Statistics and Its Interface Volume 11 (2018) 31–40

Maximum smoothed likelihood estimation for a class of semiparametric Pareto mixture densities

Mian Huang*, Shaoli Wang†, Hansheng Wang‡, and Tian Jin§

Motivated by an analysis of Return On Equity (ROE) data, we propose a class of semiparametric mixture models. The proposed models have a symmetric nonparametric component and a parametric component of Pareto distribution with unknown parameters. However, situations with general parametric components other than Pareto distribution are also investigated. We prove that these mixture models are identifiable, and establish a novel estimation procedure via smoothed likelihood and profile-likelihood techniques. For ease of computation, we develop a new EM algorithm to facilitate the maximization problem. We show that this EM algorithm possesses the ascent property. A rule-of-thumb based procedure is proposed to select the bandwidth of the nonparametric component. Simulation studies demonstrate good performance of the proposed methodology. Furthermore, we analyze the ROE dataset which may consist of real and manipulated earnings. Our analysis reveals significant earning manipulation in the Chinese listed companies from a quantitative perspective using the proposed model.

AMS 2000 subject classifications: Primary 62G05.
Keywords and phrases: Semiparametric mixture models, EM algorithm, Maximum smoothed likelihood estimation, Pareto mixture densities, Bandwidth selection.

1. INTRODUCTION

Return on Equity (ROE) is a widely used measurement of a company's efficiency at generating profits per equity unit, which equals to a fiscal year's net income divided by the total equity. Earning manipulation has been noted in several previous researches. In the U.S., firms tend to manipulate earnings to turn a small loss into a small profit [5]. For the Chinese listed companies, Ding et al. [8] noted that there is strong evidence of manipulations to dramatically boost earnings. When part of the companies manipulate their earnings, the population distribution of ROE exhibits certain abnormality. Figure 1 shows the histogram of the ROEs of Chinese listed firms in the year 2010, which are obviously asymmetric, having a big jump on the right side of the zero point. The unusual jump is mainly attributed to the "Special Treatment"(ST) policy on Chinese stock markets [15]. As mandated by Chinese securities laws, a stock will be put into "Special Treatment" status if the company reports accounting losses in two consecutive years, and ST stocks are under strict regulations by the China securities regulatory commission. Therefore, in order to meet regulatory requirements and avoid ST, Chinese companies may manipulate to dramatically boost earnings.

As part of the companies may manipulate earnings, the reported ROE is actually comprised of two different components, the real earnings and the manipulated earnings. This provides a natural connection to a two-component mixture model framework. Mixture models have been widely used in economics, finance, biology, medicine, etc. As parametric mixture models may not be capable in describing complicated datasets, there have been increasing studies that focus on nonparametric and semiparametric mixture models; see among others, Hall and Zhou [13], Hunter et al. [14], Bordes et al. [2], Bordes et al. [3], Maiboroda and Sugakova [21], and Levine et al. [18]. Bordes et al. [2] considers a two-component mixture model

$$f(x) = \lambda g(x) + (1-\lambda)p(x-\mu), \tag{1}$$

where $g(\cdot)$ is a known density function, $p(\cdot)$ is an unknown nonparametric density function that is symmetric about 0, $\lambda \in (0,1)$ and $\mu \in \mathbb{R}$. It is shown in Bordes et al. [2] that model (1) is identifiable under some moment conditions. Bordes et al. [2], Bordes and Vandekerkhove [4], Maiboroda and Sugakova [21] and Maiboroda and Sugakova [22] proposed $\sqrt{n}$-consistent estimators for the parameters λ and μ, and $\sqrt{nh}$-consistent estimators for the nonparametric density $p(\cdot)$, and proved asymptotic normality of these estimators. When $g(x)$ is unknown, Ma and Yao [20] investigated a general model where $g(x) \equiv g(x;\boldsymbol{\xi})$, a parametric density with unknown parameter $\boldsymbol{\xi}$, and proposed a semiparametric efficient estimation procedure via estimating equations. However, the model identifiability depends on the form of $g(x;\boldsymbol{\xi})$, and results are only obtained when $\boldsymbol{\xi}$ does not appear in $g(x;\boldsymbol{\xi})$.

In this paper, we propose analyzing the Chinese ROE data using model (1) with Pareto distribution $g(x) \equiv g(x;\boldsymbol{\xi})$, that is, a semiparametric mixture of a symmetric

*Associate Professor at Shanghai University of Finance and Economics.
†Corresponding author. Associate Professor at Shanghai University of Finance and Economics.
‡Professor at Peking University.
§Research Scientist at Shanghai Stock Exchange.

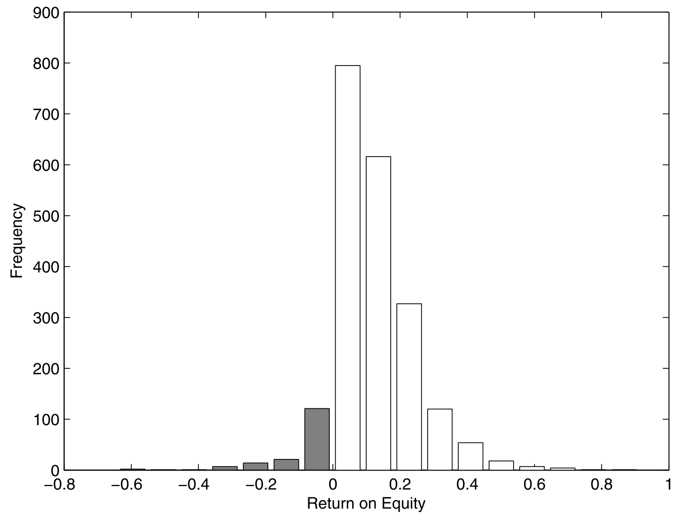


Figure 1. Histogram of Return On Equity (ROE) of Chinese Firms, Year 2010.

nonparametric distribution and a Pareto distribution with unknown parameters; the nonparametric component $p(\cdot)$ characterizes the real earnings, while the Pareto component $g(x;\boldsymbol{\xi})$ describes the manipulated earnings. Note that Pareto distribution has a jump at the left ending point of its support, so does the resulting mixture distribution, which fits the ROE data nicely.

