MINIMUM HELLINGER DISTANCE ESTIMATION FOR LOCALLY STATIONARY PROCESSES

Souleymane Fofana
Ecole Nationale de la Statistique et de l'Analyse Economique (ENSAE), Sénégal
e-mail: *fof_sn@yahoo.fr*

Aliou Diop [1]
LERSTAD, Université Gaston Berger (UGB), Saint-Louis, Sénégal
e-mail: *aliou.diop@ugb.edu.sn*

Ouagnina Hili
Institut National Polytechnique Félix Houphouët-Boigny (INPHB), Yamoussoukro, Côte d'Ivoire
e-mail: *o_hili@yahoo.fr*

Key words: Hellinger distance, Locally stationary processes, Local periodogram, Stationarity, Time varying spectral density.

Abstract: In this paper, we are interested in the estimation of locally stationary processes by the minimum Hellinger distance estimator (Beran, 1977) in spectral framework. This distance is originally applied to probability distributions. Here we apply this distance to spectral density functions belonging to a specified parametric spectral family. We generalize the minimum Hellinger distance estimation method to processes that only show a locally stationary behaviour. Asymptotic properties of the estimator are shown. The robustness of the estimator is investigated through a simulation study. An application on real data is carried out.

1. Introduction

Stationarity is the basic assumption for a general asymptotic theory for identification, estimation and forecasting in time series analysis. However, in several situations, a nonstationary behaviour is observed, in practice, on many series. For example, Stărică and Granger (2005), dropped the usual assumption of global stationarity in the S&P 500 absolute returns, approximated locally nonstationary data generating process by stationary models, and identified the intervals on which stationary processes provide a good approximation. The method of estimation that they used was that of quasi-maximum likelihood.

[1] Corresponding author. Tel +221 77 553 37 70

AMS: 62F10, 62F12, 62M10

Suppose we observe data $X_1, \ldots, X_T$ from some locally stationary process (X_t) with time varying spectral density f belonging to a specified parametric spectral family $\mathscr{F} = \{f_\theta : \theta \in \Theta \subset \mathbb{R}^p\}$ with Θ a compact subset of $\mathbb{R}^p$. Our objective is to estimate θ. In order to estimate θ for stationary time series, Hosoya (1974) and Taniguchi (1979) proposed minimising $\int_{-\pi}^{\pi}\Big(log f_\theta(\lambda) + I_T(\lambda)/f_\theta(\lambda)\Big)d\lambda$ with respect to θ, where $I_T(\lambda)$ is the periodogram of the sample $X_1,...,X_T$ and $\lambda \in [-\pi, \pi]$. This function is introduced by Whittle (1953, 1954) and is an approximate Gaussian likelihood function. For a locally stationary process $X_{t,T}$, Dahlhaus (1997) generalizes Whittle's method based on the minimisation of this function, by replacing the periodogram $I_T(\lambda)$ in Whittle's function by a local version and integrating over time. As in the stationary case, he approximates the new version of Whittle's function by the exact Gaussian likelihood function and estimates the parameter θ by maximum likelihood estimation.

In fact, for many parametric families of distributions of interest in applications, the maximum likelihood estimator has full asymptotic efficiency among regular estimators. But, recently, it has been recognised that maximum likelihood estimators are not, in general, stable under small perturbations in the underlying model. However, an alternative solution to this problem can be found with Beran (1977). He introduced a new efficient parametric estimator which is intrinsically stable under small perturbations. He proposed as an estimator of θ the value $\hat{\theta}_T$ in the parameter space Θ which minimizes the Hellinger distance between f_θ and $\hat{f}_T$, where f_T, is a suitable non-parametric estimator of f. He has investigated the asymptotic properties of $\hat{\theta}_T$, showing that it is asymptotically efficient under $\mathscr{F}$ and is also minimax robust in a small Hellinger-metric neighbourhood of $\mathscr{F}$. For the robustness of the minimum Hellinger estimator, see also Hili (1999). Similarly, by using this estimator for finite mixture models, Cutler and Cordero-Brãna (1996) show that the minimum Hellinger distance can give sensible results when likelihood fails.

The problem of estimation of the parameter has been the subject of a plentiful literature. However the case currently studied is the temporal case based on the probability distributions, see Beran (1977), and Hili (1995, 1996). But the spectral case in presence of nonstationary time series has been rarely studied, although it is important for many applications (economics, finance, etc.). The first theoretical results are due, on the one hand, to Dahlhaus (1997) who proposed the Whittle estimator and, on the other hand, to Ludeña (2000) who proposed one method of estimation for stationary Gaussian long-range dependent processes based on the log-periodogram.

In this paper, we consider the minimum Hellinger distance estimators (MHD) method to estimate the processes that only show locally stationary behaviour in a spectral framework. The MHD estimator is introduced by Beran (1977), for independent samples, as a method for simultaneously achieving robustness and first-order efficiency, see also Tamura and Boos (1986), Simpson (1987, 1989), and Basu and Lindsay (1994). Hili (1995, 2003) extended this method of parametric estimation to the case of dependent samples. However, all these studies are realised in probabilistic approach. Here we use similar arguments to study the local stationary models in a spectral framework.

The paper is organised as follows. Section 2 introduces the notion of local stationarity in the sense of Dahlhaus (1997). Section 3 presents the minimum Hellinger distance estimation method as an approach applicable to locally stationary models. In this section, we estimate the parameters by the minimisation of the Hellinger distance where the usual kernel spectral estimator in global stationarity is replaced by local kernel spectral estimators. Section 4 contains a simulation study.

The robustness and the performance of the method are investigated in this section. Section 5 is used to illustrate our methods on real data. Section 6 provides a discussion of our methods.

2. Locally stationary processes

A triangular array[2] of stochastic processes $\{X_{t,T}\} = \{X_{t,T},\ t = 1,\ldots,T\}_{T\in\mathbb{N}}$ is called locally stationary if $X_{t,T}$ has time varying infinite order moving average representation

$$X_{t,T} = \sum_{j=-\infty}^{\infty} \alpha_{t,T}(j)\varepsilon_{t-j}, \tag{1}$$

where the ε_t are independent, identically distributed random variables with $\mathbb{E}\varepsilon_t = 0$, $\mathbb{E}\varepsilon_t^2 = 1$ and $\mathbb{E}\varepsilon_t^4 < \infty$ and the time varying coefficients $\alpha_{t,T}(j)$ satisfy the following smoothness conditions :

$\mathbf{H_1}$: there exists a sequence $\{\ell(j),\ j\in\mathbb{Z}\}$ satisfying

$$\sum_{j=-\infty}^{\infty} \frac{|j|}{\ell(j)} < \infty$$

such that

$$\sup_{1\le t\le T} |\alpha_{t,T}(j)| \le \frac{K}{\ell(j)}.$$

$\mathbf{H_2}$: there exists a function $\alpha(.,j):\ (0,1]\to\mathbb{R}$ satisfying, for the same $\ell(j)$ as in $\mathbf{H_1}$,

$$\sup_{0<u\le 1} |\alpha(u,j)| \le \frac{K}{\ell(j)},$$

$$\sup_{0<u,v\le 1} |\alpha(u,j)-\alpha(v,j)| \le \frac{K|u-v|}{\ell(j)}$$

and

$$\sup_{1\le t\le T} |\alpha_{t,T}(j)-\alpha(\frac{t}{T},j)| \le \frac{K}{T\ell(j)}.$$

K denotes an arbitrary positive constant which does not depend on T and which can vary from line to line.
Thus we can define the following stationary process $\tilde{X}_t(u)$ which is an approximation of $X_{t,T}$ in a local neighbourhood around $u = \frac{t}{T}$:

$$\tilde{X}_t(u) = \sum_{j=-\infty}^{\infty} \alpha(u,j)\varepsilon_{t-j}.$$

The time varying spectral density (or spectral evolutive density) of the locally stationary process $X_{t,T}$ at time $u\in[0,1]$ and frequency $\lambda\in[-\pi,\pi]$ is defined by

$$f(u,\lambda) = \frac{1}{2\pi}|A(u,\lambda)|^2, \tag{2}$$

[2] Whose rows correspond to different stochastic processes.

where $A(u,\lambda)=\sum_{j=-\infty}^{\infty}\alpha(u,j)\exp(-i\lambda j)$. The function $A(u,\lambda):[0,1]\times[-\pi,\pi]\to\mathscr{C}$ with $A(u,\lambda)=\overline{A(u,\lambda)}$ is assumed to be a smooth function. See Dahlhaus and Subba Rao (2006), and Sergides and Paparoditis (2007) for more details.

Consider the observations $X_{1,T},...,X_{T,T}$ from the local stationary process $(X_{t,T})_{t\in Z}$ defined in (1). In inferring properties of the underlying locally stationary process, a useful class of statistics is obtained as functional of the local periodogram. The local periodogram is the periodogram of a segment of length N of consecutive observations around a time point $[uT]$, $u\in(0,1)$ and is defined by

$$I_N(u,\lambda)=\frac{1}{2\pi N}\left|\sum_{s=0}^{N-1}X_{[uT]-\frac{N}{2}+s+1,T}e^{-i\lambda s}\right|^2, \quad -\pi\leq\lambda\leq\pi.$$

I_N is calculated over segments of length N with midpoints $t_j=S(j-1)+N/2$, $j=1,\ldots,M$, where $T=S(M-1)+N$, S is the shift from segment to segment and M is the number of segments. Specifically S is an integer number and N a multiple of S. Following Dahlhaus (1997), the shift S should, in general, be as small as possible; so that the theoretical results hold even for $S=1$ and N, S and T fulfil the relation $\frac{TS^4}{N^4}\to 0$.

