Statistics and Probability Letters 79 (2009) 851–856

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Statistical signal extraction using stable processes

N. Balakrishna *, G. Hareesh

Department of Statistics, Cochin University of Science and Technology, Kochi – 682 022, India

ARTICLE INFO

Article history:
Received 21 February 2008
Received in revised form 10 November 2008
Accepted 10 November 2008
Available online 19 November 2008

ABSTRACT

The standard models for statistical signal extraction assume that the signal and noise are generated by linear Gaussian processes. The optimum filter weights for those models are derived using the method of minimum mean square error. In the present work we study the properties of signal extraction models under the assumption that signal/noise are generated by symmetric stable processes. The optimum filter is obtained by the method of minimum dispersion. The performance of the new filter is compared with their Gaussian counterparts by simulation.

1. Introduction

The statistical signal extraction theory developed by Wiener (1949) and Kolmogorov (1941) assumes that the data are generated by a stationary stochastic process and that they form a lengthy sequence. The observed data process Y_t is often depicted as a combination of signal X_t and noise N_t as follows:

$$Y_t = X_t + N_t. \quad (1)$$

The objective here is to use the data on Y_t to estimate the unobserved component series X_t and N_t. Most of the theoretical developments in statistical signal extractions assume that the signal and/or noise follow certain Auto-Regressive Moving Average (ARMA) models with Gaussian or other exponential family of distributions as marginal with finite second and higher order moments. Then the signals are expressed as linear filters of the observations, where the optimum filter weights are obtained using the method of Minimum Mean Square Error (MMSE), see Bell and Martin (2004) for details. However, in many practical instances such as communication, economics and finance, network traffic, data shows sharp spikes or occasional bursts of outlying observations and heavy tailed distributions such as symmetric stable can be used to model such series. Detailed discussion on such distributions and their applications in the cited areas may be found in Alder et al. (1998).

Our objective in this paper is to discuss the properties of the signal extraction model specified by (1) when the signal and noise are assumed to follow stationary ARMA models with symmetric stable marginal distributions. We use the method of minimum dispersion introduced by Cline and Brockwell (1985) to obtain the optimal filter for the signal since the MMSE technique cannot be used due to the non-existence of moments of the symmetric stable distribution.

A random variable X is said to have a symmetric stable distribution and we denote it by $X \sim S_\alpha(\lambda)$ if its characteristic function is of the form

$$\varphi_x(t) = \exp(-\lambda|t|^\alpha),$$

where, $\alpha \in (0, 2]$, measuring the tail thickness, $\lambda > 0$ the scale (dispersion) parameter.

We are interested in studying the aspects of signals X_t using ARMA (p, q) model with symmetric stable innovations introduced by Brockwell and Davis (1991) and is defined by

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-1} + \cdots + \theta_q e_{t-q}, \quad (2)$$

* Corresponding author. Tel.: +91 484 2575893; fax: +91 484 2577595.
E-mail addresses: nb@cusat.ac.in, balajicusat@yahoo.com (N. Balakrishna).

0167-7152/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2008.11.006

where $\{e_t\}_{t=-\infty}^{\infty}$ is an iid sequence of random variables and the polynomials

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \quad \text{and} \quad \theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q, \tag{3}$$

satisfy the condition $\phi(B)\theta(B) \neq 0$ such that $|B| \leq 1$, where B is the shift operator defined as $B^k X_t = X_{t-k}$. The polynomial $\phi(B)$ is the auto-regressive operator and $\theta(B)$ is the moving average operator. It follows that (2) has a unique stationary solution as discussed in Cline and Brockwell (1985), namely

$$X_t = \sum_j \psi_j e_{t-j}. \tag{4}$$

The weights $\{\psi_j\}$ are determined by comparing the coefficients of B^j in the power series expansion $\sum_j \psi_j B^j = \frac{\theta(B)}{\phi(B)}$, $|B| \leq 1$. If e_t has finite variance, predictors $\hat{X}_t$ are determined by minimizing $E(X_t - \hat{X}_t)^2$, the expected squared error. For processes with infinite variance however, an alternative criterion such as minimum dispersion for selection of the best predictor is needed, which we discuss in Section 3.

Section 2 introduces the mathematical representation of signal and noise processes. Section 3 includes the minimum dispersion signal extraction criteria and we discuss how symmetric stable distribution can be embedded into this frame work. Section 4 derives a finite length filter using the state space form of the model. Section 5 contains some simulation results.