As there is an unknown nonparametric component in the mixture density, it is very difficult to estimate the parameters and the unknown distribution jointly. To overcome this difficulty, we employ the maximum smoothed likelihood estimation (MSLE) for density that was firstly introduced by Eggermont and LaRiccia [11]. In MSLE, the unknown density is smoothed by a nonlinear smoothing operator, resulting in a smoothed likelihood. Eggermont and LaRiccia [10] and Eggermont and LaRiccia [11] investigated the consistency of the MSLE and showed that the estimator converges to the true density in probability under suitable conditions. Eggermont [9] proved that the smoothing operator is strictly concave, and the maximization of the smoothed log-likelihood admits a unique smooth solution under some regularity conditions. Levine et al. [18] studied the MSLE for multivariate mixtures, where all the components are nonparametric unknown densities. For model (1), since $g(\cdot)$ is a parametric density, we only smooth the nonparametric component $p(\cdot)$ and obtain the smoothed likelihood. To ensure a symmetric estimator for the nonparametric component, we employ a symmetric smoothing kernel [16] in the nonlinear smoothing operator.

The maximization of the smoothed likelihood of the proposed model can be carried out via an EM algorithm. EM algorithm consists of iterative steps to maximize a likelihood with missing data, and the resulting likelihood is nondecreasing on each iteration. Comprehensive references include Dempster et al. [7], Wu [30], and McLachlan and Krishnan [23].

In this paper, we propose an EM algorithm to facilitate the maximization of the smoothed likelihood. We further prove that the proposed EM algorithm enjoys the ascent property. We conduct simulation studies to examine the performance of the proposed methodology, and propose a two-step procedure for bandwidth selection. To summarize, we attempt to make the following three major contributions to the existing literature. Firstly, we propose a semiparametric Pareto mixture density, and prove the identifiability under mild conditions; Secondly, we derive an smoothed likelihood as the objective function by applying a nonlinear smoothing operator on the nonparametric component of the mixture which ensures nonnegative and symmetric estimation of the component; and lastly, we propose an efficient estimation procedure with the aid of an EM algorithm, and prove the ascent property of the EM algorithm.

The rest of the article is organized as follows. Section 2 is devoted to the identifiability problem and the smoothed likelihood method. In Section 3, we propose an estimation procedure for the model, and show that the corresponding EM algorithm possesses the ascent property with respect to the smoothed likelihood. Simulations and an application to the ROE dataset are given in Section 4. A discussion is given in Section 5 and technical proofs are given in the Appendix.

2. IDENTIFIABILITY AND SMOOTHED LIKELIHOOD

2.1 Model and identifiability

Consider a two-component semiparametric mixture model

$$f(x) = \lambda g(x;\boldsymbol{\xi}) + (1-\lambda)p(x-\mu), \tag{2}$$

where $\lambda \in (0,1)$ is the mixing proportion parameter, $g(x;\boldsymbol{\xi})$ is a parametric density with parameter $\boldsymbol{\xi}$, $p(\cdot)$ is an unknown nonparametric density that is symmetric about 0, and μ is a location parameter. Model (2) with a general parametric density g will be discussed. However, as our model is motivated by the ROE data of Chinese stocks, we will mainly focus on a special case of model (2) with Pareto density $g(x;\boldsymbol{\xi})$, i.e.,

$$g(x;\boldsymbol{\xi}) \equiv g(x;\alpha,\beta) = \beta\alpha^{\beta}/x^{\beta+1}I(x \geq \alpha), \quad \alpha,\beta > 0. \tag{3}$$

For ease of presentation, we refer to model (2) with a general parametric density g as the general case, and refer to model (2) with the Pareto density g as the semiparametric Pareto mixture density (SPMD) model.

Identifiability is an important issue for most mixture models. Some theoretical results exist for mixture models with nonparametric components. Identifiability in multivariate mixture models was discussed by Hall and Zhou [13] and Allman et al. [1]. For univariate mixture models with symmetric nonparametric components, Bordes et al. [3] and Hunter et al. [14] gave conditions that guarantee identifiability. For our model (2), the following theorem ensures that the SPMD model is identifiable under mild conditions.

Theorem 2.1. *Suppose that density function $p(x)$ is continuous and symmetric about 0, and $g(x;\alpha,\beta)$ is the Pareto density function (3). If $\lim_{x\to\infty} p(x)x^{\beta+1} = 0$, then model (2) is identifiable.*

Theorem 2.1 implies that for Pareto density $g(x;\alpha,\beta)$, mixture model (2) is identifiable if the the symmetric component has a lighter tail than that of the Pareto component. Such symmetric densities include a wide class of symmetric densities $p(x)$, including normal, Laplace, Student's t densities with a degree of freedom greater than β, etc. This result broadens our knowledge on identifiability of mixture models.

2.2 Smoothed likelihood

The idea of smoothing likelihoods goes back to Silverman [27]. Before introducing the concept of smoothed likelihood, we first define some notation. Let Ω be a compact subset of $\mathbb{R}$, and define

$$\mathcal{F} = \{f : 0 < f \in L_1(\Omega), \log f \in L_1(\Omega)\},$$

where $L_1(\Omega)$ denotes the set of all integrable functions on Ω. Let $K(\cdot)$ be some kernel density function on $\mathbb{R}$, and $K_h(t) = h^{-1}K(t/h)$ be the rescaled version. Eggermont and LaRiccia [11] define a nonlinear smoothing operator $\mathcal{N}^*$ for $f \in \mathcal{F}$ as

$$\mathcal{N}^* f(x) = \exp\left\{\int_\Omega K_h(x-t)\log f(t)dt\right\}.$$

For independent and identically distributed observations $X_1, \ldots, X_n$ from a nonparametric density function $f(x)$, the smoothed log-likelihood is defined as

$$L_n(f) = \frac{1}{n}\sum_{i=1}^n \log\{\mathcal{N}^* f(X_i)\},\ f \in \mathcal{F}. \tag{4}$$

By using entropy inequality, one can show that the maximizer of (4) subject to $\int f(t)dt = 1$ is indeed a kernel density estimator.

In order to obtain a symmetric estimator of $p(\cdot)$ about μ in model (2), we define a symmetric smoothing kernel $K_{h,\mu}(x,t)$ as

$$K_{h,\mu}(x,t) = \frac{1}{2h}\left\{K\left(\frac{x-t}{h}\right) + K\left(\frac{2\mu-x-t}{h}\right)\right\}.$$

The corresponding nonlinear smoothing operator $\mathcal{N}_\mu$ for $p(\cdot)$ is then defined as

$$\mathcal{N}_\mu p(x) = \exp\left\{\int_\Omega K_{h,\mu}(x,t)\log p(t)dt\right\}.$$

After smoothing $f(\cdot)$ through $p(\cdot)$ in model (2), we can now construct the smoothed log-likelihood as

$$\ell(\boldsymbol{\theta}, p(\cdot)) = \sum_{i=1}^n \log\{\lambda g(X_i;\boldsymbol{\xi}) + (1-\lambda)\mathcal{N}_\mu p(X_i)\}, \tag{5}$$

where $\boldsymbol{\theta} = (\mu, \boldsymbol{\xi}, \lambda)^T$ is the vector of unknown parameters. Our goal in the next section is to find a maximizer of $\ell(\boldsymbol{\theta}, p(\cdot))$ subject to the constraint that $p(\cdot)$ is a symmetric density function.

