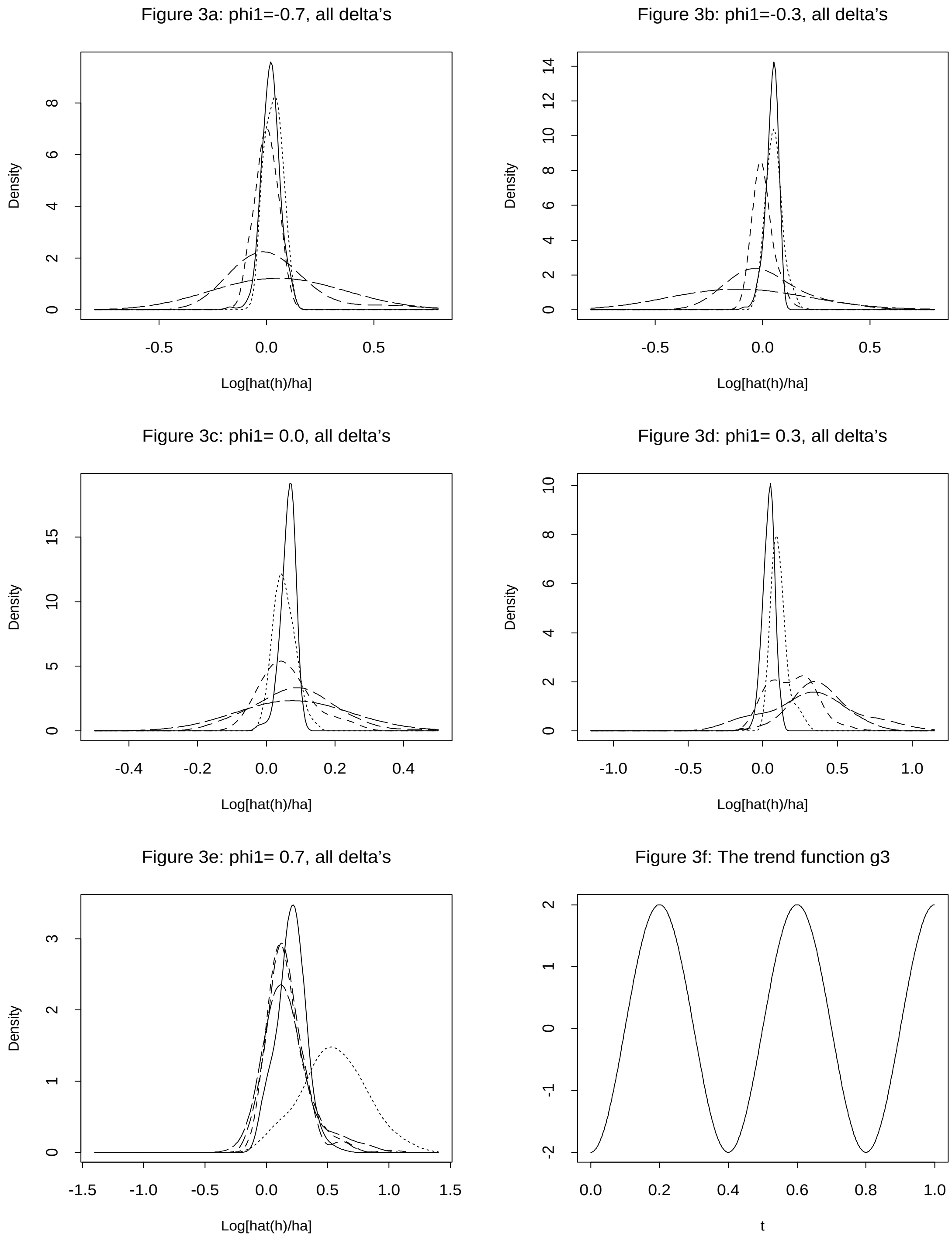

Figure 3: The same results as given in Figure 1 but for the trend function g_3.

