

# Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation

Daniel Hohmann and Hajo Holzmann\*

*Fakultät für Mathematik und Informatik, Philipps-Universität Marburg,  
Hans-Meerwein-Str., 35032 Marburg, Germany*

*e-mail:* [hohmann@mathematik.uni-marburg.de](mailto:hohmann@mathematik.uni-marburg.de); [holzmann@mathematik.uni-marburg.de](mailto:holzmann@mathematik.uni-marburg.de)

**Abstract:** We show how the multivariate two-component mixtures with independent coordinates in each component by Hall and Zhou (2003) can be studied within the framework of conditional mixtures as recently introduced by Henry, Kitamura and Salanié (2010). Here, the conditional distribution of the random variable  $Y$  given the vector of regressors  $Z$  can be expressed as a two-component mixture, where only the mixture weights depend on the covariates. Under appropriate tail conditions on the characteristic functions and the distribution functions of the mixture components, which allow for flexible location-scale type mixtures, we show identification and provide asymptotically normal estimators. The main application for our results are bivariate two-component mixtures with independent coordinates, the case not previously covered by Hall and Zhou (2003). In a simulation study we investigate the finite-sample performance of the proposed methods. The main new technical ingredient is the estimation of limits of quotients of two characteristic functions in the tails from independent samples, which might be of some independent interest.

**Keywords and phrases:** Characteristic function, conditional mixture, finite mixture, nonparametric estimation.

**AMS 2000 subject classifications:** 62G05, 62G20.

Received August 2012.

## 1. Introduction

Finite mixtures are frequently used to model populations with unobserved heterogeneity. While the component distributions are most often chosen from some parametric family, e.g. the normal or  $t$ -distributions, cf. McLachlan and Peel (2000), in recent years there has been quite some interest in finite mixtures with nonparametric components, see below for a review of some of the literature.

A prominent example is the multivariate two-component mixture with independent coordinates in each component by Hall and Zhou (2003, HZ in what follows), which they introduced for modeling results of repeated tests on a single person with unknown disease status.

---

\*Corresponding author.

In this paper we show how the model by HZ can be cast into the framework of two-component conditional mixtures by Henry, Kitamura and Salanié (2010, HKS in the following). In particular, our results imply that for the HZ-model in the (in general only partially identified) two-dimensional case, an appropriate representation of the factors in the independent components can still be identified and estimated under some additional tail assumptions.

Suppose that the conditional distribution of the random variable  $Y$  given the vector of regressors  $Z$  can be expressed as the two-component mixture

$$F(y|z) = (1 - \lambda(z))F_0(y) + \lambda(z)F_1(y), \quad (1.1)$$

where only the mixture weights depend on the covariates. Apart from actual dependence of  $\lambda(z)$  on the covariates, identification in model (1.1) requires additional assumptions on the component distributions  $F_0$  and  $F_1$ . HKS investigate identifiability and estimation under tail conditions of the distribution functions themselves, which are tailored to location-type mixtures, but do not work for scale mixtures. We focus on identification and estimation results of (1.1) under appropriate tail conditions on the characteristic functions of  $F_0$  and  $F_1$ , which shall allow for more flexible location-scale-type mixtures. Indeed, our main technical contribution is the derivation of the large-sample theory for characteristic function-based estimators.

Let us review some of the literature on mixtures with nonparametric components. In the simple case of finite mixtures of univariate distributions, most theoretical work assumes symmetry of each component distribution. For example, Bordes, Mottelet and Vandekerkhove (2006) and Hunter, Wang and Hettmansperger (2007) present results on identifiability and asymptotically normal estimation in a two-component location mixture of a single symmetric distribution, and Bordes and Vandekerkhove (2010) and Hohmann and Holzmann (2012) have similar results in a two-component mixture model with two symmetric components one of which is completely specified while the other is unknown with unknown location parameter. For mixtures of regressions, there is a series of work which exploits the additional information provided by covariates. Kasahara and Shimotsu (2009) extend the HZ-approach to the context of switching regressions. Kitamura (2004, unpublished) considers identifiability issues for univariate mixtures of regressions on the mean functions and obtains full nonparametric identification of the components under some tail assumptions of either their characteristic functions or their moment generating functions. Vandekerkhove (2010) considers the more specific case of a linear switching regression model, where the switching error distributions are only assumed to be symmetric.