2. Statistical models for signal extraction

Suppose that an observed stationary linear time series $\{Y_t\}$ can be written using (1) with X_t and N_t following the ARMA models of the type

$$\begin{aligned} \phi(B)X_t &= \theta(B)a_t, \\ \phi(B)N_t &= \theta(B)b_t, \end{aligned} \tag{5}$$

where $\phi(B)$ and $\theta(B)$ are the polynomials defined by (3). We assume that $\{a_t\}$ and $\{b_t\}$ are mutually independent symmetric stable noise processes with scale parameters λ_a and λ_b respectively. This in turn implies that X_t and N_t are independent. Our objective is to obtain an estimate $\hat{X}_t$ of X_t and $\hat{N}_t$ of N_t by filtering Y_t as

$$\begin{aligned} \hat{X}_t &= W(B)Y_t, \\ \hat{N}_t &= (1 - W(B))Y_t, \end{aligned} \tag{6}$$

where $W(B) = \sum_j w_j B^j$. We summarize the above discussion in the following proposition.

Proposition 2.1. *Suppose that an unobserved series $\{Y_t\}$ has the representation: $Y_t = X_t + N_t$. If we estimate X_t by $\hat{X}_t$ of the form (6), the estimation error is given by*

$$\zeta_t = W(B)X_t - (1 - W(B))N_t. \tag{7}$$

The two components of ζ_t are mutually independent if and only if the components of Y_t are mutually independent.

3. Minimum dispersion criteria for signal estimation

Signal extraction procedure consists of finding an optimal filter which minimizes the signal extraction error. In finite variance case optimal filter is the one which minimizes the MSE where as in symmetric stable process we propose minimum dispersion criteria. When ζ_t has a symmetric stable distribution, the minimization of error dispersion is equivalent to minimization of the scale parameter of the error distribution (see Brockwell and Davis (1991)). Thus if e_t's are iid symmetric stable random variables with index α and if $\sum_{j=-\infty}^{\infty} |\beta_j|^\alpha < \infty$, then $Y = \sum_{j=-\infty}^{\infty} \beta_j e_j$ is also symmetric stable with

$$\text{disp}(Y) = \sum_{j=-\infty}^{\infty} |\beta_j|^\alpha. \tag{8}$$

Theorem 1.1 by Cline and Brockwell (1985) indicates that the prediction error dispersion is roughly proportional to the probability of a large prediction error. From (4) and (5) we can write the moving average representation for the signal and noise respectively as

$$X_t = \sum_{j=0}^{\infty} \psi_j^x a_{t-j}, \qquad N_t = \sum_{j=0}^{\infty} \psi_j^n b_{t-j},$$

where ψ_j^x and ψ_j^n are the weights obtained for X_t and N_t respectively.

Thus by (7), the error process is

$$\zeta_t = \sum_{j=0}^{\infty} \left(\sum_{k=0}^{j} w_k \psi^x_{j-k} - \psi^x_j \right) a_{t-j} + \sum_{j=0}^{\infty} w_k \psi^n_{j-k} b_{t-j}.$$

Using (8) and the distributional properties of a_t and b_t we can show that the dispersion of the error process is

$$\mathrm{Disp}(\zeta_t) = \sum_{j=0}^{\infty} \left| \sum_{k=0}^{j} w_k \psi^x_{j-k} - \psi^x_j \right|^{\alpha} \lambda_a + \sum_{j=0}^{\infty} \left| \sum_{k=0}^{j} w_k \psi^n_{j-k} \right|^{\alpha} \lambda_b. \tag{9}$$

Finding optimal filter is equivalent to finding the weights w_k in $W(B)$ which minimizes (9). In general the solution does not have a closed form, but it gives some satisfactory results for some special cases. For this, however, we need the following Theorem.

Theorem 3.1. *For $1 < \alpha \le 2$, the optimal filter weights, $\{w_j\}$ which minimizes (9), is the solution of the system of equations,*

$$\frac{\partial \mathrm{Disp}(\zeta_t)}{\partial w_k} = 0, \quad k = 0, 1, 2, \ldots. \tag{10}$$

When $\alpha \le 1$, general expressions do not exist.

Proof. The proof follows from Lemma 3.1 of Cline and Brockwell (1985). □

From the above discussion it is clear that, we have to adopt some numerical methods for getting optimum filter weights and the signal estimate. This filter reduces to Gaussian Filter when $\alpha = 2$ and the details on the latter may be found in Cline and Brockwell (1985).