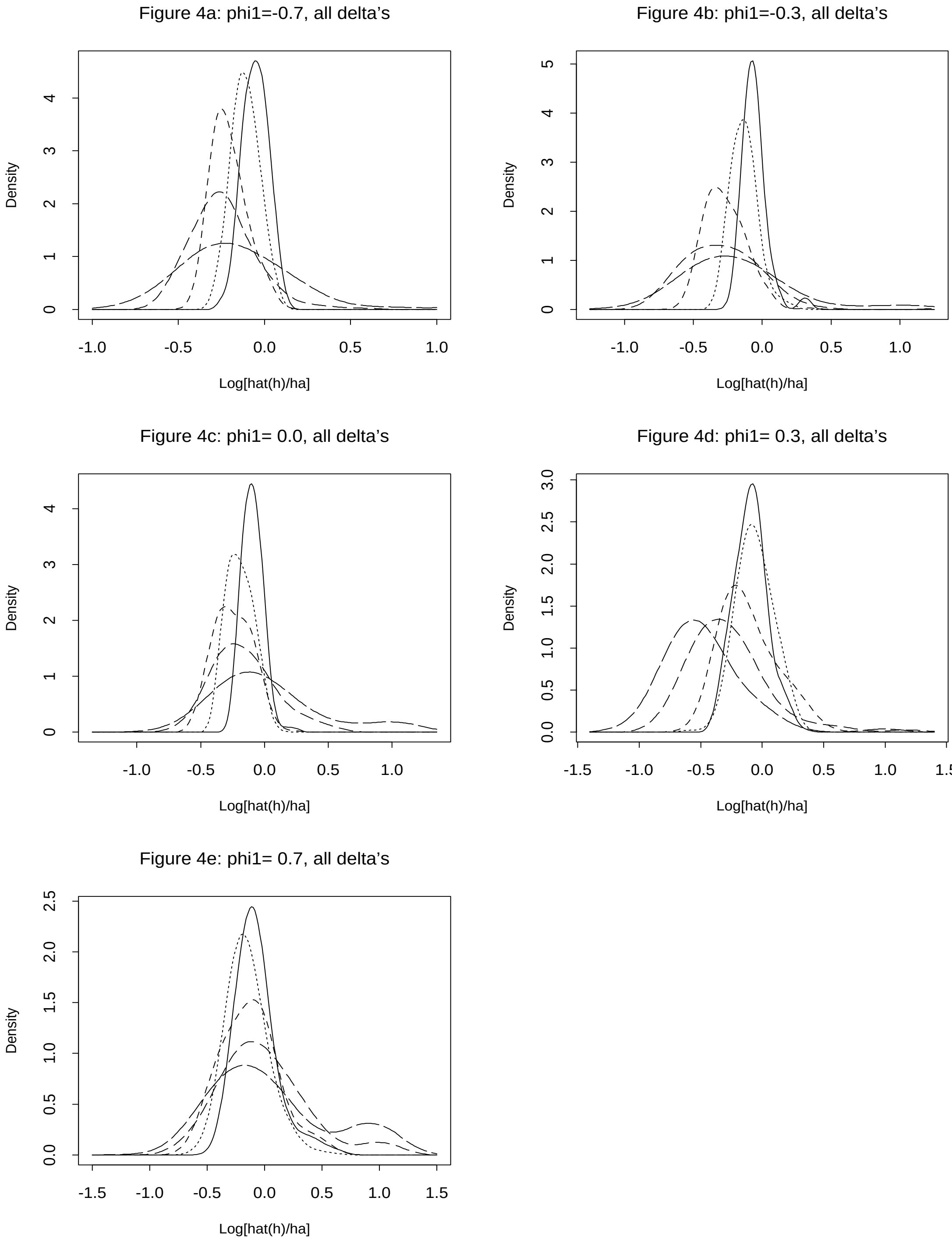

Figure 4: The same results as given in Figures 1a through 1e but for $m^0 = 1$.

