Elsevier Editorial System(tm) for Statistics and Probability Letters
Manuscript Draft

Manuscript Number:

Title: Robust parameter estimation for stationary processes by an exotic disparity from prediction problem

Article Type: Letter

Keywords: stationary process; spectral density; minimum contrast estimation; prediction problem; asymptotic efficiency

Corresponding Author: Dr. Yan Liu,

Corresponding Author's Institution: Waseda University

First Author: Yan Liu

Order of Authors: Yan Liu

Abstract: A new class of disparities from the point of view of prediction problem are proposed for minimum contrast estimation of spectral densities of stationary processes. We investigate asymptotic properties of the minimum contrast estimators based on the new disparities for stationary processes with both finite and infinite variance innovations. The relative efficiency and the robustness against randomly missing observations are shown in our numerical simulations.

Dr. Yan Liu
Dept. Applied Mathematics
Waseda University, Japan
Okubo 3-4-1, Shinjuku, Tokyo 169-8555
Phone: (+81) 03 5286 3190
Email: yan.liu@aoni.waseda.jp

June 7, 2016, Tokyo

Dear Editors,

I am writing to submit my manuscript entitled, "Robust parameter estimation for stationary processes by an exotic disparity from prediction problem", for consideration of publication in Statistics & Probability Letters as a letter.

This letter contains a new class of disparities, originating from prediction problem, for estimation of spectral densities of stationary processes with finite/infinite variance innovations. We believe the new method in time series analysis will be of interest to the readers of your journal.

This manuscript has not been published and is not under consideration for publication elsewhere.

Thank you for receiving my manuscript and considering it for review. I appreciate your time and look forward to your response.

Yours sincerely,

Yan Liu
Research Associate
Department of Applied Mathematics,
Waseda University, Japan

Manuscript
Click here to view linked References

Robust parameter estimation for stationary processes by an exotic disparity from prediction problem

Yan Liu*

Department of Applied Mathematics, Waseda University, Tokyo 169-8555, Japan

Abstract

A new class of disparities from the point of view of prediction problem are proposed for minimum contrast estimation of spectral densities of stationary processes. We investigate asymptotic properties of the minimum contrast estimators based on the new disparities for stationary processes with both finite and infinite variance innovations. The relative efficiency and the robustness against randomly missing observations are shown in our numerical simulations.

Keywords: Stationary process, Spectral density, Minimum contrast estimation, Prediction problem, Asymptotic efficiency
2010 MSC: 62M10, 62F12, 62F35

1. Introduction

Methods of fitting parametric models to linear time series have been investigated for a long time. One method for parameter estimation is to minimize a certain disparity measure $D(f_\theta, \hat{g}_n)$ between the model f_θ and the estimated spectral density $\hat{g}_n$. Two disparity measures, the location disparity and the scale disparity, have been mainly considered so far. The *location* disparity $D(f_\theta, \hat{g}_n) = \int_{-\pi}^{\pi}[\Phi(f_\theta(\omega))^2 - 2\Phi(f_\theta(\omega))\Phi(\hat{g}_n(\omega))]d\omega$ with a bijective function $\Phi(\cdot)$ was proposed in Taniguchi [13]. The *scale* disparity $D(f_\theta, \hat{g}_n) = \int_{-\pi}^{\pi} K(f_\theta(\omega)/\hat{g}_n(\omega))d\omega$ with a sufficiently smooth contrast function $K(\cdot)$ was proposed in Taniguchi [14]. Both methods for parameter estimation are consistent.

In this paper, we propose a new consistent disparity as follows.

$$D(f_\theta, I_{n,X}) = \int_{-\pi}^{\pi} a(\theta) f_\theta^\alpha(\omega) I_{n,X}(\omega) d\omega, \quad \alpha \neq 0, \tag{1}$$

where $a(\theta)$ is

$$a(\theta) = \begin{cases} \left(\int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\omega)d\omega\right)^{-\frac{\alpha}{\alpha+1}}, & \text{if } \alpha \neq -1, \\ C, & \text{if } \alpha = -1. \end{cases}$$

*Corresponding author. Tel.: +81 03 5286 3190.
Email address: yan.liu@aoni.waseda.jp (Yan Liu)

The disparity (1) originates from the notice of minimizing prediction error of prediction problem in L^p for stationary processes. Our new disparity is neither included in the class of location disparities nor scale disparities.

The notations and symbols used in this paper are listed in the following: the constant C denotes some real number which varies from context; ∂ denotes the differentiation with respect to the parameter θ; A_j and A_{ij} denote the jth and the (i, j)th element of corresponding vector and matrix; A^{T} denotes the transpose of a matrix A; $\mathrm{cum}(X_1, \dots, X_n)$ denotes the nth cumulant of the random variables $\{X_1, \dots, X_n\}$; $\mathbb{1}(\cdot)$ denotes the indicator function.

The remaining of the paper is organized as follows. In Section 2, we derive the disparity (1) from the prediction problem in L^p. Asymptotic properties of the proposed estimator for stationary processes with both finite and infinite variance innovations are investigated in Section 3. In section 4, we investigate our new disparity (1) in numerical simulations. Especially, we apply our new disparity (1) to the irregular observed stationary processes as an application for robust parameter estimation.

