Computational Statistics and Data Analysis 57 (2013) 125–140

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Robust estimation for the covariance matrix of multivariate time series based on normal mixtures

Byungsoo Kim, Sangyeol Lee *

Department of Statistics, Seoul National University, Seoul, 151-742, Republic of Korea

ARTICLE INFO

Article history:
Received 24 November 2011
Received in revised form 7 May 2012
Accepted 12 June 2012
Available online 21 June 2012

Keywords:
Density-based divergence measures
Robust estimation
Autocovariance function
Consistency
Asymptotic normality

ABSTRACT

In this paper, we study the robust estimation for the covariance matrix of stationary multivariate time series. As a robust estimator, we propose to use a minimum density power divergence estimator (MDPDE) designed by Basu et al. (1998). To supplement the result of Kim and Lee (2011), we employ a multivariate normal mixture family instead of a multivariate normal family. As a special case, we consider the robust estimator for the autocovariance function of univariate stationary time series. It is shown that the MDPDE is strongly consistent and asymptotically normal under regularity conditions. Simulation results are provided for illustration. A real data analysis applied to the portfolio selection problem is also considered.

1. Introduction

The maximum likelihood estimator (MLE) for normal mixture models has been studied by many authors; for instance, we can refer the reader to Sundberg (1974), Laird (1978), Redner (1981), Lindsay (1983), Redner and Walker (1984), Hathaway (1985), and the articles cited therein. It is widely appreciated that MLE shows very poor performance either when outliers exist or the likelihood function explodes as in such a case that one of the means in the model equals one of the data and the corresponding variance is close to 0. To cope with such defect, the research has developed the minimum distance estimators based on the Wolfowitz distance (Choi, 1969), the Cramer–von Mises distance (Woodward et al., 1984), the squared L_2 norm of cumulative distribution function (Clarke and Heathcote, 1994), the minimum Hellinger distance (Cutler and Cordero-Braña, 1996) and L_2 distance of the density function (Scott, 2001). Unlike the others, the minimum Hellinger distance has the asymptotic efficiency as MLE achieves when observations follow hypothesized models under consideration. However, this method has a drawback of requiring to use some nonparametric smoothing methods, where one possibly encounters rather a demanding problem like the selection of bandwidth.

Estimation of the autocovariance function (ACF) has been a core issue in time series analysis since ACF stands for the dependence structure of time series and the estimation of ACF is closely connected with a model selection problem. Due to its importance, some authors studied the robust estimation for ACF in univariate time series; for instance, see Ma and Genton (2000). Ma and Genton's estimator is proven to produce a highly robust estimator for the ACF. However, it also has a shortcoming that the normalizing factor in their estimator must be chosen differently according to the underlying distribution of given data. To overcome this defect, Kim and Lee (2011) proposed to use the minimum density power divergence estimator (MDPDE) designated by Basu et al. (1998) (BHHJ) in the ACF estimation problem and demonstrated a superiority to Ma and Genton's estimator. The MDPDE is proven to have strong robust properties with low loss in the asymptotic efficiency relative to the MLE under various circumstances. As a relevant paper, we refer the reader to

* Corresponding author.
E-mail addresses: sylee@stats.snu.ac.kr, jpgrslee@yahoo.com (S. Lee).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.06.012

Fujisawa and Eguchi (2006), who show that the objective function of MDPDE is bounded under mild conditions in iid univariate normal mixture models.

The objective of this paper is to provide a robust estimator for the mean and covariance of multivariate time series. In this study, we use the MDPDE method but employ a multivariate normal mixture family instead of a multivariate normal family since according to the result of Kim and Lee (2011), the normal distribution approach does not perform well when the distributions of data is far from a normal distribution. It will be shown that the normal mixture approach performs more properly in terms of both the efficiency and robustness. Although we emphasize the robust estimation for the mean and covariance matrix of multivariate time series, if the true distribution of data belongs to the multivariate normal mixture family, we can also provide the robust estimators for normal mixture parameters.

This paper is organized as follows. In Section 2, we introduce the construction of the robust estimator using the BHHJ's procedure. In Section 3, we show asymptotic properties of the MDPDE and its robustness by analyzing the influence function. In Section 4, we apply our method to the estimation of parameters of multivariate normal mixture time series and compare its performance with that of the MLE. Further, we conduct a simulation study to compare the performance of the proposed estimator for the ACF with the sample autocovariance function (SACF). In Section 5, we apply our method to the portfolio optimization problem by using Dow Jones Industrial average data. In Section 6, we provide the proofs.