The semi-infinite filter discussed so far in this section may be generalized to doubly infinite and asymmetric filters studied in the literature. The former uses future as well as the past of $\{Y_t\}$ for estimating X_t but the latter filter estimates X_t based on given data Y_u up through $u = t - m$, for finite m. In order to apply this method we can modify our filter as:

$$W(B) = \sum_{-m}^{\infty} w_j B^j.$$

Similar to (10) the optimum filter minimizes the dispersion of error process,

$$\mathrm{Disp}(\zeta_t) = \sum_{j=-m}^{\infty} \left| \sum_{k=-m}^{j} w_k \psi^x_{j-k} - \psi^x_j \right|^{\alpha} \lambda_a + \sum_{j=-m}^{\infty} | \sum_{k=-m}^{j} w_k \psi^n_{j-k} |^{\alpha} \lambda_b. \tag{11}$$

Doubly infinite filter is a symmetric filter, which can be obtained by letting $m \to \infty$.

When $\alpha = 2$ the error dispersion in (11) reduces to the mean square error and the optimal filter reduces to the asymmetric Wiener–Kolmogorov filter (see Bell and Martin (2004)).

4. State space representation and Kalman–Levy filtering

So far we have discussed the infinite length filter, but in practice we have only finite length of observations. In this section we introduce a finite length filtering algorithm based on state space representation and Kalman–Levy filtering. This can be considered as an improvement over the stable filter defined in Section 3. The classical approach of Kalman filtering assumes that the underlying models are linear and the innovations are Gaussian. Kalman–Levy filter is a generalized version of Kalman filter for heavy tailed processes. In the present section we discuss the finite length signal extraction filter for such processes with symmetric stable noise. A linear dynamic system of state variable x_t can be represented as

$$x_{k+1} = Z_k x_k + \eta_k, \quad k = 0, 1, \ldots, \tag{12}$$

and the observations y_k follow the equation:

$$y_k = T_k x_k + \varepsilon_k, \tag{13}$$

where $\{Z_k\}$ and $\{T_k\}$ are assumed to be known sequences of real numbers. Further we assume that the dynamic noise $\{\eta_k\}$ and the observational noise $\{\varepsilon_k\}$ are mutually independent iid symmetric stable sequences with scale factors λ^{η} and λ^{ε} respectively. For $1 < \alpha \le 2$, the predictor of the state variable is defined as $x_{k|k-1} = E(x_k|y_{k-1})$ and the filter is $x_{k|k} = E(x_k|y_k)$. The Kalman–Levy filtering algorithm by Sornette and Ide (2001) provides a sequential procedure for estimating the unobserved state variable x_t and the solution is obtained by sequential prediction and filtering as

$$x_{k|k-1} = Z_{k-1} x_{k-1|k-1}, \tag{14}$$

which determines the forecast of x_k from a given initial condition $x_{0|0}$. The forecast is based on the filtering performed at the previous step. This forecast is then used to find a new filter $x_{k|k}$ which will be mixed with the observed information y_k and given by

$$x_{k|k} = x_{k|k-1} + K_k(y_k - T_k x_{k|k-1}), \tag{15}$$

where K_k is called the Kalman–Levy gain which is obtained by minimizing the scale factor of the filtering error process and is given by $K_k = T_k^{-1} / \left(1 + (\Delta_k)^{\alpha/(\alpha-1)}\right)$, with modified relative error ratio, $\Delta_k = (\lambda_k^{\varepsilon})^{1/\alpha} / \left[T_k(\lambda_{k|k-1})^{1/\alpha}\right]$.

From the models (12) and (13) the finite length filter may be defined as

$$\hat{x}_n = w_0 + \sum_{j=1}^{n} w_j Y_j, \tag{16}$$

where $w_j, j = 0, 1, 2, \ldots, n$ are the filter weights whose expressions are to be obtained using Kalman–Levy filter. In this case the signal extraction problem can be divided into that of prediction and filtering. Under the prediction problem we estimate the future state of the signal from a given initial value and the observed signal. The Kalman–Levy predictor for the models (12) and (13) is given by (14) and may be expressed as

$$x_{k+1|k} = \sum_{j=1}^{n} L_j K_j y_j + L_0 x_{1|0}, \tag{17}$$

where, $L_n = 1, L_j = N_n N_{n-1} \ldots N_{j+1}$, with $N_j = Z_j - K_j T_j, j = 0, 1, 2, \ldots, n-1$.

Comparing (16) and (17) we get, $w_j = L_j K_j$ for $j = 1, 2, \ldots, n$ and $w_0 = L_0 x_{1|0}$.