2. A new class of disparities from prediction problem

In this section, we derive our disparity (1) from the prediction problem of stationary processes in L^p. Suppose $\{X(t), t = 1, 2, \dots\}$ is a stationary process with spectral density $g(\omega)$. Let $\mathbb{Z}$ denote the set of all integers, $\mathbb{Z}_0 = \mathbb{Z} \setminus \{0\}$ and $\mathcal{M}$ denote the closed linear manifold generated by $\{e^{ij\omega},\ j \in \mathbb{Z}_0\}$. Consider the problem of minimizing the prediction error in L^p on $\mathcal{M}$ for $1 < p < \infty$, that is,

$$\inf_{\phi \in \mathcal{M}} \int_{-\pi}^{\pi} |1 - \phi(\omega)|^p g(\omega) d\omega. \tag{2}$$

For example, interpolation problem for time point 0 is formulated by (2) with $p = 2$ (cf. Grenander and Rosenblatt [5], Rosenblatt [11]). As shown in Miamee and Pourahmadi [9], the best predictor $\phi(\omega)$ is given by

$$\phi(\omega) = 1 - \Big(\frac{1}{2\pi} \int_{-\pi}^{\pi} g^{\alpha+1}(\omega) d\omega\Big)^{-1} g^{\alpha+1}(\omega),$$

where $\alpha = -p/(p-1)$. In practice, it is usually difficult to know the true density $g(\omega)$ a priori. Suppose we fit a parametrized density $f_\theta(\omega)$ with d-dimensional parameter $\theta \in \Theta \subset \mathbb{R}^d$ to the true spectral density for the prediction problem. Then, the error in L^p of the prediction by the best predictor $\phi(\lambda)$ is

$$\Big(\frac{1}{2\pi} \int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\omega) d\omega\Big)^{-\frac{\alpha}{\alpha+1}} \Big(\frac{1}{2\pi} \int_{-\pi}^{\pi} f_\theta^{\alpha}(\omega) g(\omega) d\omega\Big). \tag{3}$$

Motivated by (3), let us consider an estimation procedure to estimate the parameter θ by minimizing the following disparity,

$$D(f_\theta, g) = \int_{-\pi}^{\pi} a(\theta) f_\theta^{\alpha}(\omega) g(\omega) d\omega, \quad \alpha \neq 0, \tag{4}$$

where $a(\theta)$ is

$$a(\theta) = \begin{cases} \left(\int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\omega) d\omega\right)^{-\frac{\alpha}{\alpha+1}}, & \text{if } \alpha \neq -1, \\ C, & \text{if } \alpha = -1. \end{cases}$$

We call (4) the *exotic* disparity.

Next, let us investigate the basic property of the exotic disparity. The disparity is not included in the class of either location disparities or scale disparities since the parametrized spectral density and the true density is not homogenous in most cases. However, the definition of the disparity can be motivated by the following two examples: (i) the Whittle disparity when $\alpha = -1$; (ii) the estimation procedure minimizing the interpolation error when $\alpha = -2$. The comparison of the efficiency between these two methods are considered in Suto et al. [12]. In addition to the efficiency, the robustness against randomly missing observations by the exotic disparity (4) is also considered in Section 4 in this paper.

In the following, we impose Assumption 1.

Assumption 1.

(i) *The parameter space Θ is a compact subset of $\mathbb{R}^d$.*

(ii) *If $\theta_1 \neq \theta_2$, then $f_{\theta_1} \neq f_{\theta_2}$ on a set of positive Lebesgue measure.*

Let $\mathcal{F}(\Theta)$ denote a set of spectral densities indexed by parameter θ. Under Assumption 1, even if the parametric spectral density f_θ is not differentiable, there is a maximizer (or a minimizer) achieved by the true value θ_0.

Theorem 2.1. *Under Assumption 1, we have the following results:*

(i) *If $\alpha > 0$, then θ_0 maximizes the exotic disparity $D(f_\theta, f_{\theta_0})$ for any $f_{\theta_0} \in \mathcal{F}(\Theta)$.*

(ii) *If $\alpha < 0$, then θ_0 minimizes the exotic disparity $D(f_\theta, f_{\theta_0})$ for any $f_{\theta_0} \in \mathcal{F}(\Theta)$.*

From Theorem 2.1, it is shown that there exists a maximum (or minimum) $\theta_0 \in \Theta$ in the criterion $D(f_\theta, g)$. The results are different at the two sides of $\alpha = 0$. To keep uniformity of the context, we suppose $\alpha < 0$. Accordingly, we can define the functional T by

$$D(f_{T(g)}, g) = \min_{t \in \Theta} D(f_t, g), \quad \text{for every } g \in \mathcal{F}. \tag{5}$$

Thus, if $g \in \mathcal{F}(\Theta)$, $\theta_0 = T(g)$.

In addition, we assume the following assumption in Assumption 1.

(iii) The parametric spectral density $f_\theta(\omega)$ is three times continuously differentiable with respect to θ and the second derivative $\partial^2/\partial\theta^2 f_\theta(\omega)$ is continuous in ω.

3. Parameter estimation based on exotic disparity

In this section, we investigate asymptotic behavior of the estimation procedure (5).

3.1. Finite variance innovations

Suppose $\{X(t);\, t = 1, 2, \dots\}$ is a stationary process with mean zero, which is generated by

$$X(t) = \sum_{j=0}^{\infty} G(j)\epsilon(t-j), \quad t = 1, 2, \dots,$$

where the fourth-order stationary innovation process $\{\epsilon(t)\}$ satisfies $E\epsilon(t) = 0$ and $E\epsilon(s)\epsilon(t) = \delta(s,t)\sigma^2$ with $\sigma^2 > 0$. We impose the following regularity conditions.