3. ESTIMATION PROCEDURE

In this section a maximization procedure for the smoothed log-likelihood (5) is developed for the estimation of model (2). We first describe the procedure for general parametric distribution $g(x;\boldsymbol{\xi})$, and then discuss the situation of Pareto distribution g.

3.1 Estimation with general density $g(x;\boldsymbol{\xi})$

We first consider estimation of model (2) with general density $g(x;\boldsymbol{\xi})$. We use the following procedure: First, assuming that μ is known, an EM algorithm is established to maximize (5); second, a profile-likelihood for parameter μ is constructed and then maximized by a searching algorithm. We will briefly explain why μ has to be estimated separately in the end of the section.

Assuming that μ is known, we can rewrite the objective function (5) as $\ell(\boldsymbol{\xi}, \lambda, p(\cdot), \mathbf{X})$. In the EM framework, the observed data X_is are treated as incomplete, and the unobserved latent variables δ_is are introduced to indicate the group membership of X_is, i.e., let

$$\delta_i = \begin{cases} 1, & \text{if } X_i \text{ is from the parametric component;} \\ 0, & \text{otherwise.} \end{cases}$$

The complete data are $(\mathbf{X}, \boldsymbol{\delta}) = \{(X_i, \delta_i), i = 1, \ldots, n\}$, and the complete smoothed log-likelihood is

$$\begin{aligned}\ell^c(\boldsymbol{\xi}, \lambda, p(\cdot); \mathbf{X}) = \sum_{i=1}^n [&\delta_i \log\{\lambda g(X_i;\boldsymbol{\xi})\} \\ &+ (1-\delta_i)\log\{(1-\lambda)\mathcal{N}_\mu p(X_i)\}].\end{aligned} \tag{6}$$

In the l-th cycle of the EM algorithm, we have $(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot))$. In the E-step of $(l+1)$-th cycle, we calculate the expectation of $\ell^c(\boldsymbol{\xi}, \lambda, p(\cdot); \mathbf{X})$ given $(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot))$,

$$\begin{aligned}\mathcal{L} = \sum_{i=1}^n [&r_i^{(l+1)} \log\{\lambda g(X_i;\boldsymbol{\xi})\} \\ &+ (1-r_i^{(l+1)})\log\{(1-\lambda)\mathcal{N}_\mu p(X_i)\}],\end{aligned} \tag{7}$$

where

$$\begin{aligned}r_i^{(l+1)} &= \mathrm{E}(\delta_i | X_i; \boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot)) \\ &= \frac{\lambda^{(l)} g(X_i;\boldsymbol{\xi}^{(l)})}{\lambda^{(l)} g(X_i;\boldsymbol{\xi}^{(l)}) + (1-\lambda^{(l)})\mathcal{N}_\mu p^{(l)}(X_i)}.\end{aligned} \tag{8}$$

In the M-step of $(l+1)$-th cycle, we maximize $\mathcal{L}$ with respect to $\boldsymbol{\xi}, \lambda$, and $p(\cdot)$. It is easy to see that $\mathcal{L}$ can be partitioned into three parts, i.e., $\mathcal{L} = \mathcal{L}_1(\lambda) + \mathcal{L}_2(\boldsymbol{\xi}) + \mathcal{L}_3\{p(\cdot)\}$, where

$$(9)\qquad \mathcal{L}_1(\lambda) = \sum_{i=1}^n \{r_i^{(l+1)} \log \lambda + (1 - r_i^{(l+1)}) \log(1-\lambda)\},$$

$$(10)\qquad \mathcal{L}_2(\boldsymbol{\xi}) = \sum_{i=1}^n \{r_i^{(l+1)} \log g(X_i, \boldsymbol{\xi})\},$$

$$(11)\quad \mathcal{L}_3\{p(\cdot)\} = \sum_{i=1}^n \{(1 - r_i^{(l+1)}) \log \mathcal{N}_\mu p(X_i)\}.$$

Maximization of $\mathcal{L}$ is equivalent to maximizing $\mathcal{L}_1(\lambda)$, $\mathcal{L}_2(\boldsymbol{\xi})$, and $\mathcal{L}_3\{p(\cdot)\}$, separately. The solution to (9) is

$$(12)\qquad \lambda^{(l+1)} = \sum_{i=1}^n r_i^{(l+1)}/n.$$

The solution to (10) is a weighted MLE:

$$(13)\qquad \boldsymbol{\xi}^{(l+1)} = \text{argmax} \sum_{i=1}^n \{r_i^{(l+1)} \log g(X_i, \boldsymbol{\xi})\}.$$

If the solution (13) does not admit an explicit form, we may consider using a gradient method [17] or conditional maximization [24]. The solution to the maximization of $\mathcal{L}_3(p(\cdot))$ is characterized in the following theorem.

Theorem 3.1. *Subject to the condition that $p(\cdot)$ is a symmetric density function that belongs to $\mathcal{F}$, $\mathcal{L}_3$ has a unique maximizer $p^{(l+1)}(\cdot)$, up to changes on a set of zero Lebesgue measure:*

$$(14)\qquad p^{(l+1)}(\cdot) = \frac{\sum_{i=1}^n (1 - r_i^{(l+1)}) K_{h,\mu}(X_i, \cdot)}{\sum_{i=1}^n (1 - r_i^{(l+1)})}.$$

Repeating (8), (12), (13) and (14) iteratively until convergence, we obtain $\hat{\boldsymbol{\xi}}$, $\hat{\lambda}$ and $\hat{p}(\cdot)$. The employment of EM algorithm separates the estimation of λ, $\boldsymbol{\xi}$ and $p(\cdot)$ and makes the computation of each part simpler. Moreover, estimation via the proposed EM algorithm automatically guarantees a nonnegative and symmetric estimation of $p(\cdot)$ as well as the constraint on λ.

It is of interest to investigate whether the ascent property holds for the proposed EM algorithm. To derive the ascent property, we first show that $\mathcal{L}$ minorizes ℓ up to a shifted constant.

Theorem 3.2. *For any $\boldsymbol{\xi}$, $\lambda \in (0,1)$ and $p \in \mathcal{F}$, the functions ℓ and $\mathcal{L}$ defined by (5) and (7) satisfies*

$$\mathcal{L}(\boldsymbol{\xi}, \lambda, p(\cdot)) + C^{(l)} \le \ell(\boldsymbol{\xi}, \lambda, p(\cdot)),$$

where $C^{(l)} = \ell(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot)) - \mathcal{L}(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot))$, and the equality holds when $(\boldsymbol{\xi}, \lambda, p(\cdot)) = (\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot))$.