One of the difficulties of nonstationary time series modelling is the development of a satisfactory asymptotic theory, which is needed in time series to draw statistical inference using finite size samples. If $X_1,...,X_T$ is an arbitrary finite segment of a nonstationary time series, then letting T tend to infinity, information may be lost at the beginning or middle of the time series. Therefore, a different type of asymptotic setup is needed for nonstationary time series.

As in the theory of nonparametric regression, it seems natural to develop the asymptotic theory for nonstationary time series by developing inference procedures over a finite grid. In the following sections, we adapt the work of Beran (1977) to develop parameter estimation techniques and an asymptotic theory for stationary processes in the time case to locally stationary processes in the frequency case. The interest for this method of parametric estimation is that the MHD estimation method gives efficient and robust estimators (see Beran, 1977; Hili, 1999). Thus the following section is devoted to the generalisation of Hellinger distance estimator method for stationary processes to locally stationary processes.

3. The minimum Hellinger distance estimator under local stationarity

In this section we discuss the behaviour of the minimum Hellinger distance in local stationarity. Consider a sample $X_{1,T},...,X_{T,T}$ from a locally stationary process $(X_{t,T})$ defined in (1). To apply the MHD to the local stationary process, we use the local version of the kernel density which is an approach to the kernel density in global stationarity.

3.1. Defining the estimator

We define the generalisation of the minimum Hellinger distance to a locally stationary process. We replace the usual nonparametric density estimator by a local version over data segments (possibly overlapping segment) and we integrate over time. We set

$$d_T^{HD}(f_\theta, f_N) = \frac{1}{M}\sum_{j=1}^{M}\left|\left|f_\theta^{1/2}(u_j,.) - f_N^{1/2}(u_j,.)\right|\right|_2,$$

that is, the distance between f_θ and f_N is the average of the distances over time between $f_\theta^{1/2}(u_j,.)$ and $f_N^{1/2}(u_j,.)$, where $f_\theta(u_j,.)$ and $f_N(u_j,.)$ are respectively the local parametric spectral density and the local nonparametric spectral density estimator of the process in the jth segment and $||\cdot||_2$ denotes the L_2 norm, e.g., $||f||_2 = \left(\int_{-\pi}^{\pi}|f(\lambda)|^2 d\lambda\right)^{1/2}$. In this definition, M is the number of segments depending on N as defined in previous section.

The value (or values) $\hat{\theta}_T^{HD}$, based on $X_{1,T}, X_{2,T}, ..., X_{T,T}$, is (or are) such that

$$\hat{\theta}_T^{HD} = arg\min_{\theta\in\Theta} d_T^{HD}(f_\theta, f_N). \tag{3}$$

Consider observations $X_{1,T}, ..., X_{T,T}$ from the nonstationary process $(X_{t,T})_{t\in Z}$ defined in (1). Let $f(u,\lambda)$ be the true spectral density of the process $(X_{t,T})_{t\in Z}$ and $f_\theta(u,\lambda) \in \mathscr{F}$ the local spectral density of the model chosen. We view

$$d^{HD}(f_\theta, f) = \int_0^1 \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|f_\theta^{1/2}(u,\lambda) - f^{1/2}(u,\lambda)\right|^2 d\lambda\right)^{1/2} du,$$

as the distance between the true process with spectral density $f(u,\lambda)$ and the model with spectral density $f_\theta(u,\lambda)$, $\theta \in \Theta$, a compact subset of $\mathbb{R}^p$. The best approximating parameter value from our model class then is

$$\theta_0 = arg\min_{\theta\in\Theta} d^{HD}\big(f_\theta, f\big).$$

If the model is correct, that is, $f = f_{\theta^*}$, then it is easy to show that $\theta_0 = \theta^*$.

The function $d_T^{HD}(f_\theta, f_N)$ is now obtained from $d^{HD}(f_\theta, f)$ by replacing the integral over time by $(1/M)\sum_j$ and replacing the unknown true spectral density f by the nonparametric estimate f_N. Thus, we suppose that $d_T^{HD}(f_\theta, f_N)$ is an approximation to the exact distance $d^{HD}(f_\theta, f)$ (as in the case of the minimum Whittle distance estimator, Dahlhaus, 1997).

We now prove under certain regularity conditions that $\hat{\theta}_T^{HD}$ converges to θ_0. To simplify the notation, we set in the sequel $d_T^{HD}(\theta) = d_T^{HD}(f_\theta, f_N)$ and $d^{HD}(\theta) = d^{HD}(f_\theta, f)$.

For $f_N(u,\lambda)$ the local nonparametric spectral density estimator of the process, we consider here Dahlhaus's (1997) approach by using the kernel estimator. In this case $f_N(u,\lambda)$, called the local kernel spectral density, is given by

$$f_N(u,\lambda) = \frac{1}{Nh_N}\sum_{k=-N_1}^{N_1} K\left(\frac{\lambda-\lambda_k}{h_N}\right) I_N(u,\lambda_k). \tag{4}$$

N_1 denotes the largest integer less than or equal to $\frac{N}{2}$. The discrete frequencies λ_k are given by

$$\frac{2\pi k}{N}, \qquad -N_1 \le k \le N_1.$$

The asymptotic properties of the nonparametric kernel spectral estimator for the local stationary processes have been studied by, among others, Dahlhaus (1996), Theorem 2.2 and Sergides and Paparoditis (2007). We will need the following assumptions:

$\mathbf{H_3}$: $K(\cdot)$ is a symmetric and non-negative function on $\mathbb{R}$ such that

$$\int_{-\infty}^{\infty} K(u)du = 1, \qquad \int_{-\infty}^{\infty} u^2 K(u)du = 1 \text{ and } \sup_x |x||K(x)| < \infty$$

where K has compact support $[-c, c]$ and K is uniformly Lipschitz with constant L_K i.e. for all $x, y \in \mathbb{R}, |K(x) - K(y)| \leq L_K|x - y|$.

$\mathbf{H_4}$: The smoothing bandwidth h_N satisfies $h_N \to 0$ such that $Nh_N^3 \to \infty$ and $Nh_N^4 \to 0$ as $N \to \infty$.

We remark that the condition $Nh_N^3 \to \infty$ is classical in nonparametric estimation while $Nh_N^4 \to 0$ is the same as in Beran (1977).

3.2. Asymptotic properties of the MHD estimator

The asymptotic distribution of the MHD estimator is important for making theoretical comparisons with other estimators and for making valid approximated inferences. We show here the consistency and asymptotic normality of the MHD estimator $\widehat{\theta}_T^{HD}$ defined in (3) and we discuss the robustness of the estimator.

Let $\mathscr{G}$ be the set of all spectral densities with respect to the Lebesgue measure on the real line. The functional I is defined on $\mathscr{G}$ by the requirement that, for any $f \in \mathscr{G}$,

$$||f_{I(f)}^{1/2} - f^{1/2}||_2 = \min_{t \in \Theta} ||f_t^{1/2} - f^{1/2}||_2.$$

Following the approach of Beran (1977), we view the minimum Hellinger distance estimator (MHDE) of θ as the value at f_N of the functional I, where for any $f \in \mathscr{G}$, $I(f)$ is defined by

$$I(f) = \{\theta \in \Theta : ||f_\theta^{1/2} - f^{1/2}||_2 = \min_{t \in \Theta} ||f_t^{1/2} - f^{1/2}||\}. \tag{5}$$

$I(f)$ may be multiple values and we use the notation $I(f)$ to indicate any one of the possible values, chosen arbitrarily. As in Theorem 1 of Beran (1977), to ensure existence of $I(f)$, continuity of I and uniqueness of $I(f_\theta) = \theta$, the following assumptions are needed on the parametric family $\{f_\theta,\ \theta \in \Theta\}$:

- Θ is compact subset of $\mathbb{R}^p$
- $\theta_1 \neq \theta_2$ implies $f_{\theta_1} \neq f_{\theta_2}$ on a set of a positive Lebesgue measure
- for almost every u and λ, $f_\theta(u, \lambda)$ is continuous in θ.

In the sequel, all limits are taken as $T \to \infty$, unless otherwise specified.

3.2.1. Consistency

Suppose that our data X consists of a sample with spectral density f. The consistency of the minimum Hellinger distance estimator follows from the continuity of I. The result is formally stated in the following theorem.

Theorem 1 Suppose that $f \in \mathscr{G}$ is the true spectral density function of $X_{t,T}$ defined in (1), differentiable in u and λ, that $f_\theta(u,\lambda)$ is continuous in θ for each u and λ and also that $I(f)$ is unique. Then, under the assumptions $H_1 - H_4$,

$$\hat{\theta}_T^{HD} \rightarrow I(f) \quad \text{in probability.}$$

In particular, if $f = f_{\theta_0}$, then $\hat{\theta}_T^{HD} \rightarrow \theta_0$ in probability.