Assumption 2. *For all $|z| \le 1$, there exist $C < \infty$ and $\delta > 0$ such that*

(i) $\sum_{j=0}^{\infty}(1 + j^2)|G(j)| \le C$;

(ii) $\left|\sum_{j=0}^{\infty} G(j)z^j\right| \ge \delta$;

(iii) $\sum_{t_1,t_2,t_3=-\infty}^{\infty}|Q_\epsilon(t_1,t_2,t_3)| < \infty$.

Assumption 2 (iii) guarantees the existence of a fourth-order spectral density. The fourth-order spectral density $\tilde{Q}_\epsilon(\omega_1, \omega_2, \omega_3)$ is

$$\tilde{Q}_\epsilon(\omega_1, \omega_2, \omega_3) = \left(\frac{1}{2\pi}\right)^3 \sum_{t_1,t_2,t_3=-\infty}^{\infty} Q_\epsilon(t_1,t_2,t_3)e^{-i(\omega_1 t_1 + \omega_2 t_2 + \omega_3 t_3)}.$$

Let $\mathcal{F}$ denote the set of all spectral densities with respect to the Lebesgue measure on $[-\pi, \pi]$. More specifically, we define $\mathcal{F}$ as $\mathcal{F} = \{g : g(\omega) = \sigma^2|\sum_{j=0}^{\infty} G(j)\exp(-ij\omega)|^2/(2\pi)\}$. For the linear process $\{X(t);\, t = 1, 2, \dots\}$, $I_{n,X}(\omega)$ denotes the periodogram constructed from a partial realization $\{X(1), \dots, X(n)\}$, that is,

$$I_{n,X}(\omega) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X(t)e^{it\omega}\right|^2, \quad -\pi \le \omega \le \pi.$$

Now, the parameter estimator based on (5) for the process is

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} D(f_\theta, I_{n,X}). \tag{6}$$

Let $\mathcal{B}(t)$ denote the σ-field generated by the uncorrelated process $\{\epsilon(s)\}_{s=-\infty}^{t}$.

Assumption 3. (i) *For each nonnegative integer m and $\eta_1 > 0$,*

$$\mathrm{var}[E\{\epsilon(n)\epsilon(n+m)|\mathcal{B}(t_1-\tau)\} - \delta(m,0)\sigma^2] = O(\tau^{-2-\eta_1})$$

uniformly in t.

(ii) *For $\eta_2 > 0$,*

$$E|E\{\epsilon(t_1)\epsilon(t_2)\epsilon(t_3)\epsilon(t_4)|\mathcal{B}(t_1-\tau)\} - E\{\epsilon(t_1)\epsilon(t_2)\epsilon(t_3)\epsilon(t_4)\}| = O(\tau^{-1-\eta_2}),$$

uniformly in t_1, where $t_1 \le t_2 \le t_3 \le t_4$.

(iii) *For any $\eta_3 > 0$ and for any integer $L \geq 0$, there exists $B_\rho > 0$ such that*

$$E[T(n,s)^2 \mathbb{1}\{T(n,s) > B_{\eta_3}\}] < \eta_3$$

uniformly in n and s, where

$$T(n,s) = \Big[n^{-1/2} \sum_{r=0}^{L} \Big\{ \sum_{t=1}^{n} \epsilon(t+s)\epsilon(t+s+r) - \sigma^2 \delta(0,r) \Big\}^2 \Big]^{1/2}.$$

Theorem 3.1. *Under Assumptions 1–3, if $T(g)$ exists uniquely in the interior of the parameter space Θ, then for the estimator $\hat{\theta}_n$ defined by (6), it holds that*

(i) *$\hat{\theta}_n$ converges to θ_0 in probability;*

(ii) *The distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean 0 and variance $H(\theta_0)^{-1} V(\theta_0) H(\theta_0)^{-1}$, where*

$$\begin{aligned}
H(\theta_0) &= \Big(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\omega) \partial f_{\theta_0}(\omega) d\omega\Big)\Big(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\omega) \partial f_{\theta_0}(\omega) d\omega\Big)^{\mathrm{T}} \\
&\quad - \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\omega) d\omega \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha-1}(\omega) \Big(\partial f_{\theta_0}(\omega)\Big)\Big(\partial f_{\theta_0}(\omega)\Big)^{\mathrm{T}} d\omega, \\
V(\theta_0) &= 4\pi \int_{-\pi}^{\pi} \Big(f_{\theta_0}^{\alpha}(\omega) \partial f_{\theta}(\omega) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\lambda) d\lambda - f_{\theta_0}^{\alpha+1}(\omega) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\lambda) \partial f_{\theta_0}(\lambda) d\lambda \Big) \\
&\quad \Big(f_{\theta_0}^{\alpha}(\omega) \partial f_{\theta}(\omega) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\lambda) d\lambda - f_{\theta_0}^{\alpha+1}(\omega) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\lambda) \partial f_{\theta_0}(\lambda) d\lambda \Big)^{\mathrm{T}} d\omega \\
&\quad + 2\pi \iint_{-\pi}^{\pi} \Big(f_{\theta_0}^{\alpha-1}(\omega_1) \partial f_{\theta_0}(\omega_1) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\lambda) d\lambda - f_{\theta_0}^{\alpha}(\omega_1) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\lambda) \partial f_{\theta_0}(\lambda) d\lambda \Big) \\
&\quad \Big(f_{\theta_0}^{\alpha-1}(\omega_2) \partial f_{\theta_0}(\omega_2) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\lambda) d\lambda - f_{\theta_0}^{\alpha}(\omega_2) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\lambda) \partial f_{\theta_0}(\lambda) d\lambda \Big)^{\mathrm{T}} \\
&\quad \tilde{Q}_X(-\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2.
\end{aligned}$$