By Theorem 3.2, we have

$$\mathcal{L}(\boldsymbol{\xi}, \lambda, p(\cdot)) - \mathcal{L}(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot)) \le \ell(\boldsymbol{\xi}, \lambda, p(\cdot)) - \ell(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot)).$$

Therefore, $\mathcal{L}(\boldsymbol{\xi}^{(l+1)}, \lambda^{(l+1)}, p^{(l+1)}(\cdot)) \ge \mathcal{L}(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot))$ implies that

$$\ell(\boldsymbol{\xi}^{(l+1)}, \lambda^{(l+1)}, p^{(l+1)}(\cdot)) \ge \ell(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot)),$$

which gives the ascent property.

Now we explain why μ has to be estimated separately. If we treat μ as unknown parameter in $\mathcal{L}_3$, there appear to be several methods that might be used to estimate μ and $p(\cdot)$ simultaneously. If density $p(\cdot)$ have finite expectation, the weighted sample mean $\hat{\mu}_M = \sum_{i=1}^n (1-r_i) X_i / \sum_{i=1}^n (1-r_i)$, and sample median $\hat{\mu}_{Med}$ of the weighted data $X_1, \cdots, X_n$ with weights $w_i = 1 - r_i$ are two natural candidate estimators for μ, since μ is the symmetry point. If $p(\cdot)$ is unimodal, the maximum point $\hat{\mu}_U$ of the ordinary kernel estimator of $p(\cdot)$ can serve as an estimator for μ. However, none of these estimators can be proven to maximize $\mathcal{L}_3(\mu, p(\cdot))$, and thus can not guarantee the ascent property if we estimate μ simultaneously in the EM algorithm. Hence the estimation of μ is left to a searching scheme. Letting $\hat{\lambda}_\mu, \hat{\boldsymbol{\xi}}_\mu, \hat{p}_\mu(\cdot)$ denote the estimators of λ, $\boldsymbol{\xi}$ and $p(\cdot)$ given μ, we obtain the profile-likelihood

$$(15)\qquad \hat{\ell}_p(\mu) = \ell\{\mu, \hat{\lambda}_\mu, \hat{\boldsymbol{\xi}}_\mu, \hat{p}_\mu(\cdot)\}.$$

This becomes a one-dimensional maximization problem, and its maximizer $\hat{\mu}$ can be obtained by searching the maximum of $\hat{\ell}_p$ with respect to μ. It can be solved using some advanced numerical methods implemented in computing softwares, e.g., the function "fminsearch" in Maltab. The final estimates can be obtained by plugging in $\hat{\mu}$, i.e., $\hat{\lambda} = \hat{\lambda}_{\hat{\mu}}$, $\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_{\hat{\mu}}$, and $\hat{p}(\cdot) = \hat{p}_{\hat{\mu}}(\cdot)$.

Remark. It is very difficult to simultaneously maximize $\mathcal{L}_3$ with respect to both nonparametric density $p(\cdot)$ and parameter μ. However, in special cases where $p(\cdot)$ is parametric, μ can be estimated in the EM algorithm without breaking the ascent property. See the estimation procedure of (19) in the appendix.

3.2 Estimation for SPMD

When $g(x; \boldsymbol{\xi})$ is a Pareto distribution and $\boldsymbol{\xi} = (\alpha, \beta)$, the solution to (10) is

$$(16)\qquad \alpha^{(l+1)} = \min_i \{X_i : r_i^{(l+1)} > 0, i = 1, \cdots, n\},$$

$$(17)\qquad \beta^{(l+1)} = \frac{\sum_{i=1}^n r_i^{(l+1)}}{\sum_{i=1}^n r_i^{(l+1)} (\log X_i - \log \alpha^{(l+1)})}.$$

Let i_0 be the index corresponding to the minimum of (16), i.e., $\alpha^{(l+1)} = X_{i_0}$. Then for any $X_i < X_{i_0}$, the estimated density $g^{(l+1)}(X_i; \boldsymbol{\xi}) = 0$, and therefore the corresponding expectation of latent variable in the following cycle $r_i^{(l+2)}$ is also zero by (8). On the other hand, for any $X_i > X_{i_0}$, it is clear that the corresponding $r_i^{(l+2)} > 0$. Hence in the next iteration $\alpha^{(l+2)}$ is still X_{i_0}. It means that if α is updated in the EM algorithm, it will remain to be the initial value.

Hence the estimation of α should not be included in the EM algorithm.

Since the location parameter α for the Pareto component is the only discontinuity point of the mixture density, a number of methods could be used for its estimation, e.g., see Chu et al. [6] and Gayraud [12]. Gayraud [12] proposed minimax estimators based on differences of histograms, and showed that their estimators could achieve the optimal convergence rate n^{-1}. For convenience, we use the estimator with uniform kernels in Chu et al. [6], which is similar to the first stage estimator of Gayraud [12]. Once α is estimated, the rest estimation procedure follows as in Section 3.1. The complete estimation procedure is summarized as follows:

1. Estimate α using the estimator in Chu et al. [6], and denote the estimator as $\hat{\alpha}$.
2. Estimate μ by searching the maximum of profile-likelihood (15) with respect to μ, where the EM algorithm with fixed μ is as follows:
   (a) Let μ be fixed and with given initial values $(\beta^{(0)}, \lambda^{(0)}, p^{(0)}(\cdot))$, iterate E-step and M-step for $l = 0, 1, \cdots$, until convergence.
   (b) E-step. For each i, compute
   $$r_i^{(l+1)} = \frac{\lambda^{(l)} g(X_i; \hat{\alpha}, \beta^{(l)})}{\lambda^{(l)} g(X_i; \hat{\alpha}, \beta^{(l)}) + (1-\lambda^{(l)}) \mathcal{N}_\mu p^{(l)}(X_i)}.$$
   (c) M-step. Compute
   $$\lambda^{(l+1)} = \frac{\sum_{i=1}^n r_i^{(l+1)}}{n},$$
   $$\beta^{(l+1)} = \frac{\sum_{i=1}^n r_i^{(l+1)}}{\sum_{i=1}^n r_i^{(l+1)} (\log X_i - \log \hat{\alpha})},$$
   $$p^{(l+1)}(\cdot) = \frac{\sum_{i=1}^n (1 - r_i^{(l+1)}) K_{h,\mu}(X_i, \cdot)}{\sum_{i=1}^n (1 - r_i^{(l+1)})}.$$
3. Obtain the final estimates by plugging in $\hat{\mu}$, i.e., $\hat{\lambda} = \hat{\lambda}_{\hat{\mu}}$, $\hat{\beta} = \hat{\beta}_{\hat{\mu}}$, and $\hat{p}(\cdot) = \hat{p}_{\hat{\mu}}(\cdot)$.