Proof. To prove the consistency of $\hat{\theta}_T^{HD}$, we start by proving that

$$\sup_{\theta\in\Theta} |d_T^{HD}(\theta) - d^{HD}(\theta)| \rightarrow 0 \qquad \text{in probability.}$$

Indeed

$$\big|d_T^{HD}(\theta) - d^{HD}(\theta)\big| = \Big|\frac{1}{M}\sum_{j=1}^{M} \big|\big|f_\theta^{1/2}(u_j,.) - f_N^{1/2}(u_j,.)\big|\big| - \int_0^1 \big|\big|f_\theta^{1/2}(u,.) - f^{1/2}(u,.)\big|\big| du\Big|.$$

Clearly, as $M \rightarrow +\infty$, the sum and integral here should converge to the same limit, i.e.

$$\frac{1}{M}\sum_{j=1}^{M} \big|\big|f_\theta^{1/2}(u_j,.) - f_N^{1/2}(u_j,.)\big|\big| \sim \int_0^1 \big|\big|f_\theta^{1/2}(u,.) - f^{1/2}(u,.)\big|\big| du.$$

In fact,

$$\begin{aligned}\big|d_T^{HD}(\theta) - d^{HD}(\theta)\big| \;\leq\; & \Big|\frac{1}{M}\sum_{j=1}^{M}\Big(\big|\big|f_\theta^{1/2}(u_j,.) - f_N^{1/2}(u_j,.)\big|\big| - \big|\big|f_\theta^{1/2}(u_j,.) - f^{1/2}(u_j,.)\big|\big|\Big)\Big| \\ & + \Big|\frac{1}{M}\sum_{j=1}^{M} \big|\big|f_\theta^{1/2}(u_j,.) - f^{1/2}(u_j,.)\big|\big| - \int_0^1 \big|\big|f_\theta^{1/2}(u,.) - f^{1/2}(u,.)\big|\big| du\Big|.\end{aligned}$$

It is obvious that the second term of the above inequality tends to 0 as $M \rightarrow \infty$. Following the triangle inequality, the first term of the right-hand side is less than

$$\Big|\frac{1}{M}\sum_{j=1}^{M} \big|\big|f^{1/2}(u_j,.) - f_N^{1/2}(u_j,.)\big|\big|\Big|$$

which goes to zero as $N \rightarrow \infty$ in probability. Then

$$\sup_{\theta} \big|d_T^{HD}(\theta) - d^{HD}(\theta)\big| \rightarrow 0 \quad \text{in probability.} \tag{6}$$

Since $d^{HD}(\theta)$ and $d_T^{HD}(\theta)$ are respectively minimised by θ_0 and $\hat{\theta}_T^{HD}$, we have

$$d^{HD}(\theta_0) \leq d^{HD}(\hat{\theta}_T^{HD}) \quad \text{and} \quad d_T^{HD}(\hat{\theta}_T^{HD}) \leq d_T^{HD}(\theta_0).$$

The convergence in (6) implies $d^{HD}(\hat{\theta}_T^{HD}) \rightarrow d^{HD}(\theta_0)$ and therefore

$$\hat{\theta}_T^{HD} \rightarrow \theta_0 \qquad \text{in probability.}$$

■

3.2.2. Asymptotic Normality

In this section, we examine the large sample behaviour of $I(f_N)$ i.e $\hat{\theta}_T^{HD}$, where I is the functional introduced above in (5) and f_N is a suitable spectral density estimator. In order to establish the asymptotic normality of the appropriately normalised estimator $\hat{\theta}_T^{HD}$, as in Beran (1977), we impose some smoothness conditions on the model. We need to make the following assumption for $s_\theta = f_\theta^{1/2}$. Suppose for $\theta \in \Theta^0 \subset \mathbb{R}^p$ (Θ^0 interior of Θ) that s_θ is twice differentiable in L_2; that is, suppose that there exist $\dot{s}_\theta(p\times 1)$, the vector of first partial derivative with components in L_2, and $\ddot{s}_\theta(p\times p) \in L_2$, the matrix of second partial derivative with components in L_2 with respect to θ and which satisfy for every β in a neighbourhood of zero,

$$s_{\theta+\beta}(u,\lambda) = s_\theta(u,\lambda) + \beta \dot{s}_\theta(u,\lambda) + \beta u_\beta(u,\lambda) \tag{7}$$

and

$$\dot{s}_{\theta+\beta}(u,\lambda) = \dot{s}_\theta(u,\lambda) + \beta \ddot{s}_\theta(u,\lambda) + \beta v_\beta(u,\lambda), \tag{8}$$

where u_β and v_β tend to zero in L_2 as $\beta \to 0$.

Here we extend the results of Hosoya and Taniguchi (1982, p.150) for stationary processes to the framework of locally stationary processes. Hence the proof is omitted.

Lemma 1 Suppose $X_{1,T},\ldots,X_{T,T}$ are realisations of locally stationary process, Assumption $\mathbf{H}_1 - \mathbf{H}_2$ are fulfilled and $f(u,\lambda)$ is the Lipschitz continuous function in u. For a 2π-periodic function $\psi : [0,1]\times[-\pi,\pi] \to \mathbb{C}$,

$$J_T = \sqrt{T}\int_0^1\int_{-\pi}^{\pi} \psi(u,\lambda)\Big(I_N(u,\lambda) - f(u,\lambda)\Big)d\lambda du$$

has, asymptotically, a normal distribution with zero mean vector and covariance matrix

$$V = 2\pi\int_0^1\int_{-\pi}^{\pi} \psi(u,-\lambda)\psi'(u,\lambda)f^2(u,\lambda)d\lambda du + 2\pi\int_0^1\int_{-\pi}^{\pi} \psi^2(u,-\lambda)f^2(u,\lambda)d\lambda du.$$

The following theorem is an extension of the results stated in Theorem 2 of Beran's paper to the case of locally stationary processes.

Theorem 2 Suppose that (7) and (8) hold for $\theta \in \Theta^0$, for $f \in \mathscr{G}$,
$I(f)$ exists, is unique, and $I(f) \in \Theta^0$. Also, suppose that $\int_0^1\int_{-\pi}^{\pi} \ddot{s}_{I(f)}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du$ is a nonsingular matrix, and the functional I is continuous at f in the Hellinger topology. Then for every sequence of spectral densities $\{f_N\}$ converging to f in the Hellinger metric,

$$\begin{aligned} I(f_N) \;=\; & I(f) + \frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi} \rho_f(u_j,\lambda)f_N^{1/2}(u_j,\lambda)d\lambda - \int_0^1\int_{-\pi}^{\pi} \rho_f(u,\lambda)f^{1/2}(u,\lambda)d\lambda du \\ & + a_N\Big(\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi} \dot{s}_{I(f)}(u_j,\lambda)f_N^{1/2}(u_j,\lambda)d\lambda - \int_0^1\int_{-\pi}^{\pi} \dot{s}_{I(f)}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big), \end{aligned} \tag{9}$$

where

$$\rho_f(u,\lambda) = -\Big[\int_0^1\int_{-\pi}^{\pi} \ddot{s}_{I(f)}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big]^{-1}\dot{s}_{I(f)}(u,\lambda)$$

and a_N is a real $p \times p$ matrix which tends to zero as $N \to \infty$.
In particular, for $f = f_\theta$,

$$\rho_{f_\theta}(u,\lambda) = \Big[\int_0^1 \int_{-\pi}^{\pi} \dot{s}_\theta(u,\lambda)\dot{s}'_\theta(u,\lambda)d\lambda du \Big]^{-1} \dot{s}_\theta(u,\lambda). \tag{10}$$

Proof. We have, following the proof of Theorem 1, the sum $\frac{1}{M}\sum_{j=1}^{M} ||f_\theta^{1/2}(u_j,.) - f_N^{1/2}(u_j,.)||$ and the integral $\int_0^1 ||f_\theta^{1/2}(u,.) - f^{1/2}(u,.)||du$ converge to the same limit.