If we impose a simple but stronger assumption on the fourth order cumulant of $\{\epsilon(t)\}$, then the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is much simpler. We compare the asymptotic variance with the Fisher information matrix for Gaussian process, which is defined as

$$\mathcal{F}(\theta_0) = \frac{1}{4\pi} \int_{-\pi}^{\pi} f_{\theta_0}^{-2}(\omega) \partial f_{\theta_0}(\omega) \partial f_{\theta_0}(\omega)^{\mathrm{T}} d\omega.$$

Theorem 3.2. *If the fourth order cumulant of $\epsilon(t)$ satisfies*

$$\mathrm{cum}\{\epsilon(t_1), \epsilon(t_2), \epsilon(t_3), \epsilon(t_4)\} = \begin{cases} \kappa_4 & \textit{if } t_1 = t_2 = t_3 = t_4, \\ 0 & \textit{otherwise,} \end{cases}$$

and Assumptions 1–3 hold, then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *is asymptotically Gaussian with mean 0 and variance* $H(\theta_0)^{-1}\tilde{V}(\theta_0)H(\theta_0)^{-1}$*, where*

$$\tilde{V}(\theta_0) = 4\pi \int_{-\pi}^{\pi} \Big(f_{\theta_0}^{\alpha}(\omega)\partial f_{\theta_0}(\omega) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\lambda)d\lambda - f_{\theta_0}^{\alpha+1}(\omega) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\lambda)\partial f_{\theta_0}(\lambda)d\lambda \Big)$$
$$\Big(f_{\theta_0}^{\alpha}(\omega)\partial f_{\theta_0}(\omega) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\lambda)d\lambda - f_{\theta_0}^{\alpha+1}(\omega) \int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\lambda)\partial f_{\theta_0}(\lambda)d\lambda \Big)^{\mathrm{T}} d\omega.$$

Furthermore, it holds that

$$H(\theta_0)^{-1}\tilde{V}(\theta_0)H(\theta_0)^{-1} \geq \mathcal{F}(\theta_0)^{-1}.$$

The equality holds when $\alpha = -1$ *or the spectral density does not depend on* ω*.*

3.2. Infinite variance innovations

Next, we consider the problem for the linear processes with infinite variance innovations. Suppose $\{X(t); t = 1, 2, \dots\}$ is a stationary process

$$X(t) = \sum_{j=0}^{\infty} G(j)\epsilon(t-j), \quad t = 1, 2, \dots,$$

where $\{\epsilon(t)\}$ satisfy the following assumptions for infinite variance innovation process.

Assumption 4. *For some* $d > 0$, $\delta = 1 \wedge d$ *and positive sequence* a_n *satisfying* $a_n \uparrow \infty$*, the coefficient* $G(j)$ *and the innovation process* $\{\epsilon(t)\}$ *has the properties of*

(i) $\sum_{j=-\infty}^{\infty} |j||G(j)|^{\delta}$;

(ii) $E|\epsilon(t)|^d < \infty$;

(iii) *as* $n \to \infty$, $n/a_n^{2\delta} \to 0$;

(iv) $\lim_{x\to 0} \limsup_{n\to\infty} P\Big(a_n^{-2}\sum_{t=1}^{n}\epsilon(t)^2 \leq x\Big) = 0.$

(v) *For* $0 < q < 2$*, the distribution of* $\epsilon(t)$ *is in the domain of normal attraction of a symmetric q-stable random variable* Y*. The characteristic function of q-stable distribution* Y *is* $E(e^{itY}) = \exp(-\sigma^q|t|^q)$*.*

We follow the results in Mikosch et al. [10], Klüppelberg and Mikosch [8] to see asymptotic properties of minimum contrast estimators with the exotic disparity in the case of infinite variance innovations. Let $\tilde{I}_{n,X}$ denote the self-normalized periodogram, that is, $\tilde{I}_{n,X} = I_{n,X}/\sum_{i=1}^{n} X(t)^2$. Now, minimum contrast estimator based on (5) for the process is defined by

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} D(f_\theta, \tilde{I}_{n,X}). \tag{7}$$

Also, we introduce the scale constant C_q, i.e.,

$$C_q = \begin{cases} \frac{1-q}{\Gamma(2-q)\cos(\pi q/2)}, & \text{if } q \neq 1, \\ \frac{2}{\pi}, & \text{if } q = 1. \end{cases}$$

Theorem 3.3. *Under Assumptions 1 and 4, if $T(g)$ exists uniquely in the interior of the parameter space Θ, then for the estimator $\hat{\theta}_n$ defined by (7), it holds that*

(i) *$\hat{\theta}_n$ converges to θ_0 in probability;*

(ii) *It holds that*

$$\left(\frac{n}{\log n}\right)^{1/q}(\hat{\theta}_n - \theta_0) \to 4\pi H^{-1}(\theta_0)\sum_{k=1}^{\infty}\frac{Y_k}{Y_0}V_k$$

in law, where $H(\theta_0)$ is the same as in Theorem 3.1,

$$V_k = \left(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\omega)\partial f_{\theta_0}(\omega)d\omega\right)\left(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\omega)\cos(k\omega)d\omega\right) - \left(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\omega)d\omega\right)\left(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha}(\omega)\partial f_{\theta_0}(\omega)\cos(k\omega)d\omega\right),$$

and $\{Y_k\}_{k=0,1,\ldots,}$ are mutually independent random variables. Y_0 is $q/2$-stable with scale $C_{q/2}^{-2/q}$ and Y_k $(k \geq 1)$ is q-stable with scale $C_q^{-1/q}$.