4. SIMULATION AND APPLICATION

4.1 Bandwidth selection

The choice of the bandwidth h is a challenging problem. For kernel density estimation, several methods were introduced in Li and Racine [19], including Rule-of-Thumb, Plug-in, Least Square CV and Likelihood CV. For computation consideration, we use a refinement of the Rule-of-Thumb method. In Silverman [28], it is advocated to use the following Rule-of-Thumb bandwidth for nonparametric density,

$$h = C^* n^{-1/5} \min(\sigma_y, IQR_y/1.34),$$

where $C^* = 0.9$, σ_y and IQR_y are the standard deviations and the interquartile range of the sample, respectively. Another common variation is given by Scott [26], using factor $C^* = 1.06$ instead of 0.9. As our model is a mixture model, not all the data come from the nonparametric component, therefore a direct calculation of σ_y or IQR_y in (18) will be biased, and the deviation depends on the location and proportion of the two components. For semiparametric mixture model (2), a reasonable modification is

$$h = C^* (n(1-\lambda))^{-1/5} \min(\sigma_p, IQR_p/1.34), \tag{18}$$

where σ_p and IQR_p are the standard deviations and the interquartile range of the nonparametric component. In this paper, we use the following steps to deal with this problem.

1. Approximate the nonparametric component by a normal distribution, and maximize the log-likelihood
   $$\ell(\boldsymbol{\theta}_1) = \sum_{i=1}^n \log \left\{ \lambda g(X_i; \alpha, \beta) + (1-\lambda) \phi(X_i | \mu, \sigma^2) \right\}, \tag{19}$$
   with respect to $\boldsymbol{\theta}_1 = (\lambda, \alpha, \beta, \mu, \sigma^2)$, where $\phi(x|\mu, \sigma^2)$ is the density function of $N(\mu, \sigma^2)$. The estimation procedure of (19) is given in the appendix.
2. With λ, σ_p and IQR_p estimated by step 1, select a bandwidth h_0 by (18).
3. Given bandwidth h_0, estimate λ, σ_p and IQR_p in model (2) using the procedure in section 3.2, and then select bandwidth by (18).

4.2 Numerical simulation

We conduct simulation studies to demonstrate the performance of the proposed model. The estimation of parameters is evaluated via the root of mean square error (RMSE)

$$RMSE(\hat{\theta}) = \sqrt{\frac{1}{S} \sum_{s=1}^S (\hat{\theta}^{(s)} - \theta)^2},$$

and the performance of nonparametric density estimation is evaluated via the root of mean intergraded squared error (RMISE)

$$RMISE(\hat{p}(\cdot)) = \sqrt{\frac{1}{S} \sum_{s=1}^S \int \{\hat{p}^{(s)}(u) - p(u)\}^2 du}, \tag{20}$$

where $\hat{\theta}^{(s)}$ and $\hat{p}^{(s)}(\cdot)$ are the estimates in the s-th replication.

Example 1. The simulated data is generated from the mixture of a normal distribution $N(0.4, 0.5^2)$ and a Pareto distribution with parameters $\alpha = 0.1$ and $\beta = 3$. The mixing proportion λ takes values of 0.25, 0.5, and 0.75. We ran $S = 500$ replications of n observations ($n = 400, 800$). We estimate the parameters by two models, the SPMD model where the normal distribution is regarded as the nonparametric component, and the parametric model defined in (19). For SPMD model, the bandwidth h is selected for each simulated sample by using the modified Rule-of-Thumb

Table 1. RMSE for parameters and RMISE for $p(\cdot)$ when $p(\cdot)$ is a normal distribution

n	true λ	α	Model (19)				SPMD			
			λ	β	μ	$p(\cdot)$	λ	β	μ	$p(\cdot)$
	0.25	0.0015	0.0324	0.6590	0.0299	0.0389	0.0373	0.6447	0.0334	0.0657
400	0.5	0.0005	0.0337	0.3786	0.0369	0.0462	0.0358	0.3868	0.0439	0.0821
	0.75	0.0003	0.0300	0.2604	0.0510	0.0670	0.0300	0.2628	0.0669	0.1129
	0.25	0.0007	0.0221	0.4543	0.0206	0.0260	0.0248	0.4497	0.0220	0.0481
800	0.5	0.0004	0.0242	0.2528	0.0255	0.0322	0.0260	0.2776	0.0279	0.0598
	0.75	0.0002	0.0201	0.1846	0.0384	0.0469	0.0210	0.1896	0.0438	0.0840

Table 2. RMSE for parameters and RMISE for $p(\cdot)$ when $p(\cdot)$ is a t-distribution

n	true λ	α	Model (19)				SPMD			
			λ	β	μ	$p(\cdot)$	λ	β	μ	$p(\cdot)$
	0.25	0.0018	0.1449	1.5099	0.0529	0.2539	0.0573	0.9227	0.0243	0.0951
400	0.5	0.0005	0.1094	0.9487	0.0528	0.2655	0.0473	0.4979	0.0301	0.1154
	0.75	0.0003	0.0589	0.4939	0.0859	0.2708	0.0368	0.3019	0.0479	0.1588
	0.25	0.0011	0.1404	1.5185	0.0332	0.2592	0.0417	0.7121	0.0170	0.0720
800	0.5	0.0003	0.1121	0.9625	0.0492	0.2691	0.0344	0.3626	0.0211	0.0850
	0.75	0.0002	0.0535	0.4630	0.0504	0.2630	0.0264	0.2324	0.0323	0.1195

Table 3. Standard error estimation when $p(\cdot)$ is a t-distribution

	$\lambda = 0.25$		$\lambda = 0.50$		$\lambda = 0.75$	
	SD	SE(STD)	SD	SE(STD)	SD	SE(STD)
α	0.0009	0.0224(0.0264)	0.0004	0.0006(0.0004)	0.0002	0.0003(0.0002)
λ	0.0439	0.0395(0.0058)	0.0445	0.0390(0.0042)	0.0331	0.0317(0.0026)
β	0.6869	0.7363(0.2747)	0.3661	0.4057(0.0768)	0.2531	0.2769(0.0363)
μ	0.0245	0.0240(0.0030)	0.0286	0.0314(0.0062)	0.0467	0.0520(0.0142)

method described in Section 4.1. The simulation results are shown in Table 1.