Thus, let $\theta_0 = I(f), \quad \theta_N = I(f_N)$. Since $\theta_0 = I(f) \in \Theta^0$ maximise $\int_0^1 \int_{-\pi}^{\pi} s_\theta(u,\lambda) f^{1/2}(u,\lambda) d\lambda du$ and since from (7)

$$\lim_{\beta \to 0} \beta^{-1} \int_0^1 \int_{-\pi}^{\pi} \big(s_{\theta+\beta}(u,\lambda) - s_\theta(u,\lambda)\big) f^{1/2}(u,\lambda) d\lambda du = \int_0^1 \int_{-\pi}^{\pi} \dot{s}_\theta(u,\lambda) f^{1/2}(u,\lambda) d\lambda du$$

for every $\theta \in \Theta^0$, it follows that

$$\int_0^1 \int_{-\pi}^{\pi} \dot{s}_{\theta_0}(u,\lambda) f^{1/2}(u,\lambda) d\lambda du = 0.$$

A similar conclusion applies to $\dot{s}_{\theta_N}$. Then, using (8)

$$\begin{aligned} 0 &= \frac{1}{M}\sum_{j=1}^{M} \int_{-\pi}^{\pi} \dot{s}_{\theta_N}(u_j,\lambda) f_N^{1/2}(u_j,\lambda) d\lambda \\ &= \frac{1}{M}\sum_{j=1}^{M} \int_{-\pi}^{\pi} \big(\dot{s}_{\theta_0}(u_j,\lambda) + \beta \ddot{s}_{\theta_0}(u_j,\lambda) + \beta v_\beta(u_j,\lambda)\big) f_N^{1/2}(u_j,\lambda) d\lambda. \end{aligned}$$

Set $\beta = \theta_N - \theta_0$,

$$0 = \frac{1}{M}\sum_{j=1}^{M} \int_{-\pi}^{\pi} \big(\dot{s}_{\theta_0}(u_j,\lambda) + (\theta_N - \theta_0)\ddot{s}_{\theta_0}(u_j,\lambda) + (\theta_N - \theta_0) v_N(u_j,\lambda)\big) f_N^{1/2}(u_j,\lambda) d\lambda,$$

where v_N tends to zero in L_2 as $\beta \to 0$ since $\theta_N \to \theta_0$.

$$0 = \frac{1}{M}\sum_{j=1}^{M} \Big(\int_{-\pi}^{\pi} \dot{s}_{\theta_0}(u_j,\lambda) f_N^{1/2}(u_j,\lambda) d\lambda + \int_{-\pi}^{\pi} \big(\ddot{s}_{\theta_0}(u_j,\lambda) + v_N(u_j,\lambda)\big)(\theta_N - \theta_0) f_N^{1/2}(u_j,\lambda) d\lambda \Big).$$

Thus, for N sufficiently large,

$$\theta_N - \theta_0 = -\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\dot{s}_{\theta_0}(u_j,\lambda)f_N^{1/2}(u_j,\lambda)d\lambda \times \Big[\Big(\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\big(\ddot{s}_{\theta_0}(u_j,\lambda)+v_N(u_j,\lambda)\big)f_N^{1/2}(u_j,\lambda)d\lambda\Big)^{-1}$$
$$-\Big(\int_0^1\int_{-\pi}^{\pi}\dot{s}_{I(f)}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big)^{-1}+\Big(\int_0^1\int_{-\pi}^{\pi}\dot{s}_{I(f)}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big)^{-1}\Big]$$
$$= -\Big(\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\dot{s}_{\theta_0}(u_j,\lambda)f_N^{1/2}(u_j,\lambda)d\lambda - \int_0^1\int_{-\pi}^{\pi}\dot{s}_{\theta_0}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big)$$
$$\times\Big(\int_0^1\int_{-\pi}^{\pi}\dot{s}_{I(f)}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big)^{-1}$$
$$+\Big(\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\dot{s}_{\theta_0}(u_j,\lambda)f_N^{1/2}(u_j,\lambda)d\lambda - \int_0^1\int_{-\pi}^{\pi}\dot{s}_{\theta_0}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big)$$
$$\times\Big[\Big(\int_0^1\int_{-\pi}^{\pi}\dot{s}_{I(f)}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big)^{-1}-\Big(\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\big(\ddot{s}_{\theta_0}(u_j,\lambda)+v_N(u_j,\lambda)\big)f_N^{1/2}(u_j,\lambda)d\lambda\Big)^{-1}\Big].$$

If $f_N \to f$, we have $\theta_N = I(f_N) \to \theta_0 = I(f)$ and $\ddot{s}_\theta$ is continuous in θ.
Thus, we have

$$\theta_N = \theta_0 + \frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\rho_f(u_j,\lambda)f_N^{1/2}(u_j,\lambda)d\lambda - \int_0^1\int_{-\pi}^{\pi}\rho_f(u,\lambda)f^{1/2}(u,\lambda)d\lambda du$$
$$+a_N\Big(\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\dot{s}_{\theta_0}(u_j,\lambda)f_N^{1/2}(u_j,\lambda)d\lambda - \int_0^1\int_{-\pi}^{\pi}\dot{s}_{\theta_0}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big),$$

where

$$\rho_f(u,\lambda) = -\Big[\int_0^1\int_{-\pi}^{\pi}\ddot{s}_{\theta_0}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big]^{-1}\dot{s}_{\theta_0}(u,\lambda)$$

and

$$a_N = \Big(\int_0^1\int_{-\pi}^{\pi}\ddot{s}_{I(f)}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big)^{-1}-\Big(\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\big(\ddot{s}_{\theta_0}(u_j,\lambda)+v_N(u_j,\lambda)\big)f_N^{1/2}(u_j,\lambda)d\lambda\Big)^{-1}.$$

which goes to zero as $N \to \infty$. ■

Using the representation (9), we now show by the next theorem, under stronger assumptions, that the MHD estimator $\hat{\theta}_T^{HD} = I(f_N)$ has an asymptotically normal distribution. Its proof follows the main lines of the proof of Theorem 4 in Beran (1977). Hence, it is omitted.

Theorem 3 Suppose $f \in \mathscr{G}$ is the true spectral density of $(X_t)_{t\in\mathbb{Z}}$ defined in (1). Assume that (7) and (8) hold and ρ has compact support $\mathscr{C} = [0,1]\times[-\pi,\pi]$ on which it is continuous. Suppose f is twice absolutely continuous and f'' with respect to λ is bounded. Suppose $I(f)$ exists, is unique and $I(f) \in \Theta^0$, $\int \ddot{s}_{I(f)}(u_j,\lambda)f^{1/2}(u_j,\lambda)d\lambda > 0$ for every $j = 1,...,M$, and $\dot{s}_{I(f)} \in L_2$. Then, under assumption H_2 and the hypotheses of Theorem 1,

$$\sqrt{T}(I(f_N) - I(f)) \to \mathscr{N}(0,V)$$

where

$$V = \frac{\pi}{2}\int_0^1\int_{-\pi}^{\pi}\rho(u,-\lambda)\rho'(u,\lambda)f(u,\lambda)d\lambda du + \frac{\pi}{2}\int_0^1\int_{-\pi}^{\pi}\rho^2(u,-\lambda)f(u,\lambda)d\lambda du$$

with

$$\rho(u,\lambda) = -\Big[\int_0^1\int_{-\pi}^{\pi}\ddot{s}_{I(f)}(u,\lambda)f^{1/2}(u,\lambda)d\lambda du\Big]^{-1}\dot{s}_{I(f)}(u,\lambda).$$

In particular, if $f = f_{\theta_0}$, then

$$\sqrt{T}(\hat{\theta}_T^{HD} - \theta_0) \to \mathcal{N}(0,V)$$

where

$$V = \frac{\pi}{2}\int_0^1\int_{-\pi}^{\pi}\big[\dot{s}_{\theta_0}(u,-\lambda)\dot{s}'_{\theta_0}(u,\lambda)\big]^{-1}s_{\theta_0}(u,\lambda)d\lambda du + \frac{\pi}{2}\int_0^1\int_{-\pi}^{\pi}\big[\dot{s}^2_{\theta_0}(u,-\lambda)\big]^{-1}s_{\theta_0}(u,\lambda)d\lambda du.$$

3.3. Robustness

The robustness of an estimator is often studied by means of the influence function, IF (Hampel, 1974), defined by

$$IF(x,f) = \lim_{t\to 0} t^{-1}\big(I((1-t)f + t\delta_x) - I(f)\big),$$

where δ_x is unit mass at x, $I(f)$ the functional based on f is defined in (5) and f the true spectral density of (X_t). This function characterizes local robustness and the breakdown point, which relates to global robustness. Following Lindsay (1994), the influence function is a very misleading robustness measure for minimum Hellinger distance estimation. An alternative measure of robustness is the α-influence function, (Beran, 1977). Let $f_{\theta,\alpha}(x) = (1-\alpha)f_\theta(x) + \alpha\delta_{[0,1]}(x)$, where $\delta_{[0,1]}(x)$ denotes the uniform density on the interval $[0,1]$, $\theta \in \Theta$ and $\alpha \in [0,1]$. The α-influence function, denoted here by α-*IF* is defined as

$$\alpha\text{-}IF = \alpha^{-1}\big(I(f_{\theta,\alpha}) - \theta\big),$$

where $\theta = I(f_\theta)$. The influence function is obtained by taking the limit as $\alpha \to 0$. It was shown in Beran (1977) that in order to assess the robustness of a functional with respect to the gross-error model it is necessary to examine the α-influence curve rather than the influence curve except when the influence curve provides a uniform approximation to the α-influence curve. Analytical evaluation of α-*IF* is very difficult, so we calculated the α-*IF* numerically.

4. Simulation study

In this section, we illustrate the finite-sample performance of the proposed estimator. We present simulation studies to a time varying autoregressive model which is obviously locally stationary. Let $X_{t,T}$ be a solution of the system of difference equations

$$\sum_{j=0}^{p} a_j\Big(\frac{t}{T}\Big)X_{t-j,T} = \sigma\Big(\frac{t}{T}\Big)\varepsilon_t \qquad \text{for} \quad t \in \mathbb{Z}.$$

The objective is to investigate the accuracy of the proposed estimator and the quality of the Gaussian approximation of its asymptotic distribution. We also compare the MHD estimator to alternative approaches.