4. Numerical Results

In this section, we perform two numerical simulations for minimum contrast estimators based on the exotic disparity (1) for parameter estimation. One is to see the relative efficiency and the other is to see the robustness with different α.

4.1. Relative efficiency

In the first simulation, we investigate the empirical relative efficiency between the different choices of α when the true model is specified. Let us consider the autoregressive process

$$X(t) - bX(t-1) = \epsilon(t), \tag{8}$$

where the innovation process $\{\epsilon(t);\, t = 1, 2, \ldots\}$ is assumed to be independent and identically distributed random variables. The distributions of random variables in innovation process are assumed to be Gaussian distribution, Laplace distribution and standard Cauchy distribution. Laplace distribution is used as an example for a distribution with fourth cumulant and Cauchy distribution is used as an example for a distribution with infinite variance. All distributions are symmetric around 0. Gaussian distribution and Laplace distribution are set to be unit variance. We are interested in estimating the coefficient b in the model (8) by the exotic disparity (1) with different values of α. The parametric spectral density is assumed to be

$$f_\theta(\omega) = \frac{1}{2\pi}|1 - \theta e^{-i\omega}|^{-2}. \tag{9}$$

We generate 100 samples from the model (8) with coefficients $b = 0$ and 0.9. The estimation for θ by the exotic disparity (1) are repeated via 100 Monte Carlo simulations.

Let $\hat{\theta}_\alpha^{(i)}$ be the estimate by the exotic disparity with α in ith simulation. We define the empirical relative efficiency (ERE) by

$$\text{ERE} = \frac{\sum_{i=1}^{100}(\hat{\theta}_{-1}^{(i)} - b)^2}{\sum_{i=1}^{100}(\hat{\theta}_{\alpha}^{(i)} - b)^2}.$$

EREs for $b = 0$ and 0.9 are reported in Tables 1 and 2, respectively.

For the case $b = 0$, we can see that the exotic disparities achieve more than 74% in relative efficiency when $\alpha = -2, -3$. When $\alpha \leq -4$, the exotic disparities in Gaussian and Laplace cases are much better.

Table 1: The empirical relative efficiency of AR(1) model (8) when $b = 0$

	$\alpha = -1$	$\alpha = -2$	$\alpha = -3$	$\alpha = -4$	$\alpha = -5$	$\alpha = -6$	$\alpha = -7$	$\alpha = -8$
Gaussian	1.000	0.955	0.853	0.447	0.091	0.061	0.043	0.038
Laplace	1.000	0.910	0.749	0.441	0.099	0.050	0.042	0.037
Cauchy	1.000	0.990	0.918	0.074	0.057	0.029	0.020	0.018

On the other hand, ERE of the exotic disparities in Cauchy case are better than the other two finite variance cases when $b = 0.9$. However, the overall relative efficiencies are not as good as the case $b = 0$.

Table 2: The empirical relative efficiency of AR(1) model (8) when $b = 0.9$

	$\alpha = -1$	$\alpha = -2$	$\alpha = -3$	$\alpha = -4$	$\alpha = -5$	$\alpha = -6$	$\alpha = -7$	$\alpha = -8$
Gaussian	1.000	0.604	0.203	0.097	0.067	0.051	0.042	0.037
Laplace	1.000	0.578	0.162	0.075	0.047	0.035	0.029	0.025
Cauchy	1.000	0.847	0.304	0.118	0.073	0.057	0.048	0.041

4.2. Robustness of exotic disparity

In the second simulation, we compare the robustness of the exotic disparity (1) with different α. As pointed out in Fujisawa and Eguchi [4], the exotic disparity has the robustness under the heavy contaminated models.

Let us consider the spectral analysis with randomly missing observations as Bloomfield [1] and Brillinger [2]. The time series with randomly missing observations are generally modeled by the amplitude modulated series $\{Y(t)\}$ such that

$$Y(t) = X(t)Z(t),$$

where

$$Z(t) = \begin{cases} 1, & Y(t) \text{ is observed} \\ 0, & \text{otherwise.} \end{cases}$$

If we define $P(Z(t) = 1) = p$ and $P(Z(t) = 0) = 1 - p$, then the spectral density f_Y for $\{Y(t)\}$ is represented by

$$f_Y(\omega) = p^2 f_X(\omega) + p \int_{-\pi}^{\pi} a(\omega - \alpha) f_X(\alpha) d\alpha, \tag{10}$$

where $a(\omega) = (2\pi)^{-1} \sum_r a_r e^{ir\omega}$ with $a_r = p^{-1}\mathrm{Cov}(Z(t), Z(t + r))$. Not difficult to see from (10), the spectral density f_Y can be considered as f_X heavily contaminated by the spectral density $\int_{-\pi}^{\pi} a(\omega - \alpha) f_X(\alpha) d\alpha$.