The parameter estimation of α has the smallest RASE, which shows that the estimator based on difference of histogram works quite well. We see that the estimation procedure of model (2) works almost as well as the parametric model (19). Both the methods give good estimation of λ and μ. The RMSE of β is relatively large as compared to λ and μ. However, it shrinks when sample size increases from $n = 400$ to $n = 800$ for both models. For the RMISE of $p(\cdot)$, the result of semiparametric model are about 1.8 times of the results of parametric model. Also note that the RMISE of $p(\cdot)$ increases when λ increases from 0.25 to 0.75. This is because $(1-\lambda)$ is the proportion of nonparametric component and then larger λ means fewer information of $p(\cdot)$.

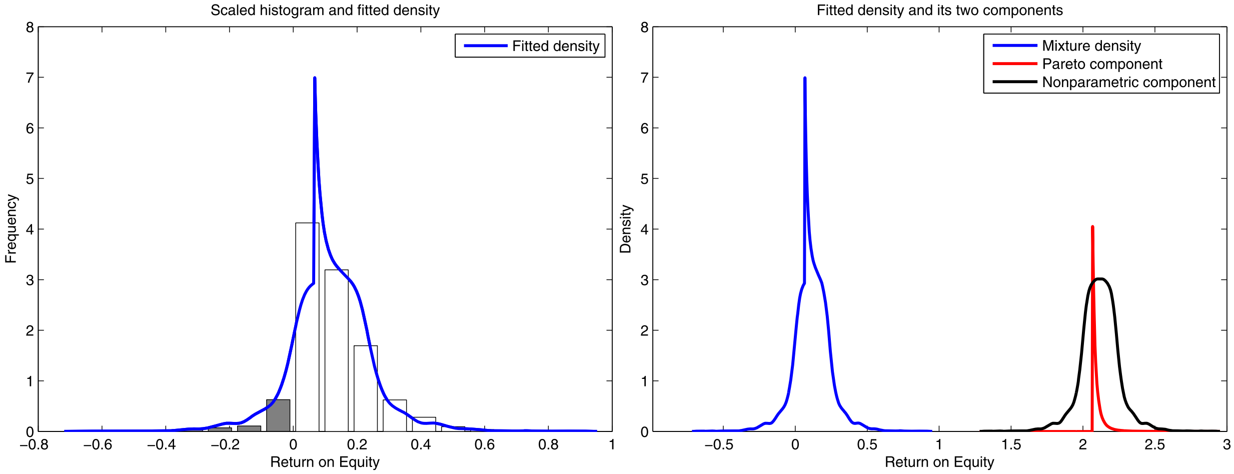
Example 2. In this example we generate data from a mixture of a Pareto distribution with parameters $\alpha = 0.1$ and $\beta = 3$, and a scaled t-distribution. The scaled t-distribution is assumed to be $T/\sqrt{12}+\mu$, where T follows a t-distribution with $df = 3$ and the symmetry point $\mu = 0.4$. We estimate the parameters by the two models as in example 1, and the results are given in Table 2.

Since the normal assumption is not satisfied, the parameters estimated by parametric model (19) are much worse than that of the SPMD model as expected. The RMSE of λ and β are quite large, and increasing the sample size does not yield significant improvements. On the other hand, estimates by SPMD model have smaller RMSE for all parameters, and the RMSE reduces as sample size increases. As to the estimation of nonparametric part, the estimate of SPMD model has much smaller RMISE than that of the parametric model. Similar to Example 1, the RMISE of $p(\cdot)$ increase as λ increase for SPMD model. However, for Model (19), the RMISE of $p(\cdot)$ show little difference when λ varies.

We use a nonparametric bootstrap method for standard error estimation, and conduct simulation to test its performance. In 500 simulations, we calculate the standard deviation of 500 estimates as a proxy of the true standard error, denoted by SD. In each simulation, the standard error is estimated by bootstrap. The mean and standard deviation of the 500 estimated standard errors are recorded, denoted by SE and STD. From the results in Table 3, we see that the bootstrap procedure provides reasonable estimates for the true standard deviation, except for α when the sample size of the Pareto component is small in the case of true $\lambda = 0.25$.

4.3 Analysis of Chinese ROE data

In this section we analyze a dataset on ROE of Chinese listed companies. The dataset is obtained from the Wind Fi-



(a) (b)

Figure 2. (a) Scaled histograms and the fitted density; (b) Two components (shifted) of the fitted density.

nancial Terminal Database (http://www.wind.com.cn). We collect a total of 2,110 observations in 2010 for all Chinese listed companies. Return On Equity (ROE) refers to the amount of net income returned as a percentage of shareholders equity and is an important index which measures a corporation's profitability. It is also a useful indicator for fundamental analysts to price the value of stocks. Understanding the companies' ROE plays a critical role in investment analysis. Wang and Tsai [29] investigated the extreme situations of Chinese stock market by a tail index regression with the ROE of Chinese listed firms.

As shown in Figure 1, China's ROE data is an example of asymmetry of ROE data with a jump on the right of zero. This empirical phenomenon is not unique to the Chinese stock market, but exists in many other markets. This phenomenon has been noted and studied in literature such as in Burgstahler and Dichev [5], Ding et al. [8], and Jiang and Wang [15]. However, to the best of our knowledge no quantitative model has been proposed for the ROE data which has a jump near the right of the zero point and heavy-tailed distributions, and consist of probabilistic categories of real and manipulated earnings. Since empirical accounting researches show that the reported ROE values should constitute two different components [15], a tail index model is insufficient to precisely model the data patterns and provide reasonable estimates. We analyze the data using the proposed SPMD model and estimation procedure. The real and manipulated earnings are characterized by a symmetric nonparametric distribution and a Pareto distribution, respectively. Pareto distribution is well known for heavy-tail modeling, and has been demonstrated as a useful model for the size distribution of various heavy-tail phenomena, including incomes, earnings, and stock returns (See Reed and Jorgensen [25]). The underlining assumption to employ Pareto distribution for earning manipulation is that most companies manipulate loss as small profit, while a few companies manipulate loss as large profit.

Table 4 depicts the estimates and their confidence intervals.The estimated Pareto component coefficients are $\alpha = 0.065$ and $\beta = 2.911$, with the estimated proportion $\lambda = 0.102$. The estimated nonparametric component is symmetric about $\mu = 0.1172$ with proportion 0.898. The result suggests that about 89.8% of Chinese listed firms show the real earnings, with average return on equity 11.72%; about 10.2% of Chinese listed companies manipulate their earnings to show a slightly positive profit, whose earnings may actually be negative. The fitted density superimposed on a scaled version of the histogram is shown in Figure 2(a), and the two components are separately shown in Figure 2(b). We further obtain the 95% confidence intervals by a nonparametric bootstrap method. From Table 4, λ is significantly larger than zero and the proportion of manipulated earnings is between 6.27% and 13.84% with probability 95%. The above analysis showed significant earning manipulation in the Chinese listed companies from the quantitative perspective, which is consistent with the previous qualitative analyses obtained by Burgstahler and Dichev [5], Ding et al. [8], and Jiang and Wang [15].