4.1. Consistency

We suppose here that $a_0(u) = 1$ and (ε_t) is a sequence of independent random variables with mean zero and variance 1. Suppose also that $a_\theta(u) = \left(a_1^\theta(u), ..., a_p^\theta(u)\right)$ and $\sigma_\theta(u)$ depend on a finite dimensional parameter.

We now carry out some simulations to establish the robustness of the estimation of the parameter θ using MHD estimator, for finite samples. We present here a simulation example for the estimate $\hat{\theta}_T^{HD}$ from a locally stationary AR(2)-process

$$X_{t,T} + a_1(\frac{t}{T})X_{t-1,T} + a_2(\frac{t}{T})X_{t-2,T} = \sigma(\frac{t}{T})\varepsilon_t \tag{11}$$

where the ε_t are Gaussian random variables with mean zero and variance 1. The autoregressive parameters $a_i = a_i(\frac{t}{T})$, $i = 1,2$ are functions which change over time. As parameters we choose as in Dahlhaus (1997), $a_1(u) = -1.8\cos(1.5 - \cos 4\pi u)$, i.e. $a_1(u)$ is time varying, $a_2(u) = 0.832$ for all $0 \le u \le 1$, i.e. $a_2(u)$ is constant over time, $\sigma(u) = 1$. We generate $T = 128$ observations from the model (11). This choice is motivated by the desire to study the behaviour of the estimator for moderate or small sample sizes.

Following Adak (1998), $X_{t,T}$ is locally stationary and its time-dependent spectral density as given at (2) is

$$f(u,\lambda) = \frac{\sigma^2(u)}{2\pi}\left|1 + \sum_{j=1}^{2} a_j(u)\exp(i\lambda j)\right|^{-2} = f_{\theta(u)}(\lambda).$$

Now we estimate, from the data generated by (11), the parameter $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ such that $a_1(u) = -\theta_1\cos(\theta_2 - \cos(4\pi u))$, $a_2(u) = \theta_3$ and $\sigma^2(u) = \theta_4$. We could also choose to model the coefficients as polynomials with different orders as in Dahlhaus (1997).

The local kernel spectral density estimator defined by (4) was used. To compute the periodogram, we choose $S = 2$, $N = 16$ (i.e., $M = 57$). For the choice of the kernel function in f_N, we use the density of the $\mathcal{N}(0,1)$ distribution. For the choice of the bandwidth h, we select the value of h which minimizes the Mean Integrated Square Error ($MISE$) defined as follows

$$MISE(h) = \int_0^1 \int_{-\pi}^{\pi} \mathbb{E}\left(f_N(u,\lambda) - f(u,\lambda)\right)^2 d\lambda du.$$

The measure $MISE(h)$ depends on the unknown function f. We are going to estimate it by adopting the cross-validation rule proposed by Rudemo (1982) and Bowman (1984). In fact, consider the Integrated square error (ISE) defined by

$$ISE(h) = \int_0^1 \int_{-\pi}^{\pi} \left(f_N(u,\lambda) - f(u,\lambda)\right)^2 d\lambda du.$$

Grégoire (1993) approximate the $ISE(h)$ by the $CV(h)$ criteria, defined by

$$CV(h) = CV_1(h) + \int_0^1 \int_0^{2\pi} f^2(u,\lambda) d\lambda du$$

where

$$CV_1(h) = \int_0^1 \int_0^{2\pi} f_N^2(u,\lambda) d\lambda du - \frac{2}{\overline{N}} \sum_{j=1}^{\overline{N}} \int_0^1 f_N^j(u,\lambda_j) I_N(u,\lambda_j) du.$$

Remark

In general, the cross validation rule may ruin the efficiency of the MHDE in practical implementation. In Leung, Marriott and Wu (1993), simulation results showed that the conventional cross-validation rule, equivalent here to

$$CV(h) \approx ISE(h) = \int_0^1 \int_{-\pi}^{\pi} (f_N(u,\lambda) - f(u,\lambda))^2 \, d\lambda du,$$

behaves unsatisfactorily when the data is contaminated with outliers. To solve this problem, a robust cross-validation rule can be defined as

$$CV(h) = \int_0^1 \int_{-\pi}^{\pi} (f_N(u,\lambda) - f(u,\lambda))^2 \, \rho(u,\lambda) d\lambda du$$

with $\rho(\cdot)$ is a weight function that we assume known and null outside $[0,1] \times [0,2\pi]$.

To select the spectral bandwidth h, we use the robust cross-validation (CV) rule, defined by :

$$CV(h) = CV_1(h) + \int_0^1 \int_0^{2\pi} f^2(u,\lambda)\rho(u,\lambda) d\lambda du$$

where

$$CV_1(h) = \int_0^1 \int_0^{2\pi} f_N^2(u,\lambda)\rho(u,\lambda) d\lambda du - \frac{2}{\overline{N}} \sum_{j=1}^{\overline{N}} \int_0^1 f_N^j(u,\lambda_j) I_N(u,\lambda_j)\rho(u,\lambda) du$$

with

$$\rho(u,\lambda) = \begin{cases} \frac{1}{2\pi} & \text{on} \quad [0,1] \times [0,2\pi] \\ \\ 0 & \text{otherwise} \end{cases}$$

where

$$\lambda_j = \frac{2\pi j}{\overline{N}}, \quad \overline{N} = \left[\frac{N-1}{2}\right] \quad \text{and} \quad f_N^j(u,\lambda) = \frac{1}{Nh_N} \sum_{k=-N_1}^{N_1} K\Big(\frac{\lambda - \lambda_k}{h_N}\Big) I_N^j(u,\lambda_k)$$

with

$$I_N^j(u,\lambda) = \begin{cases} I_N(u,\lambda) & \text{if } (u,\lambda) \notin [0,1] \times]\lambda_{j-1}, \lambda_{j+1}[\\ a(u,\lambda) I_N(0,\lambda_{j-1}) + b(u,\lambda) I_N(1,\lambda_{j-1}) + & \\ c(u,\lambda) I_N(0,\lambda_{j+1}) + d(u,\lambda) I_N(1,\lambda_{j+1}) & \text{otherwise.} \end{cases}$$

$$a(u,\lambda) = \alpha\beta, \; b(u,\lambda) = (1-\alpha)\beta, \; c(u,\lambda) = \alpha(1-\beta) \quad \text{and} \quad d(u,\lambda) = (1-\alpha)(1-\beta)$$

where

$$\alpha = \frac{u-1}{0-1} \quad \text{and} \quad \beta = \frac{\lambda - \lambda_{j+1}}{\lambda_{j-1} - \lambda_{j+1}}.$$

In this way, the spectral bandwidth will be chosen at the point $\hat{h}$ minimizing the criterion CV :

$$\hat{h} = arg \min_h CV(h) = arg \min_h CV_1(h).$$

The results with simulated data are presented in Figures 1 – 2.

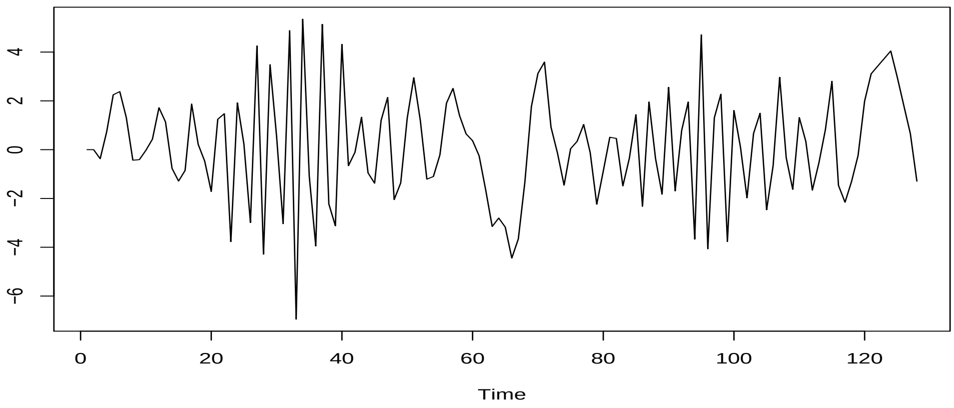

Figure 1: $T = 128$ realisations of a time varying AR-model.

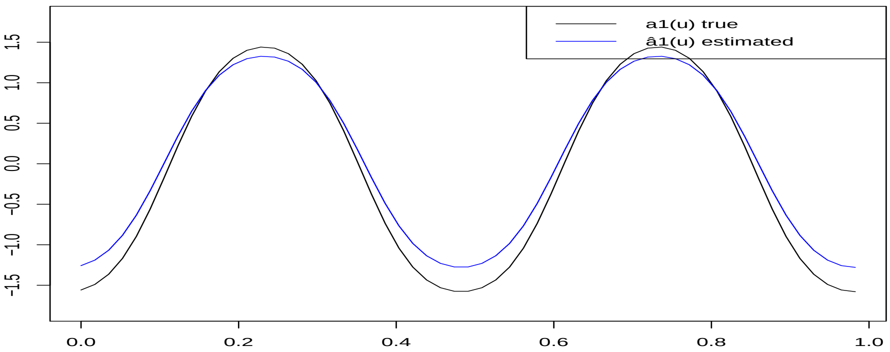

Figure 2: True and estimated time varying coefficient $a_1(u)$.