Our paper is organized as follows. In Section 2 we show how the HZ-model can be studied within the framework of model (1.1). Further, we show identification under tail conditions on the characteristic functions of  $F_0$  and  $F_1$ , and in particular conclude that for the HZ-model in the two-dimensional case, an appropriate representation of the component distributions is identified and can be estimated under our assumptions. In Section 3 we construct asymptotically normal estimators of the component distributions and the mixture weights. The main new

technical ingredient is the estimation of limits of quotients of two characteristic functions in the tails from independent samples, which we discuss in Section 4, and which might be of some independent interest. The proofs use strong approximation as well as an entropy-type bound for the characteristic process.

A simulation study is conducted in Section 5, where we focus on the HZ-model in two dimensions. We also propose a cross-validation scheme in order to select the tuning parameters of the estimators. Proofs are deferred to an appendix, while some additional technical results are given in the supplementary material Hohmann and Holzmann (2013).

## 2. Two-component conditional mixtures

In this section, we discuss the model by HZ as our main example of (1.1). Further, we briefly review the identifiability statements of HKS and extend these to cover tail conditions on the characteristic functions of the component distributions.

### 2.1. Examples and model reduction

**Example 1** (Unobserved binary status). Let  $Y$  be some endogenous variable affected by an unobservable binary status  $T \in \{0, 1\}$ . Instead of  $T$  we observe the regressor  $T^*$  which effects the status variable  $T$ , but such that  $Y$  given  $T$  is independent of  $T^*$ . Then

$$\begin{aligned} \mathrm{P}(Y \leq y | T^* = z) &= \sum_{s=0,1} \mathrm{P}(T = s | T^* = z) \mathrm{P}(Y \leq y | T = s) \\ &= (1 - \lambda(z)) F_0(y) + \lambda(z) F_1(y) \end{aligned}$$

with  $F_j(y) = \mathrm{P}(Y \leq y | T = j)$  and  $\lambda(z) = \mathrm{P}(T = 1 | T^* = z)$ . The main example in HKS is the *misclassified* binary status variable, e.g. the case of binary  $T^*$ .  $\square$

**Example 2** (Mixtures with independent components). Suppose that  $X_1, \dots, X_k$  are given results of a medical test of a single person where it is unknown if this person is indeed affected by the disease or not. Let  $T$  denote the indicator for this unknown affection status, i. e.,  $T = 1$  if and only if the person is actually diseased. It is reasonable to assume that the  $k$  test results are stochastically independent given the status  $T$ . Therefore, the observed data can be modeled by the  $k$ -variate two-component mixture

$$F(\mathbf{x}) = (1 - \alpha) \prod_{i=1}^k F_{0,i}(x_i) + \alpha \prod_{i=1}^k F_{1,i}(x_i), \quad \mathbf{x} = (x_1, \dots, x_k)' \in \mathbb{R}^k,$$

with  $\alpha = \mathrm{P}(T = 1)$  and  $F_{j,i}(x) = \mathrm{P}(X_i \leq x | T = j)$ . This model was investigated by HZ who established nonparametric identifiability of the cdfs  $F_{j,i}$  and the mixture weight  $\alpha$  under some mild *irreducibility* condition on  $F$  for the case

$k \geq 3$ . Partial identifiability results were obtained for  $k = 2$ . Let  $i_0 \in \{1, \dots, k\}$  be an arbitrary fixed index, and define

$$Y = X_{i_0}, \quad Z = (X_1, \dots, X_{i_0-1}, X_{i_0+1}, \dots, X_k)'.$$

Due to the conditional independence of the  $X_j$  given  $T$ , we obtain for  $y \in \mathbb{R}$  and  $z \in \mathbb{R}^{k-1}$  that

$$P(Y \leq y | Z = z) = (1 - \lambda(z))F_{0,i_0}(y) + \lambda(z)F_{1,i_0}(y), \quad \lambda(z) = P(T = 1 | Z = z).$$