The filtering problem deals with the estimation of the present state of the signal from a given initial condition and the observed signal at that time. As before the Kalman–Levy filter for the model (12) and (13) given by (15) may be written as

$$x_{k|k} = \sum_{j=1}^{n} L_j K_j y_j + L_0 x_{1|0}, \tag{18}$$

where, $L_n = 1, L_{n-1} = I - K_n T_n$ and $L_j = L_{n-1} N_{n-1} \ldots N_{j+1}, j = 0, 1, 2, \ldots, n-2, N_j = Z_j - K_j T_j$, then comparing (16) and (18) we get $w_j = L_j K_j$ for $j = 1, 2, \ldots, n$ and $w_0 = L_0 x_{1|0}$.

Remark. The finite length filter discussed above can be extended to the case of multivariate filter and predictor with appropriate modifications of Sornette and Ide (2001). State space representation of model (1) can be used to extract the signal and noise defined in (5) by applying the multivariate filter and predictor.

5. Simulation

Suppose that an observed time series Y_t evolves according to Eq. (1) and the unobserved signal X_t, is a moving average process which can be represented as follows:

$$X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}.$$

Also assume that $a_t \sim S_\alpha(\lambda_a)$ and $N_t \sim S_\alpha(\lambda_e)$ where λ_a, λ_e are respectively the dispersion parameter of a_t and N_t. Our main objective is to extract the signal X_t from the given observed signal Y_t. This problem can be solved by applying the methods discussed in Section 3. We simulate the above model by taking $q = 4$, and $\theta_1 = 0.7, \theta_2 = 0.4, \theta_3 = 0.2, \theta_4 = 0.1, \alpha = 1.5$, $\lambda_a = 5, \lambda_e = 3$. The symmetric stable innovation sequences $\{a_t\}$ and $\{N_t\}$ are generated using the algorithm given by Alder

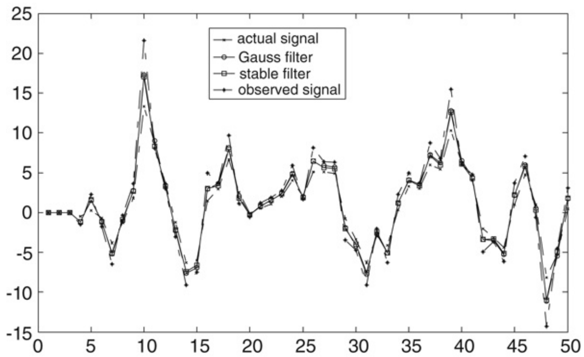

Fig. 1. Signal extraction.

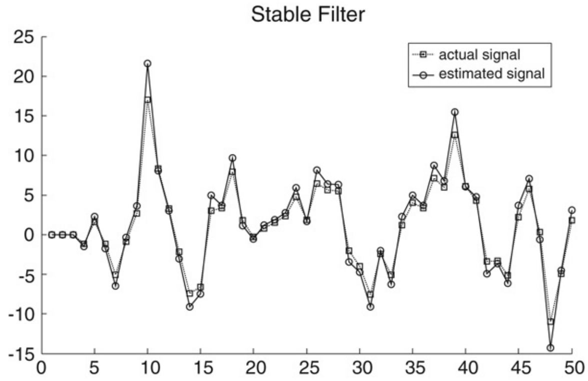

Fig. 2. Stable filter.

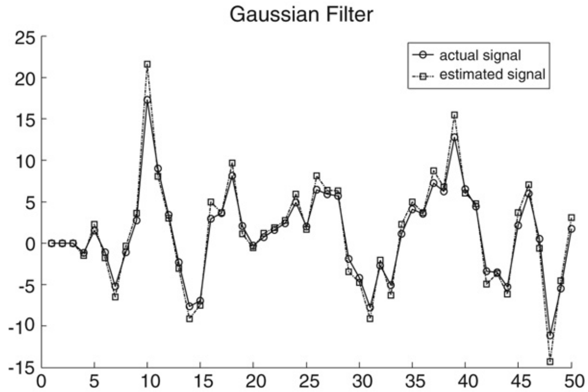

Fig. 3. Gaussian filter.