From an application of the Monte Carlo method with 3000 replications, we obtain the estimate of the first AR parameter as shown in Figure 2, the estimate of the second AR parameter as 0.819, and the estimate of the $\sigma(u)$ as 1.28.

4.2. Evaluating Gaussian approximation

In this simulation, we present normal plots for the estimators for simulated data from TV-AR(1) process that has the following form

$$X_{t,T} + a_1\left(\frac{t}{T}\right) X_{t-1,T} = \sigma\left(\frac{t}{T}\right) \varepsilon_t \tag{12}$$

where the ε_t are Gaussian random variables with mean zero and variance 1 and where Time Varying parameters have the following expressions : $a_1(u) = 2\cos(0.3 * 2\pi - \sin(0.5u))$ and $\sigma(u) = 1$ for $0 \leq u \leq 1$. This model is a modified version of TV-AR(2) in Zhao (2008). In this simulation, we draw 1000 samples of size n from a normal distribution with parameters $a_1(u) = 2\cos(0.3 * 2\pi - \sin(0.5u))$ and $\sigma(u) = 1$. The sample sizes used were $n = 100, 500, 1000$. We can see in Figure 3 that, for some selected values u, as the sample size increases the estimates tend to normality.

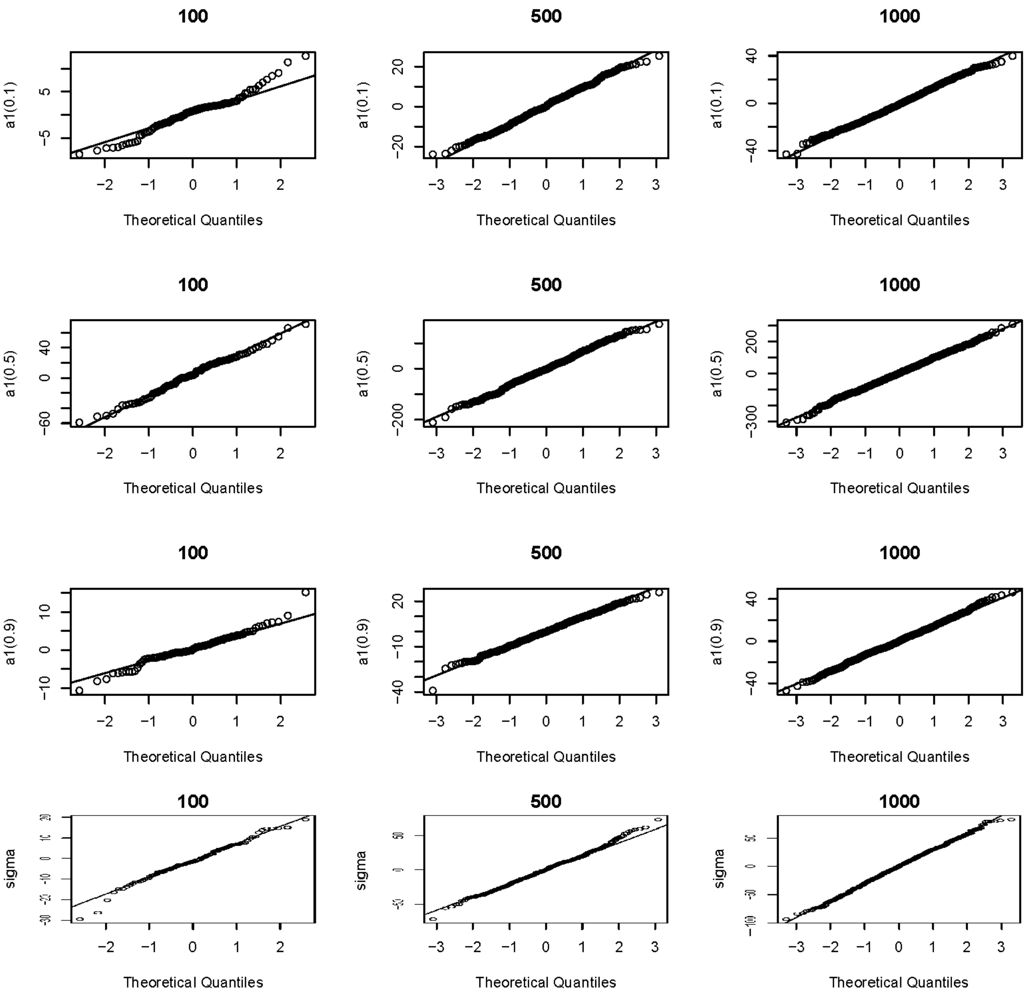

Figure 3: Normal P-P plots for a_1 and σ.

4.3. Contaminated models

To investigate the robustness of the method when the model is not correctly specified, we propose a new estimator for θ of f_θ. The estimator for $I(f)$ is defined by the requirement that

$$d^{HD}(f_{I(\hat{f}_{\alpha,N})}, f_\alpha) = \min_{t\in\Theta} d^{HD}(f_t, \hat{f}_{\alpha,N})$$

where $\hat{f}_{\alpha,N} = (1-\alpha)\hat{f}_N + \alpha\delta_{[0,1]}$ is the contaminated density and $\delta_{[0,1]}$ denotes the uniform density on the interval $[0,1]$ and $\alpha \in [0,1]$.

In the MHD estimation of the coefficients of model (11), we replace $\hat{f}_N$ by $\hat{f}_{\alpha,N}$ for six particular values of α. An application of Monte Carlo is shown in Table 1 for the values of $\hat{a}_2(u)$ and $\hat{\sigma}(u)$ and, in the following Figure 4, for $\hat{a}_1(u)$. True values are : $a_2(u) = 0.819$ and $\sigma^2(u) = 1.28$. In parentheses we have the Mean Absolute Error (MAE).

Table 1: Estimation of $a_2(u)$ and $\sigma(u)$.

α	$\hat{a}_2(u)$	$\hat{\sigma}(u)$
0.05	0.805	1.007
	(0.027)	(0.007)
0.1	0.821	1.019
	(0.011)	(0.019)
0.2	0.737	0.986
	(0.095)	(0.014)
0.3	0.818	0.962
	(0.014)	(0.038)
0.4	0.831	1.296
	(0.001)	(0.296)
0.5	0.826	1.103
	(0.006)	(0.103)

4.4. Comparison of the MHD method to other methods

To illustrate the gain provided by our estimator, we provide a comparison with two simple estimation methods. We investigate the performance of the MHD estimator for locally stationary processes and compare it to Maximum Likelihood (ML) and Cramér-von Mises (CVM) estimators by simulating 1000 different series of size 100, 500 and 1000 from the above time varying AR(2)-process in (11). The paper by Par and Schucany (1980) provides a reference on the Cramér-von Mises minimum distance estimation technique. We examine both correctly specified models and contaminated models. We calculate the root mean squared error (RMSE) of an estimate θ as

$$RMSE = \sqrt{\frac{1}{r}\sum_{i=1}^{r}(\theta_i - \theta)^2},$$

where r is the number of replications performed in the simulation study and $\hat{\theta}_i$ is the estimate obtained in the i^{th} replication.

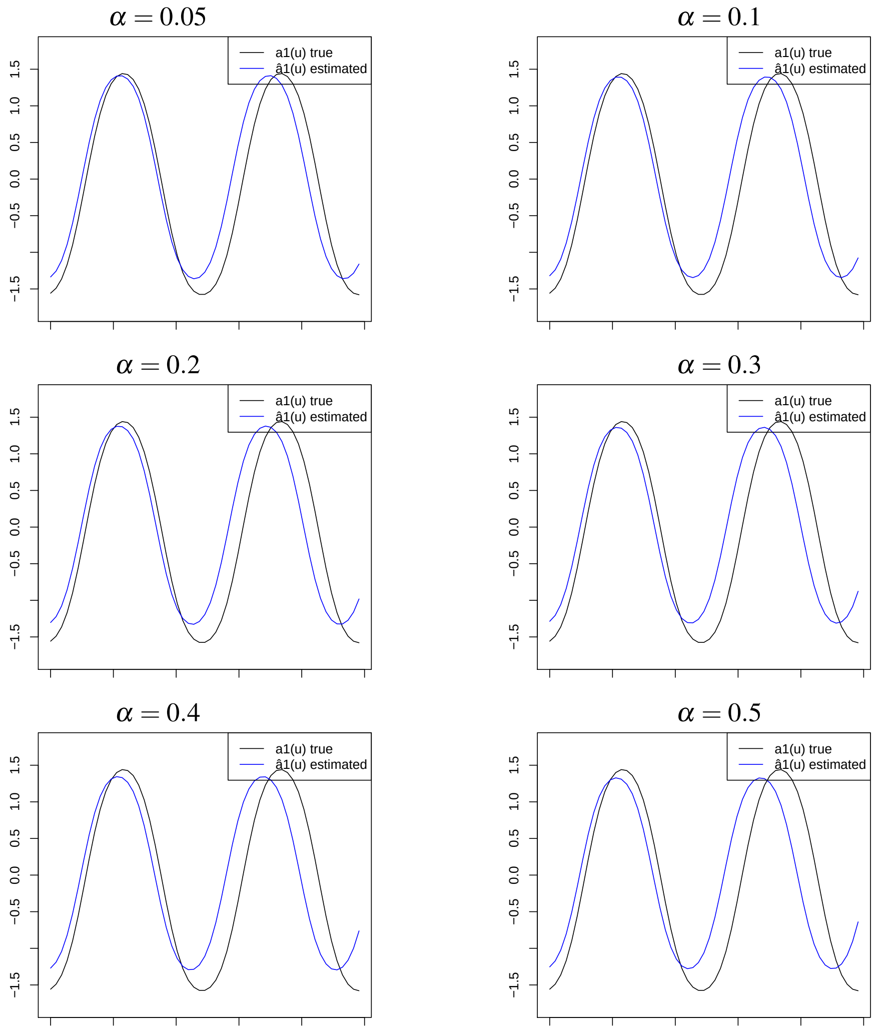

Figure 4: True and estimated time varying coefficient $a_1(u)$.