Table 4. Estimators and the 95% confidence intervals for the estimators, based on 500 *bootstrap replications, for the ROE data example*

	estimator	95 % CI
α	0.065	(0.0031, 0.1330)
λ	0.102	(0.0627, 0.1384)
β	2.911	(2.0463, 4.7851)
μ	0.1172	(0.1107, 0.1237)

One practical application of the analysis is to provide a model-based revision of the average ROE. Note that in the

2010 dataset, the naive average of ROE is 12.13%. Since manipulation exists, the value 12.13% is over estimated. Given the estimates of our model, a reasonable revision of the average ROE is the weighted average of the means of the two components, where the mean of the manipulated component can be estimated using the average of all negative ROE of the original data, which is -8.48%. Therefore, the revised average ROE is $0.898 * 11.72\% + 0.102 * (-8.48\%) = 9.66\%$.

5. DISCUSSION

In this paper, we introduced a novel semiparametric Pareto mixture density model, and proposed an efficient estimation procedure via maximum smoothed likelihood estimation and EM algorithm. We showed that the proposed model are identifiable under mild conditions. We constructed the objective function via the smoothed likelihood framework, and developed an estimation procedure via profiled likelihood and EM algorithm. Furthermore, we proved the ascent property of proposed EM algorithm. However, the proposed EM algorithm is not limited to the Pareto component. Our algorithm can be implemented for other parametric component, e.g., normal or exponential distribution, provided that the model is identifiable. Therefore, research on the identifiability of model (2) with other specific parametric component is undoubtedly of great importance. Further research on the asymptotic distribution is also of interest.

In this paper, we focused on the Pareto distribution as the parametric component, which is motivated by the jump in the density of the ROC data of Chinese firms. We analyzed the Chinese ROE dataset, which may consist of real and manipulated earning as suggested by the previous qualitative studies. Our analysis confirmed that there are significant manipulation in the ROE dataset, and provided a novel model-based analysis including a quantitative measure for the proportion of the manipulated earning, and a quantitative revision of the average ROE.

APPENDIX A. APPENDIX SECTION

A.1 Proofs

Proof of Theorem 2.1. Since the symmetric component density $p(x)$ is continuous, $x = \alpha$ is the unique discontinuity point of the mixture density $f(x)$ of (3), hence the location parameter α for Pareto density is identifiable. Moreover, the jump size $f(\alpha) - f(\alpha-)$ of $f(x)$ at $x = \alpha$ is also identifiable, while $f(\alpha-)$ denotes the left limit of $f(x)$ when x approaches α from the left side, $f(\alpha) - f(\alpha-) = \lambda\beta/\alpha$, hence the product $\lambda\beta$ is identifiable. For simplicity and without loss of generality, we assume that $\alpha = 1$.

Now for any constant $\gamma > 0$, let us consider

$$\lim_{x\to\infty} f(x)x^{\gamma+1} = \lambda\beta x^{\gamma-\beta} I\{x \geq 1\} + (1-\lambda)x^{\gamma+1}p(x-\mu).$$

It follows that

$$\lim_{x\to\infty} f(x)x^{\gamma+1} = \begin{cases} 0 & \gamma < \beta \\ \lambda\beta & \gamma = \beta \\ \infty & \gamma > \beta. \end{cases}$$

Hence $\gamma = \beta$ is the unique value such that the limit is a positive constant. Thus parameter β of the Pareto density is identifiable. Since the product $\lambda\beta$ is identifiable, the proportion parameter λ is identifiable. By the symmetry of $p(x)$, the center of the nonparametric component μ is also identifiable. This proves Theorem 2.1.

Proof of Theorem 3.2. By applying Jensen's inequality, we have

$$\begin{aligned}
&\ell(\boldsymbol{\xi}, \lambda, p(\cdot)) - \ell(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot)) \\
=&\sum_{i=1}^{n} \log \frac{\lambda g(X_i;\boldsymbol{\xi}) + (1-\lambda)\mathcal{N}_\mu p(X_i)}{\lambda^{(l)} g(X_i;\boldsymbol{\xi}^{(l)}) + (1-\lambda^{(l)})\mathcal{N}_\mu p^{(l)}(X_i)} \\
=&\sum_{i=1}^{n} \log \left\{ r_i^{(l+1)} \frac{\lambda g(X_i;\boldsymbol{\xi})}{\lambda^{(l)} g(X_i;\boldsymbol{\xi}^{(l)})} \right. \\
&\left. + q_i^{(l+1)} \frac{(1-\lambda)\mathcal{N}_\mu p(X_i)}{(1-\lambda^{(l)})\mathcal{N}_\mu p^{(l)}(X_i)} \right\} \\
\geq&\sum_{i=1}^{n} \left\{ r_i^{(l+1)} \log \frac{\lambda g(X_i;\boldsymbol{\xi})}{\lambda^{(l)} g(X_i;\boldsymbol{\xi}^{(l)})} \right. \\
&\left. + q_i^{(l+1)} \log \frac{(1-\lambda)\mathcal{N}_\mu p(X_i)}{(1-\lambda^{(l)})\mathcal{N}_\mu p^{(l)}(X_i)} \right\} \\
=&\mathcal{L}(\boldsymbol{\xi}, \lambda, p(\cdot)) - \mathcal{L}(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot)),
\end{aligned}$$

where $q_i^{(l+1)} = 1 - r_i^{(l+1)}$. Taking $C^{(l)}$ to be $\ell(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot)) - \mathcal{L}(\boldsymbol{\xi}^{(l)}, \lambda^{(l)}, p^{(l)}(\cdot))$ proves Theorem 3.2.

A.2 Estimation procedure for model (19)

We present an estimation procedure to maximize the log-likelihood (19) with respect to $\alpha, \beta, \lambda, \mu$ and σ^2. Similar to Section 3.2, α is estimated using the estimator in Chu et al. [6], denoted by $\hat{\alpha}$.