2. MDPDE with multivariate normal mixture family

Consider a parametric family of models $\{F_\theta\}$, indexed by the unknown parameter $\theta \in \Theta \subset R^\rho$, possessing densities $\{f_\theta\}$ with respect to the Lebesgue measure, and let $\mathcal{G}$ be the class of all distributions having densities with respect to the Lebesgue measure. For estimating the unknown parameter θ, BHHJ introduced a family of density power divergences

$$d_\alpha(g,f) := \begin{cases} \displaystyle\int \left\{ f^{1+\alpha}(z) - \left(1+\frac{1}{\alpha}\right) g(z) f^\alpha(z) + \frac{1}{\alpha} g^{1+\alpha}(z) \right\} dz, & \alpha > 0, \\ \displaystyle\int g(z)(\log g(z) - \log f(z))dz, & \alpha = 0, \end{cases}$$

where g and f are density functions, and defined the minimum density power divergence functional $T_\alpha(\cdot)$ for G in $\mathcal{G}$ by

$$d_\alpha(g, f_{T_\alpha(G)}) = \min_{\theta\in\Theta} d_\alpha(g, f_\theta),$$

where g is the density of G. Note that if G belongs to $\{F_\theta\}$, $T_\alpha(G) := \theta_\alpha = \theta$ for some $\theta \in \Theta$. Based on these, given the random sample $X_1, \ldots, X_n$ with unknown density g, they defined the MDPDE as

$$\hat\theta_{\alpha,n} = \operatorname*{argmin}_{\theta\in\Theta} H_{\alpha,n}(\theta), \tag{2.1}$$

where $H_{\alpha,n}(\theta) = \frac{1}{n}\sum_{t=1}^n V_\alpha(\theta; X_t)$ and

$$V_\alpha(\theta; x) = \begin{cases} \displaystyle\int f_\theta^{1+\alpha}(z)dz - \left(1+\frac{1}{\alpha}\right) f_\theta^\alpha(x), & \alpha > 0, \\ -\log f_\theta(x), & \alpha = 0. \end{cases} \tag{2.2}$$

BHHJ demonstrated that the estimator is robust against outliers but still has a high efficiency when the true density belongs to the parametric family $\{F_\theta\}$ and α is close to 0. Note that when $\alpha = 0, 1$, MDPDE is the same as the MLE and L_2 the distance estimator respectively.

In this paper, we study the MDPDE for the mean and covariance matrix of a d-dimensional strictly stationary and ergodic time series $\{\boldsymbol{X}_t, t = 1, 2, \ldots\}$. Since the $\alpha > 1$ case can cause a great loss of efficiency for some basic models as described by Basu et al., we focus on the case $0 < \alpha \le 1$. In order to obtain the MDPDE for the mean and covariance matrix of $\boldsymbol{X}_t$, we consider a d-dimensional multivariate normal mixture parametric family in BHHJ's procedure. Let $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ be the set of m-component d-dimensional multivariate normal mixture densities of the form

$$f_\theta(\boldsymbol{x}) = \sum_{j=1}^m \omega_j \phi(\boldsymbol{x}; \mu_j, \Sigma_j),$$

where m is known and $\phi(\boldsymbol{x}; \mu_j, \Sigma_j) = \frac{1}{\sqrt{2\pi}^d |\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu_j)'\Sigma_j^{-1}(\boldsymbol{x}-\mu_j)\right)$ satisfies for $j = 1, \ldots, m$,

$$\Sigma_j \text{ is symmetric, } 0 < \frac{1}{n}\frac{(1+\alpha)^{d/2+1}}{\alpha} \le \min_j \omega_j \le 1, \qquad \sum_{j=1}^m \omega_j = 1, \qquad \|\mu_j\| \le c_1 < \infty,$$

$$0 < c_2 \le \lambda_{\min}(\Sigma_j) \le \lambda_{\max}(\Sigma_j) \le c_3 < \infty \quad \text{for some positive constants } c_1, c_2, c_3, \tag{2.3}$$

where $\lambda_{\min}(\Sigma_j)$ and $\lambda_{\max}(\Sigma_j)$ denote the minimal and maximal eigenvalues of Σ_j.