Table 2: Comparison of Root Mean Squared Error for the uncontaminated data. The reported values have been calculated using 1000 replication samples.

	RMSE					
	MHDE		MLE		CVM	
T	$\hat{a}_2(u)$	$\hat{\sigma}(u)$	$\hat{a}_2(u)$	$\hat{\sigma}(u)$	$\hat{a}_2(u)$	$\hat{\sigma}(u)$
100	0.8184	1.2796	0.8091	1.3078	0.8083	1.2936
	(0.0186)	(0.0416)	(0.1052)	(0.1164)	(0.0692)	(0.0981)
500	0.8232	1.2930	0.8159	1.3104	0.8056	1.3006
	(0.0175)	(0.0291)	(0.0725)	(0.0942)	(0.0753)	(0.0917)
1000	0.8112	1.2845	0.8190	1.2981	0.8064	1.2916
	(0.0148)	(0.0273)	(0.0143)	(0.0268)	(0.0146)	(0.0281)

Table 3: Comparison of Root Mean Squared Error for the contaminated data. The reported values have been calculated using 1000 replication samples.

		RMSE					
		MHDE		MLE		CVM	
T	α	$\hat{a}_2(u)$	$\hat{\sigma}(u)$	$\hat{a}_2(u)$	$\hat{\sigma}(u)$	$\hat{a}_2(u)$	$\hat{\sigma}(u)$
100	0.01	0.8209	1.2869	0.8271	1.3164	0.8267	1.3126
		(0.0183)	(0.0365)	(0.1030)	(0.1139)	(0.1103)	(0.1163)
	0.05	0.8196	1.2776	0.8297	1.3109	0.8291	1.3106
		(0.0190)	(0.0368)	(0.1063)	(0.1189)	(0.1108)	(0.1189)
	0.1	0.8186	1.2760	0.8252	1.3069	0.8257	1.2983
		(0.0197)	(0.0392)	(0.1096)	(0.1207)	(0.1127)	(0.1219)
500	0.01	0.8189	1.2926	0.8192	1.3114	0.81904	1.3129
		(0.0172)	(0.0297)	(0.0745)	(0.0916)	(0.8152)	(0.0937)
	0.05	0.8184	1.2896	0.8187	1.3012	0.8190	1.3108
		(0.0176)	(0.0317)	(0.0663)	(0.0674)	(0.0551)	(0.0772)
	0.1	0.8175	1.3001	0.8189	1.2968	0.8227	1.3062
		(0.0280)	(0.0425)	(0.0472)	(0.0674)	(0.055)	(0.0897)
1000	0.01	0.8109	1.2535	0.8184	1.2864	0.8173	1.2906
		(0.0153)	(0.0279)	(0.0147)	(0.0265)	(0.0250)	(0.0324)
	0.05	0.8135	1.2614	0.8172	1.2720	0.8176	1.2783
		(0.0172)	(0.0284)	(0.0168)	(0.0279)	(0.0170)	(0.0296)
	0.1	0.8174	1.2745	0.8197	1.2981	0.8191	1.2815
		(0.0183)	(0.0218)	(0.0178)	(0.0213)	(0.0194)	(0.0287)

The results, for $\hat{a}_2(u)$ and $\hat{\sigma}(u)$, are summarised in Tables 2 – 3. In Figures 6 – 7 (see Appendix A), some results concerning $\hat{a}_1(u)$ are represented.
Examining the results of the three methods reported in Table 2, we can see that, for the parameters $a_2(u)$ and $\sigma(u)$, the performance of the ML and CVM estimators is inferior to that of the MHD estimator, for the small sample sizes. This observation is consistent with Hasselblad (1969) who warned that ML estimators have large errors when the sample size becomes small. While, for large sample sizes, the MHD estimator does slightly worse than the ML and the CVM estimators. The result obtained compared with the MLE agree with Lindsay (1994) who showed that, usually, the ML method works better for well-specified models when compared to the MHD method.

Table 3 shows that the MHD estimator is more robust than ML and CVM estimators when the incorrect model is hypothesised, particularly when the sample size is small. In general, the MHDE is considerably more efficient than the MLE and the CVME for contaminated data. For small α (low contamination) if the sample size is small, the MHD estimators are far better than the ML and the CVM estimators, but if the sample size becomes large the MLE and the CVM estimators are far better than the MHD estimators. For α increasing, the two methods MHD and ML work in the same manner and are better than the CVM method.

Our results suggest that the MHDE is quite resistant to large quantities of bad observations.

Concluding, we can say that the MHDE for locally stationary processes is appealing compared to MLE and CVM with respect to both efficiency and robustness.

5. Application in real data

In this section, we based our analysis on the daily returns of the S&P 500 index

$$R_t = \ln\left(\frac{X_t}{X_{t-1}}\right) = \ln(X_t) - \ln(X_{t-1}),$$

where X_t is the closing level of the index between January 2, 1957 and July 26, 2010. The local dependence structure of the S&P 500 log-returns was studied by Stărică and Granger (2005) between January 3, 1928 and May 25, 2000. Our objective here is to fit an $ARMA(1,1)$ model as local stationary approximation of the dynamics of the returns by using the MHD estimation method.
The intervals of homogeneity corresponding to the $ARMA(1,1)$ local approximation are built using a goodness-of-fit test based on the statistic defined in Stărică and Granger (2005). The method of estimation of the *ARMA* parameters was that of minimum Hellinger distance.

Figure 5 displays the results of local estimation of the $ARMA(1,1)$ process by the minimum Hellinger distance estimator. From the graph in Figure 5, we conclude that the estimated coefficients *AR* and *MA* are always very closed.

6. Discussion

We have developed a method to estimate locally stationary processes. This method is recommended since it is stable and can optimize the estimates with respect to both efficiency and robustness. Our simulation study indicates that the parameters estimated using the Hellinger distance method is very

Time-Varying AR coefficient (continuous line) and MA coefficient (dotted line).

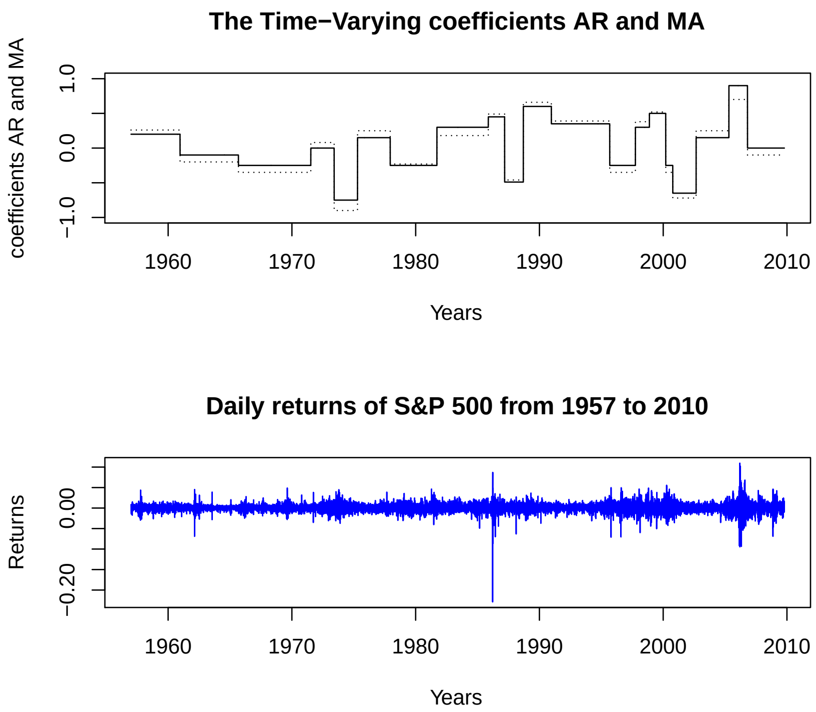

Figure 5: Results of locally estimation of the $ARMA(1,1)$ process on the dynamic of returns of the S&P 500 by the minimum Hellinger distance estimator.

close to that of the estimates using MLE and CVM methods. From the viewpoint of robustness and efficiency, our simulations show that the MHDE has high efficiency when the data are from the model family and is more robust than the MLE while still retaining acceptable efficiency when the data are not from the model family. When compared to another minimum distance estimator, the Cramér-von Mises minimum distance estimator (CVME), the MHDE is more robust when the contaminant is moderately far away from the underlying distribution. An illustration of our method to the S&P 500 returns R_t was given. We approximate the data by time varying $ARMA(1,1)$ processes and the estimation provides good results. Our method is applied to stochastic processes whose margins have time varying infinite order moving average representation. It will be nice to extend our results to more general processes.