Let us fit the parametric spectral density (9) to the following two models: (i) the model (8) and

$$\text{(ii)} \quad X(t) = \epsilon(t) + 5/4\epsilon(t - 1) + 4/3\epsilon(t - 2) + 5/6\epsilon(t - 3). \tag{11}$$

The true value for the model (11) is $\theta = 0.7$. In these two models, we suppose the innovation process $\{\epsilon(t); t = 1, \ldots, n\}$ is independent and identically distributed as standard normal distribution and standard Cauchy distribution as above.

We evaluate the robustness by the ratio of mean squared error of all estimation procedure by

$$\mathrm{RoMSE} = \frac{\sum_{i=1}^{100} (\hat{\theta}_{-1}^{(i)} - \theta)^2}{\sum_{i=1}^{100} (\hat{\theta}_{\alpha}^{(i)} - \theta)^2}.$$

We generated 128 samples from the model (8) and (11), respectively. We randomly chose four sets of time points $T_i \subset \{1, 2, \ldots, 128\}$ as the observed time points as follows. The length of T_i is 32 for $i = 1, 2, 3, 4$.

$$\begin{aligned}
T_1 &= \{1, 7, 9, 17, 19, 23, 26, 30, 34, 39, 44, 50, 54, 58, 59, 61, 66, 67, \\
&\quad 74, 75, 79, 80, 81, 87, 101, 102, 104, 112, 118, 121, 125, 128\}, \\
T_2 &= \{1, 2, 4, 6, 9, 13, 21, 22, 31, 36, 37, 38, 39, 42, 49, 50, 56, 71, \\
&\quad 76, 77, 82, 85, 93, 96, 101, 110, 112, 113, 115, 117, 126, 127\}, \\
T_3 &= \{1, 4, 5, 12, 14, 15, 18, 20, 23, 26, 27, 28, 33, 39, 41, 55, 56, 74, \\
&\quad 78, 83, 84, 85, 88, 100, 104, 106, 107, 108, 109, 114, 115, 120\}, \\
T_4 &= \{2, 4, 12, 14, 34, 38, 39, 41, 42, 43, 44, 49, 54, 55, 56, 59, 60, 63, \\
&\quad 65, 70, 72, 78, 81, 94, 98, 100, 103, 107, 110, 119, 123, 126\}.
\end{aligned}$$

Suppose we obtained the observations on the sets T_i $(i = 1, \ldots, 4)$ of (8) and (11), respectively. The ratio of mean squared errors are given in Tables

Table 3: RoMSE of AR(1) model (8) with Cauchy and Gaussian innovations when $b = 0.9$

Cauchy	$\alpha = -1$	-2	-3	-4
T_1	1.000	1.192	1.149	1.007
T_2	1.000	1.332	1.294	1.053
T_3	1.000	1.182	1.158	0.948
T_4	1.000	1.288	1.307	1.073

Gaussian	$\alpha = -1$	-2	-3	-4
T_1	1.000	1.287	1.490	1.435
T_2	1.000	1.181	1.220	1.199
T_3	1.000	1.148	1.247	1.078
T_4	1.000	1.230	1.395	1.125

Table 4: RoMSE of MA(3) model (11) with Cauchy and Gaussian innovations

Cauchy	$\alpha=-1$	-2	-3	-4	Gaussian	$\alpha=-1$	-2	-3	-4
T_1	1.000	1.053	1.093	0.959	T_1	1.000	1.061	1.140	1.181
T_2	1.000	1.107	1.107	1.230	T_2	1.000	1.099	1.294	2.085
T_3	1.000	1.129	1.266	1.260	T_3	1.000	1.136	1.605	1.533
T_4	1.000	1.068	1.078	0.961	T_4	1.000	1.052	1.144	1.291

Acknowledgements

This research was supported by Grant-in-Aid for JSPS Fellows 26·7404. The author would like to thank Professor Masanobu Taniguchi for his continued suggestions and encouragements. He also would like to thank Professor Hans R. Künsch and Professor Claudia Klüppelberg for their advice on this paper and hospitality during his stay at ETH and TUM.

Reference

[1] Bloomfield, P., 1970. Spectral analysis with randomly missing observations. Journal of the Royal Statistical Society. Series B (Methodological) 32 (3), 369–380.

[2] Brillinger, D. R., 1984. Statistical inference for irregularly observed processes. In: Time series analysis of irregularly observed data. Springer, pp. 38–57.

[3] Brockwell, P. J., Davis, R. A., 1991. Time Series: Theory and Methods. Springer-Verlag.

[4] Fujisawa, H., Eguchi, S., 2008. Robust parameter estimation with a small bias against heavy contamination. Journal of Multivariate Analysis 99 (9), 2053–2081.

[5] Grenander, U., Rosenblatt, M., 1957. Statistical Analysis of Stationary Time Series. Johm Wiley & Sons, New York.

[6] Hosoya, Y., Taniguchi, M., 1982. A central limit theorem for stationary processes and the parameter estimation of linear processes. The Annals of Statistics 10, 132–153.

[7] Kholevo, A., 1969. On estimates of regression coefficients. Theory of Probability & Its Applications 14 (1), 79–104.

[8] Klüppelberg, C., Mikosch, T., 1996. The integrated periodogram for stable processes. The Annals of Statistics 24 (5), 1855–1879.

[9] Miamee, A., Pourahmadi, M., 1988. Best approximations in $\mathrm{L}^p(\mathrm{d}\mu)$ and prediction problems of szegö, kolmogorov, yaglom, and nakazi. Journal of the London Mathematical Society 2 (1), 133–145.