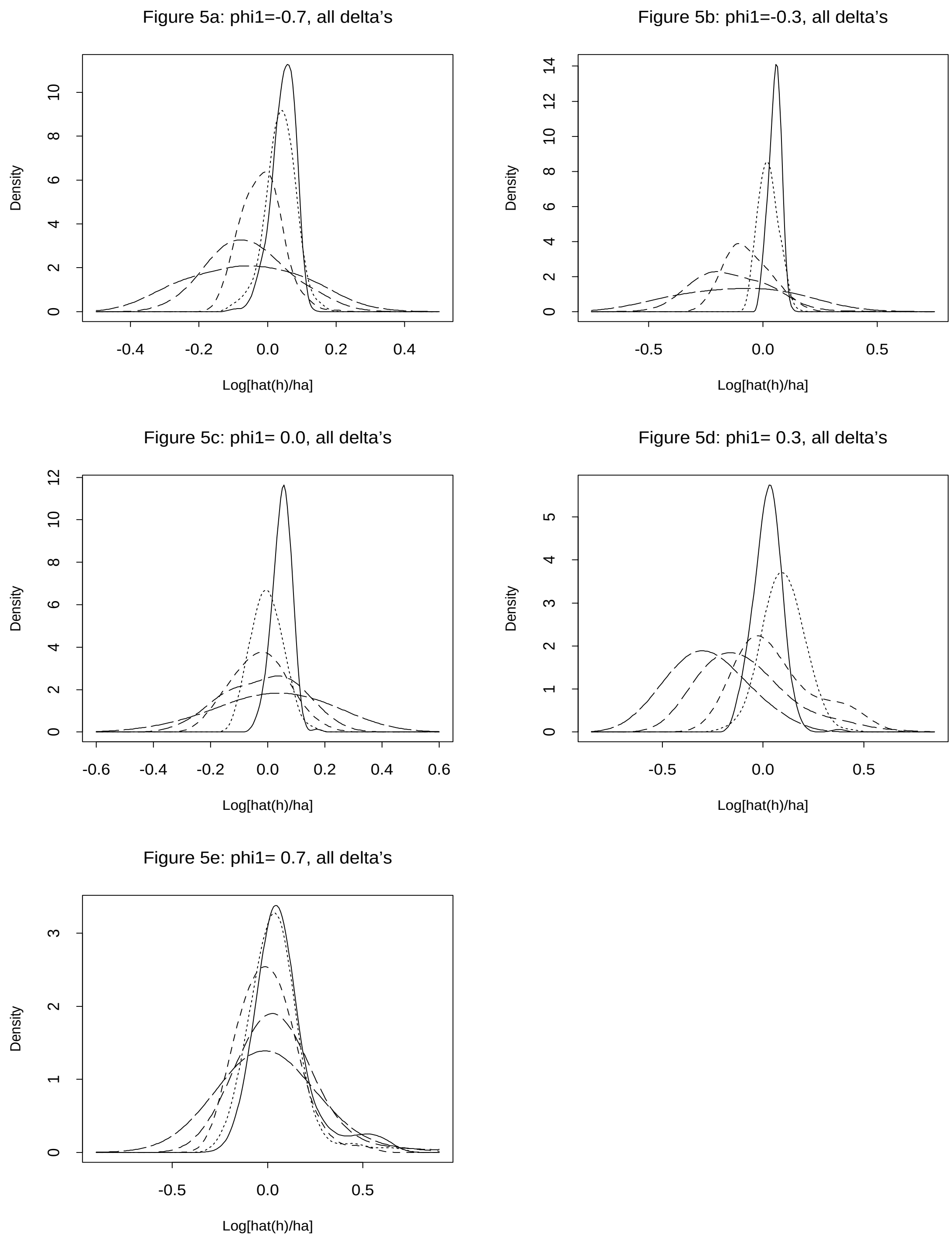

Figure 5: The same results as given in Figures 2a through 2e but for $m^0 = 1$.