References

ADAK, S. (1998). Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association*, **93**, 1488 – 1501.

BASU, A. AND LINDSAY, B. G. (1994). Minimum disparity estimation for continuous models:

efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics*, **46**, 683 – 705.

BERAN, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, **5**, 445 – 463.

BOWMAN, A. W. (1984). An alternative method of cross-validation for smoothing of density estimates. *Biometrika*, **71** (2), 353 – 360.

CUTLER, A. AND CORDERO-BRÃNA, O. I. (1996). Minimum Hellinger distance estimation for finite mixture models. *Journal of the American Statistical Association*, **91**, 1716 – 1723.

DAHLHAUS, R. (1996). Asymptotic statistical inference for nonstationary process with evolutionary spectra. *In* ROBINSON, P. M. AND ROSENBLATT, M. (Editors) *Athens Conference on Applied Probability and Time Series Analysis*, volume 2. Springer: New York, pp. 145 – 159.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, **25**, 1 – 37.

DAHLHAUS, R. AND SUBBA RAO, S. (2006). Statistical inference for locally stationary ARCH models. *The Annals of Statistics*, **34**, 1075 – 1114.

GRÉGOIRE, G. (1993). Least squares cross-validation for counting process intensities. *Scandinavian Journal of Statistics*, **20** (4), 343 – 360.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *Journal of American Statistical Association*, **69**, 383 – 393.

HASSELBLAD, V. (1969). Estimation of finite mixtures from the exponential family. *Journal of the American Statistical Association*, **64**, 1459 – 1471.

HILI, O. (1995). On the estimation of nonlinear time series models. *Stochastics and Stochastics Reports*, **52**, 207 – 226.

HILI, O. (1996). Estimateurs du minimum de distance d'Hellinger d'un modèle autorégressif multivarié, application au modèle EXPARMA. *Comptes Rendus de l'Académie des Sciences – Series I*, **323**, 1073 – 1078.

HILI, O. (1999). On the estimation of β-ARCH models. *Statistics and Probability Letters*, **45**, 285 – 293.

HILI, O. (2003). Hellinger distance estimation of nonlinear dynamical systems. *Statistics and Probability Letters*, **63**, 177 – 184.

HOSOYA, Y. (1974). *Estimation Problems On Stationary Time-series Models*. PhD thesis, Yale University: New Haven, Connecticut, USA.

HOSOYA, Y. AND TANIGUCHI, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, **10** (1), 132 – 153.

LEUNG, D., MARRIOTT, F., AND WU, E. (1993). Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics*, **2**, 333 – 339.

LINDSAY, B. (1994). Efficiency versus robustness: the case of minimum Hellinger distance and related methods. *The Annals of Statistics*, **22** (2), 1081 – 1114.

LUDEÑA, C. (2000). Parametric estimation for Gaussian long-range dependent processes based on the log-periodogram. *Bernoulli*, **6** (4), 709 – 728.

PAR, W. C. AND SCHUCANY, W. R. (1980). Minimum distance and robust estimation. *Journal of the American Statistical Association*, **75**, 616 – 624.

RUDEMO, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**, 65 – 78.

SERGIDES, M. AND PAPARODITIS, E. (2007). Bootstrapping the local periodogram of locally stationary processes. *Journal of Time Series Analysis*, **29** (2), 264 – 299.

SIMPSON, D. G. (1987). Minimum hellinger distance estimation for the analysis of count data. *Journal of American Statistical Association*, **82**, 802 – 807.

SIMPSON, D. G. (1989). Hellinger deviance test: efficiency, breakdown points, and examples. *Journal of American Statistical Association*, **84**, 107 – 113.

STĂRICĂ, C. AND GRANGER, C. W. J. (2005). Nonstationarities in stock return. *The Review of Economics and Statistics*, **81**, 553 – 574.

TAMURA, R. N. AND BOOS, D. D. (1986). Minimum Hellinger distance estimate for multivariate location and covariance. *Journal of American Statistical Association*, **81**, 223 – 229.

TANIGUCHI, M. (1979). On estimation of parameters of Gaussian stationary processes. *Journal of Applied Probability*, **16**, 575 – 591.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Arkiv för Matematik*, **2**, 423 – 434.

WHITTLE, P. (1954). Some recent contributions to the theory of stationary processes. *In* WOLDIS, H. (Editor) *A Study In The Analysis Of Stationary Time Series*, 2nd edition. Almqvist and Wiksell: Uppsala, pp. 196 – 228.

ZHAO, H. (2008). *Development of a Wireless Health Monitoring System with an Efficient Power Management Scheme Based on Localized Time-varying Data Analyses*. PhD thesis, University of New York, Stony Brook. ProQuest: Ann Arbor, MI, USA.

Appendix A

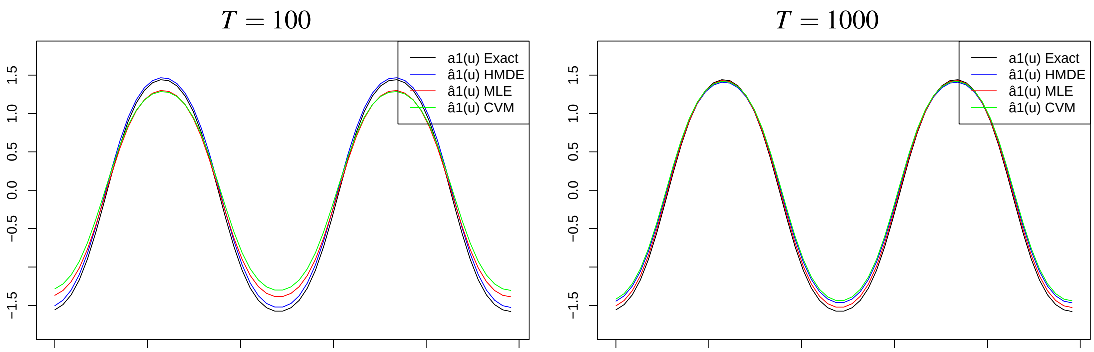

Figure 6: True and estimated time varying coefficient $a_1(u)$ by MHDE, MLE and CVM.

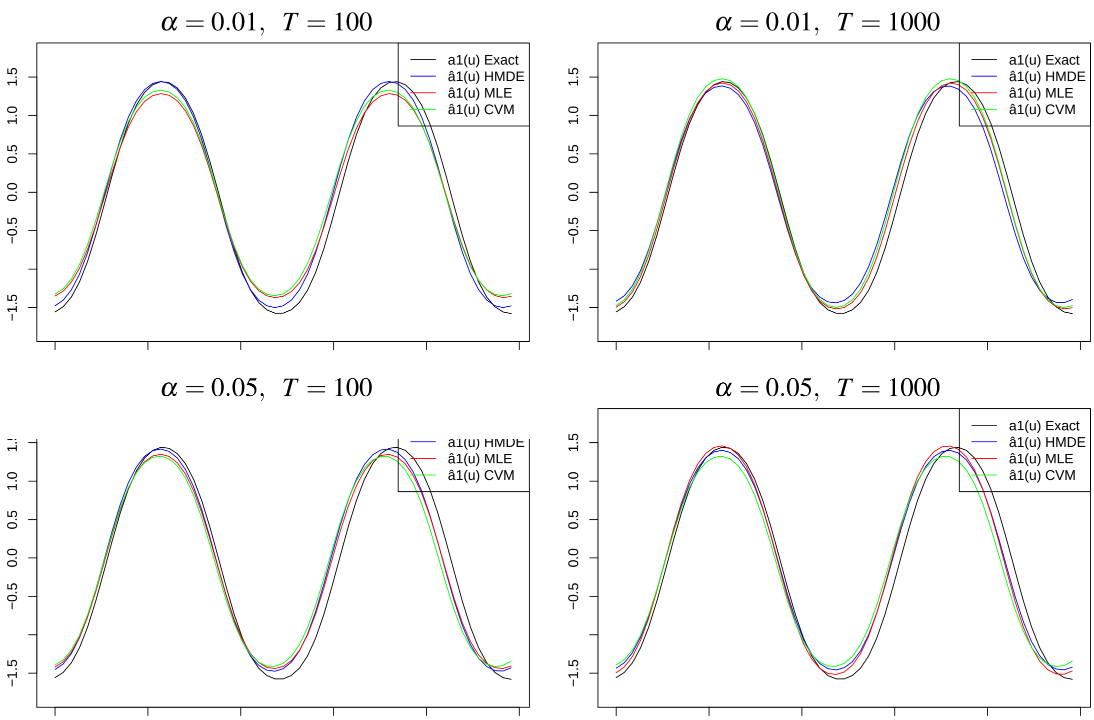

Figure 7: True and estimated time varying coefficient $a_1(u)$ by MHDE, MLE and CVM.

Manuscript received, 2014-11-05, revised, 2016-03-15, accepted, 2016-05-02.