[10] Mikosch, T., Gadrich, T., Klüppelberg, C., Adler, R. J., 1995. Parameter estimation for arma models with infinite variance innovations. The Annals of Statistics 23 (1), 305–326.

[11] Rosenblatt, M., 2000. Gaussian and non-Gaussian linear time series and random fields. Springer Science & Business Media.

[12] Suto, Y., Liu, Y., Taniguchi, M., 2016. Asymptotic theory of parameter estimation by a contrast function based on interpolation error. Statistical Inference for Stochastic Processes 19 (1), 93–110.

[13] Taniguchi, M., 1981. An estimation procedure of parameters of a certain spectral density model. Journal of the Royal Statistical Society. Series B 43, 34–40.

[14] Taniguchi, M., 1987. Minimum contrast estimation for spectral densities of stationary processes. Journal of the Royal Statistical Society. Series B 49, 315–325.

Appendix: Proofs

Proof of Theorem 2.1

Proof. First, suppose $\alpha > 0$. The exotic disparity (4) can be rewritten by

$$D(f_\theta, f_{\theta_0}) = \frac{\int_{-\pi}^{\pi} f_\theta^\alpha(\omega) f_{\theta_0}(\omega) d\omega}{(\int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\omega) d\omega)^{\frac{\alpha}{\alpha+1}}}.$$

From Hölder's inequality, the numerator then satisfies

$$\begin{aligned}\int_{-\pi}^{\pi} f_\theta^\alpha(\omega) f_{\theta_0}(\omega) d\omega &\leq \left(\int_{-\pi}^{\pi} \{f_\theta^\alpha(\omega)\}^{\frac{\alpha+1}{\alpha}} d\omega\right)^{\frac{\alpha}{\alpha+1}} \left(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\omega) d\omega\right)^{\frac{1}{\alpha+1}} \\ &= \left(\int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\omega) d\omega\right)^{\frac{\alpha}{\alpha+1}} \left(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\omega) d\omega\right)^{\frac{1}{\alpha+1}}.\end{aligned}$$

Therefore,

$$D(f_\theta, f_{\theta_0}) \leq \left(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\omega) d\omega\right)^{\frac{1}{\alpha+1}}.$$

The equality holds only when $f_\theta = f_{\theta_0}$ almost everywhere. From Assumption 1, the conclusion holds.

On the other hand, if $\alpha < 0$, then there are three cases (a) $-1 < \alpha < 0$, (b) $\alpha < -1$ and (c) $\alpha = -1$ must be considered. However, it is easy to see that both first two cases are corresponding to the case $p < 0$ and $0 < p < 1$ in Hölder's inequality since if $-1 < \alpha < 0$ then $(\alpha + 1)/\alpha < 0$ and if $\alpha < -1$ then $0 < (\alpha + 1)/\alpha < 1$. As a result, we obtain

$$D(f_\theta, f_{\theta_0}) \geq \left(\int_{-\pi}^{\pi} f_{\theta_0}^{\alpha+1}(\omega) d\omega\right)^{\frac{1}{\alpha+1}},$$

with a minimum from Assumption 1. For the case (c), the disparity is corresponding to the predictor error, which has a lower bound. (See Proposition 10.8.1 in Brockwell and Davis [3].) □

Proof of Theorem 3.1

Under Assumption 1 (iii), we simplify the notation as follows: for $1 \leq i \leq d$,

$$\begin{aligned}A_1(\theta) &= \int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\lambda) d\lambda, & B_1(\theta)_i &= f_\theta^{\alpha-1}(\omega)\partial_i f_\theta(\omega), \\ A_2(\theta)_i &= \int_{-\pi}^{\pi} f_\theta^\alpha(\lambda)\partial_i f_\theta(\lambda) d\lambda, & B_2(\theta) &= f_\theta^\alpha(\omega), \\ A_3(\theta)_{ij} &= \int_{-\pi}^{\pi} f_\theta^{\alpha-1}(\lambda)\partial_i f_\theta(\lambda)\partial_j f_\theta(\lambda) d\lambda, & C_1(\theta) &= \beta\left(\int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\lambda) d\lambda\right)^{\beta-1}.\end{aligned}$$

Proof. In view of (6), it is equivalent to consider $\hat{\theta}_n$ satisfies

$$\partial D(f_\theta, I_{n,X})\Big|_{\theta=\hat{\theta}_n} = 0.$$

The result that $\hat{\theta}_n \to \theta_0$ in probability follows that for any $\theta \in \Theta$ compact,

$$\partial D(f_\theta, I_{n,X}) \to \partial D(f_\theta, f_{\theta_0}) \quad \text{in probability},$$

which is guaranteed by Lemma 3.3A in Hosoya and Taniguchi [6]. Differentiating the disparity (4) with respect to θ, then we have

$$\partial D(f_\theta, I_{n,X}) = C_1(\theta) \int_{-\pi}^{\pi} (A_1(\theta)B_1(\theta) - A_2(\theta)B_2(\theta)) I_{n,X}(\omega) d\omega.$$