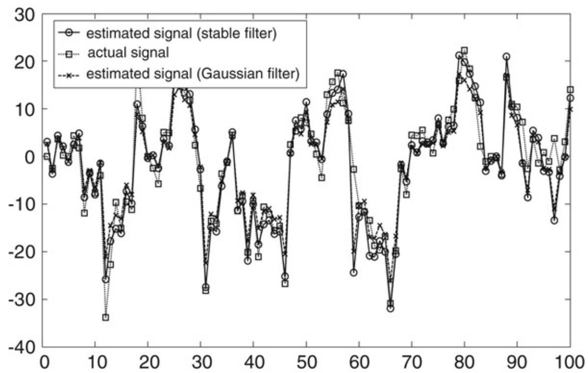

Fig. 4. Error plot.

Table 1
Comparison of error sum of squares and the dispersion.

Filter $N = 100$	Dispersion	Error sum of squares
Gaussian	225.3723	282.0869
Stable	218.7023	258.6477
$N = 50$		
Gaussian	116.0728	223.0761
Stable	113.0230	212.4247

et al. (1998). The above computations are compared with the Gaussian Filter in terms of their error sum of squares and dispersion which are summarized in Table 1 for different sample sizes.

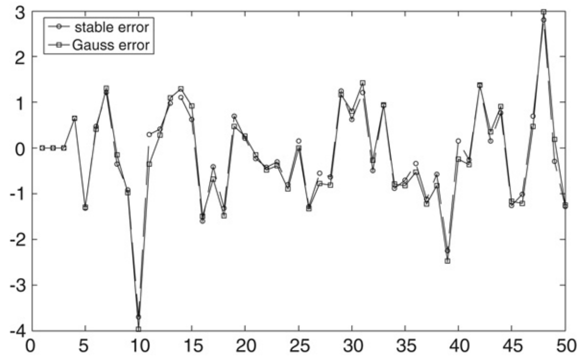

Fig. 5. Finite length filtering.

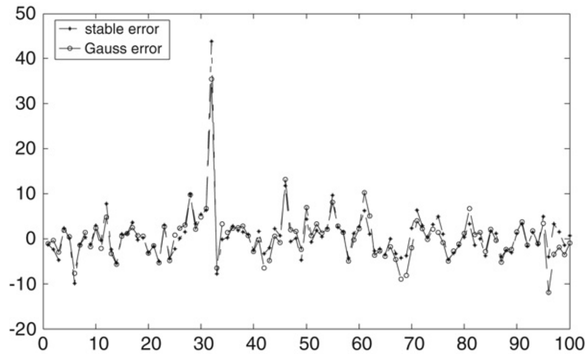

Fig. 6. Error plot.

The error sum of squares and the error dispersion of these filters show the improvement of stable filter against Gaussian filter. Figs. 1–3 give the plots corresponding to the actual signal, observed signal under noisy environment, estimated signal through stable and Gaussian filters. Fig. 4 shows the plot of signal extraction error under stable filter and Gaussian filter. Figs. 5 and 6 give the plot for the finite length filtering signals and error plot respectively. Here we assume that the signal X_t follows an AR (1) model with $\rho = 0.7$, innovation sequences $a_t \sim S_\alpha(\lambda_a)$ and noise $N_t \sim S_\alpha(\lambda_e)$.

6. Conclusion

In this work we present a linear filtering method for computing the filter weights assigned to the observation for estimating unobserved signal under general noisy environment. Here we consider both the signal and the noise as stationary processes with infinite variance innovations. Nonlinear filtering such as particle filtering and sequential Bayesian filtering are some other alternatives to these methods. We will consider these approaches in the forthcoming work.

Acknowledgement

We thank the referee for several suggestions on the earlier draft of the paper which led to the present improved version.

References

Alder, R.J, Feldman, R.E, Taqqu, M.S., 1998. A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Birkhäuser, Boston.
Bell, W.R., Martin, D.E.K., 2004. Computation of asymmetric signal extraction filters and mean squared error for ARIMA component models. Journal of Time Series Analysis 25, 603–635.
Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, 2nd ed.. Springer-Verlag, New York.
Cline, D.B.H., Brockwell, P.J., 1985. Linear prediction of ARMA processes with infinite variance. Stochastic Processes and their Applications 19, 281–296.
Kolmogorov, A.N., 1941. Interpolation and extrapolation. Bull. Acad. Sci. USSR., Ser. Math. 5, 3–14.
Sornette, D.K, Ide, 2001. The Kalman–Levy filter. Physica D 151, 142–174.
Wiener, N., 1949. Extrapolation, interpolation and smoothing of stationary time series. In: Report on the Service Research Project DIC – 6037. MIT Technology Press, Wiley, New York.