Then we use an EM algorithm as follows:

1. Give initial values $(\beta^{(0)}, \lambda^{(0)}, \mu^{(0)}, \sigma^{2(0)})$, iterate E-step and M-step until convergence.
2. E-Step. For each i,

$$r_i^{(l+1)} = \frac{\lambda^{(l)} g(X_i;\hat{\alpha},\beta^{(l)})}{\lambda^{(l)} g(X_i;\hat{\alpha},\beta^{(l)}) + (1-\lambda^{(l)})\phi(X_i;\mu^{(l)},\sigma^{2(l)})}.$$

3. M-Step.

$$\begin{aligned}
\lambda^{(l+1)} &= \frac{\sum_{i=1}^n r_i^{(l+1)}}{n}, \\
\beta^{(l+1)} &= \frac{\sum_{i=1}^n r_i^{(l+1)}}{\sum_{i=1}^n r_i^{(l+1)}(\log X_i - \log\hat{\alpha})}, \\
\mu^{(l+1)} &= \frac{\sum_{i=1}^n (1-r_i^{(l+1)})X_i}{\sum_{i=1}^n (1-r_i^{(l+1)})}
\end{aligned}$$

$$\sigma^{2(l+1)} = \frac{\sum_{i=1}^n (1 - r_i^{(l+1)})(X_i - \mu^{(l+1)})^2}{\sum_{i=1}^n (1 - r_i^{(l+1)})}.$$

ACKNOWLEDGEMENTS

Mian Huang's research is supported by National Natural Science Foundation of China (NSFC, 11301324) and Shanghai Chenguang Program. Shaoli Wang's research is supported by National Natural Science Foundation of China (NSFC, 11371235) and Program for Innovative Research Team of Shanghai University of Finance and Economics. Hansheng Wang's research is supported in part by National Natural Science Foundation of China (NSFC, 11131002, 11271032).

Received 12 October 2016

REFERENCES

[1] Allman, E., Matias, C., and Rhodes, J. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37(6A):3099–3132. MR2549554

[2] Bordes, L., Delmas, C., and Vandekerkhove, P. (2006a). Semiparametric estimation of a two-component mixture model where one component is known. *Scandinavian journal of statistics*, 33(4):733–752. MR2300913

[3] Bordes, L., Mottelet, S., and Vandekerkhove, P. (2006b). Semiparametric estimation of a two-component mixture model. *Annals of Statistics*, 34(3):1204–1232. MR2278356

[4] Bordes, L. and Vandekerkhove, P. (2010). Semiparametric two-component mixture model with a known component: An asymptotically normal estimator. *Mathematical Methods of Statistics*, 19(1):22–41. MR2682853

[5] Burgstahler, D. and Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of accounting and economics*, 24(1):99–126.

[6] Chu, C., Cheng, P., et al. (1996). Estimation of jump points and jump values of a density function. *Statistica Sinica*, 6(1):79–95. MR1379050

[7] Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38. MR0501537

[8] Ding, Y., Zhang, H., and Zhang, J. (2007). Private vs state ownership and earnings management: evidence from chinese listed companies. *Corporate Governance: An International Review*, 15(2):223–238.

[9] Eggermont, P. (1999). Nonlinear smoothing and the em algorithm for positive integral equations of the first kind. *Applied mathematics & optimization*, 39(1):75–91. MR1654562

[10] Eggermont, P. and LaRiccia, V. (1995a). Maximum smoothed likelihood density estimation. *Journal of Nonparametric Statistics*, 4(3):211–222. MR1366769

[11] Eggermont, P. and LaRiccia, V. (1995b). Maximum smoothed likelihood density estimation for inverse problems. *The Annals of Statistics*, 23:199–220. MR1331664

[12] Gayraud, G. (2002). Minimax estimation of a discontinuity for the density. *Journal of Nonparametric Statistics*, 14(1–2):59–66. MR1905585

[13] Hall, P. and Zhou, X. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Annals of Statistics*, pages 201–224. MR1962504

[14] Hunter, D., Wang, S., and Hettmansperger, T. (2007). Inference for mixtures of symmetric distributions. *Annals of Statistics*, 35(1):224–251. MR2332275

[15] Jiang, G. and Wang, H. (2008). Should earnings thresholds be used as delisting criteria in stock market? *Journal of Accounting and Public Policy*, 27(5):409–419.

[16] Kraft, C., Lepage, Y., and Van Eeden, C. (1985). Estimation of a symmetric density function. *Communications in statistics. Theory and methods*, 14(2):273–288. MR0788587

[17] Lange, K. (1995). A gradient algorithm locally equivalent to the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 425–437. MR1323348

[18] Levine, M., Hunter, D., and Chauveau, D. (2011). Maximum smoothed likelihood for multivariate mixtures. *Biometrika*, 98(2):403. MR2806437

[19] Li, Q. and Racine, J. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press, Princeton, NJ. MR2283034

[20] Ma, Y. and Yao, W. (2015). Flexible estimation of a semiparametric two-component mixture model with one parametric component. *Electronic Journal of Statistics*, 9(1):444–474. MR3326131

[21] Maiboroda, R. and Sugakova, O. (2010a). Generalized estimating equations for symmetric distributions observed with admixture. *Communications in Statistics—Theory and Methods*, 40(1):96–116. MR2747201

[22] Maiboroda, R. and Sugakova, O. (2010b). Nonparametric density estimation for symmetric distributions by contaminated data. *Metrika*, pages 1–18. MR2878111

[23] McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*, volume 382. John Wiley & Sons. MR2392878

[24] Meng, X. and Rubin, D. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, 80(2):267. MR1243503

[25] Reed, W. J. and Jorgensen, M. (2004). The double pareto-lognormal distribution—a new parametric model for size distributions. *Communications in Statistics-Theory and Methods*, 33(8):1733–1753. MR2065171

[26] Scott, D. (1992). *Multivariate density estimation*, volume 139. Wiley Online Library. MR1191168

[27] Silverman, B. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *The Annals of Statistics*, pages 795–810. MR0663433

[28] Silverman, B. (1986). *Density estimation for statistics and data analysis*, volume 26. Chapman & Hall/CRC. MR0848134

[29] Wang, H. and Tsai, C. (2009). Tail index regression. *Journal of the American Statistical Association*, 104(487):1233–1240. MR2750246

[30] Wu, C. (1983). On the convergence properties of the em algorithm. *The Annals of Statistics*, pages 95–103. MR0684867

Mian Huang
School of Statistics and Management
Shanghai University of Finance and Economics
People's Republic of China
E-mail address: huang.mian@shufe.edu.cn

Shaoli Wang
School of Statistics and Management
Shanghai University of Finance and Economics
People's Republic of China
E-mail address: swang@shufe.edu.cn

Hansheng Wang
Department of Business Statistics and Econometrics
Guanghua School of Management
Peking University
People's Republic of China
E-mail address: hansheng@gsm.pku.edu.cn

Tian Jin
Shanghai Stock Exchange
People's Republic of China
E-mail address: jintian82@gmail.com