From $\partial D(f_\theta, f_{\theta_0})\Big|_{\theta=\theta_0} = 0$, asymptotic normality of the estimator follows Assumption 3, that is,

$$\partial D(f_\theta, I_{n,X})\Big|_{\theta=\theta_0} = C_1(\theta_0) \int_{-\pi}^{\pi} (A_1(\theta_0)B_1(\theta_0) - A_2(\theta_0)B_2(\theta_0))(I_{n,X}(\omega) - f_{\theta_0}(\omega)) d\omega,$$

converges to a random variable with normal distribution with mean 0 and variance $C_1(\theta_0)^2 V(\theta_0)$. Using $\partial D(f_\theta, f_{\theta_0})\Big|_{\theta=\theta_0} = 0$ again, we see that

$$\left| \partial^2 D(f_\theta, I_{n,X}) - C_1(\theta)\partial\Big(\int_{-\pi}^{\pi} (A_1(\theta)B_1(\theta) - A_2(\theta)B_2(\theta)) I_{n,X}(\omega) d\omega\Big)\Big|_{\theta=\theta_0} \right| \to 0$$

in probability. We also have

$$\begin{aligned}
\int_{-\pi}^{\pi} B_1(\theta) f_\theta(\omega) d\omega \, \partial A_1(\theta) - \int_{-\pi}^{\pi} A_2(\theta) f_\theta(\omega) d\omega \, \partial B_2(\theta) &= A_2(\theta) A_2(\theta)^{\mathrm{T}} \\
\int_{-\pi}^{\pi} A_1(\theta) f_\theta(\omega) d\omega \, \partial B_1(\theta) - \int_{-\pi}^{\pi} B_2(\theta) f_\theta(\omega) d\omega \, \partial A_2(\theta) &= -A_1(\theta) A_3(\theta),
\end{aligned}$$

and therefore

$$C_1(\theta)\partial\Big(\int_{-\pi}^{\pi} (A_1(\theta)B_1(\theta) - A_2(\theta)B_2(\theta)) I_{n,X}(\omega) d\omega\Big)\Big|_{\theta=\theta_0} \to C_1(\theta_0) H(\theta_0).$$

in probability. As a result, we obtain

$$\partial^2 D(f_\theta, I_{n,X})\Big|_{\theta=\theta_0} \to C_1(\theta_0) H(\theta_0).$$

in probability. Canceling $C_1(\theta_0)$, the desirable result is obtained. □

Proof of Theorem 3.2

Lemma 1 (Grenander and Rosenblatt [5], Kholevo [7]). *Let $A(\omega)$, $B(\omega)$ be $r \times s$ matrix-valued functions, and $g(\omega)$ a positive function on $\omega \in [-\pi, \pi]$. If*

$$\left\{ \int_{-\pi}^{\pi} B(\omega) B(\omega)^{\mathrm{T}} g(\omega)^{-1} d\omega \right\}^{-1}$$

exists, the following inequality

$$\int_{-\pi}^{\pi} A(\omega) A(\omega)^{\mathrm{T}} g(\omega) d\omega$$
$$\geq \left\{ \int_{-\pi}^{\pi} A(\omega) B(\omega)^{\mathrm{T}} d\omega \right\} \left\{ \int_{-\pi}^{\pi} B(\omega) B(\omega)^{\mathrm{T}} g(\omega)^{-1} d\omega \right\}^{-1} \left\{ \int_{-\pi}^{\pi} A(\omega) B(\omega)^{\mathrm{T}} d\omega \right\}^{\mathrm{T}} \quad \text{(A.1)}$$

holds. In (A.1)*, the equality holds if and only if there exists a constant matrix C such that*

$$g(\omega) A(\omega) + C B(\omega) = O, \quad \textit{almost everywhere } \omega \in [-\pi, \pi]. \quad \text{(A.2)}$$

Proof. The simplified asymptotic variance $\tilde{V}(\theta_0)$ follows Lemma 2.1 in Hosoya and Taniguchi [6]. The inequality is derived as follows. Define

$$\begin{aligned} A(\omega) &= A_1(\theta_0) B_1(\theta_0) - A_2(\theta_0) B_2(\theta_0), \\ B(\omega) &= \partial f_\theta(\omega) \Big|_{\theta=\theta_0}, \\ g(\omega) &= f_\theta^2(\omega) \Big|_{\theta=\theta_0}. \end{aligned}$$

Then the inequality in Theorem 3.2 holds from Lemma 1. According to (A.2), the equality holds when

$$\int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\lambda) d\lambda \, f_\theta^{\alpha+1}(\omega) \partial f_\theta(\omega) - \int_{-\pi}^{\pi} f_\theta^{\alpha}(\lambda) \partial f_\theta(\lambda) d\lambda \, f_\theta^{\alpha+2}(\omega) - C \partial f_\theta(\omega) \Big|_{\theta=\theta_0} = 0 \quad \text{(A.3)}$$

with a constant c. If $\alpha = -1$, then the left hand side of (A.3) is

$$2\pi \partial f_\theta(\omega) - \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) \partial f_\theta(\lambda) d\lambda \, f_\theta(\omega) - C \partial f_\theta(\omega) \Big|_{\theta=\theta_0}.$$

However, Kolmogorov's Formula tells us

$$\int_{-\pi}^{\pi} \log f_\theta(\lambda) d\lambda \Big|_{\theta=\theta_0} = \log \frac{\sigma^2}{2\pi},$$

which is followed by

$$\int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) \partial f_\theta(\lambda) d\lambda \Big|_{\theta=\theta_0} = 0.$$

If we choose $c = 2\pi$, then the equality holds. If $\alpha \neq -1$, then (A.3) does not hold generally. It is easy to see that (A.2) holds if the spectral density does not depend on ω. □

Proof of Theorem 3.3

Proof. We only need to change $g(\lambda, \beta)^{-1}$ in Mikosch et al. [10] by $a(\theta) f_{\theta_0}^{\alpha}(\omega)$ to show the statements. □