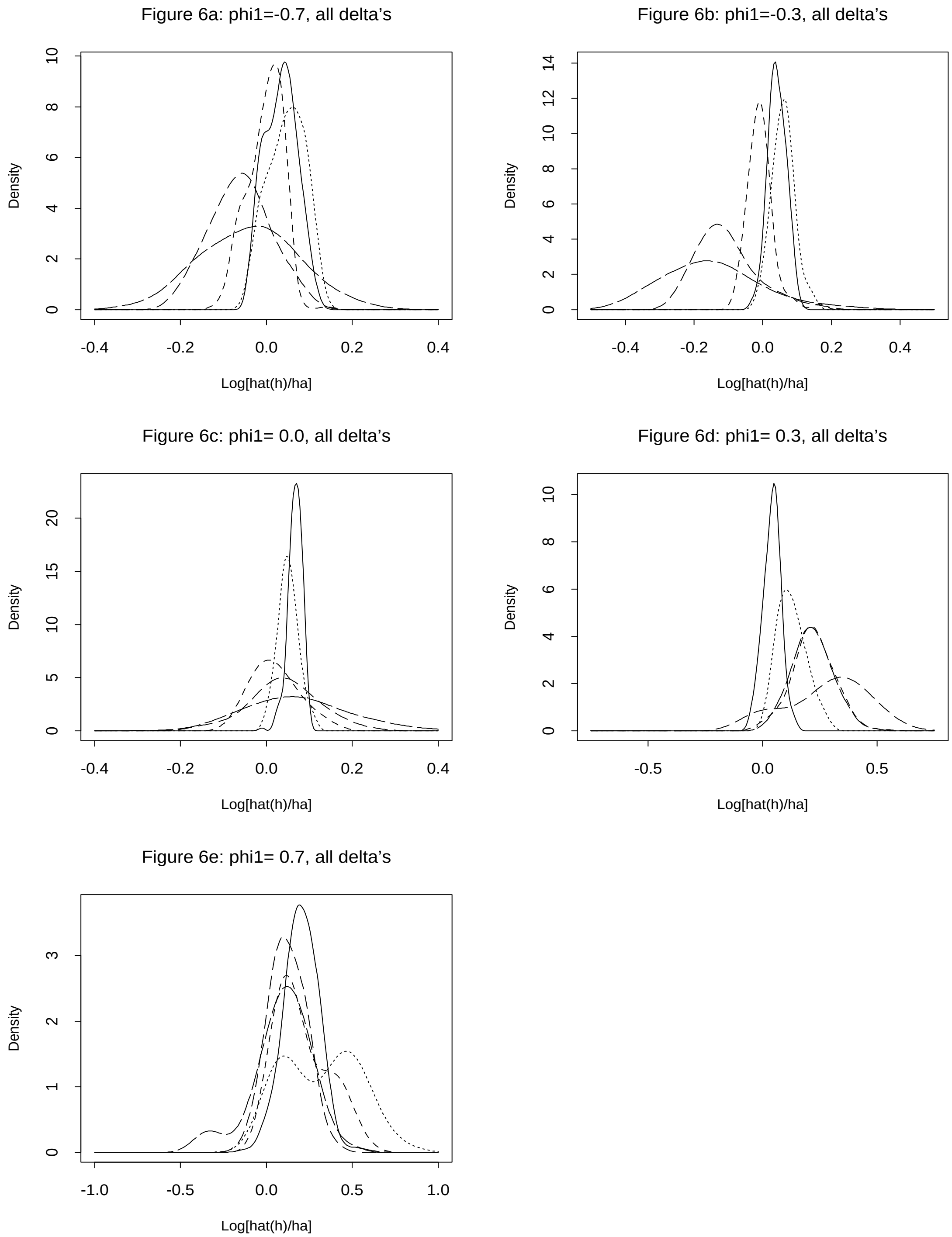

Figure 6: The same results as given in Figures 3a through 3e but for $m^0 = 1$.