Note that the function  $\lambda(z)$  is only of minor interest. We may as well condition on  $\{Z \in B\}$  for a Borel set  $B \subset \mathbb{R}^{k-1}$  to obtain

$$\begin{aligned} P(Y \leq y | Z \in B) &= (1 - \pi(B))F_{0,i_0}(y) + \pi(B)F_{1,i_0}(y), \\ \pi(B) &= P(T = 1 | Z \in B) = \alpha P(Z \in B | T = 1) / P(Z \in B). \end{aligned} \quad (2.1)$$

In particular, the weight  $\pi(\mathbb{R}^{k-1})$  determines the parameter  $\alpha$ .  $\square$

Motivated by the HZ-model as well as the case of a misclassified binary status variable, we consider conditioning in model (1.1) on events  $\{Z \in B\}$  for a Borel set  $B \in \mathcal{B}^p$  which have positive probability. Specifically, setting  $\pi(B) = E(\lambda(Z) | Z \in B)$  allows to write (1.1) as

$$F(y|B) = (1 - \pi(B))F_0(y) + \pi(B)F_1(y). \quad (2.2)$$

## 2.2. Identification

Next, we briefly revisit the results on identification in HKS, reformulated in the context of (2.2), and add a condition for identifiability on the quotients of characteristic functions which allows to identify scale mixtures. As HKS, we start with two basic assumptions.

- A1.** 1. There exist  $B_0, B_1 \in \mathcal{B}^p$  such that  $0 < \pi(B_0), \pi(B_1) < 1$ ,  $\pi(B_0) \neq \pi(B_1)$ .  
2. There exists a  $y_0 \in \mathbb{R}$  such that  $F_1(y_0) \neq F_0(y_0)$ .

Given Assumption A1 set

$$\xi = \xi(B_0, B_1) = \frac{\pi(B_1)}{\pi(B_0)} \quad \text{and} \quad \zeta = \zeta(B_0, B_1) = \frac{1 - \pi(B_1)}{1 - \pi(B_0)}.$$

Then direct computations show that

$$\begin{aligned} F_0(y) &= F(y|B_0) + \frac{F(y|B_1) - F(y|B_0)}{1 - \xi}, \\ F_1(y) &= F(y|B_0) + \frac{F(y|B_1) - F(y|B_0)}{1 - \zeta}. \end{aligned} \quad (2.3)$$

Further, set

$$\Lambda(B) = \frac{F(y_0|B) - F(y_0|B_0)}{F(y_0|B_1) - F(y_0|B_0)} = \frac{\pi(B) - \pi(B_0)}{\pi(B_1) - \pi(B_0)}.$$

Straightforward calculations give  $\pi(B_0) = (1 - \zeta)/(\xi - \zeta)$  and  $\pi(B_1) - \pi(B_0) = -(1 - \xi)(1 - \zeta)/(\xi - \zeta)$ , and thus

$$\pi(B) = \frac{1 - \zeta}{\xi - \zeta} - \frac{(1 - \xi)(1 - \zeta)}{\xi - \zeta} \Lambda(B). \tag{2.4}$$

From (2.3) and (2.4), HKS observe that  $F_0, F_1$  and  $\pi$  (and in particular  $\lambda$ ) can be identified from the quantities  $\xi$  and  $\zeta$  and the observable cdf  $F$ . So, under Assumption A1, the identification and estimation of  $\xi$  and  $\zeta$  is the crucial part for the mixture (2.2).

In order to achieve full identification, consider the following tail dominance conditions concerning the component cdfs  $F_0$  and  $F_1$  and their Fourier transforms, say  $\tilde{F}_0$  and  $\tilde{F}_1$ , respectively.

- C1.  $\lim_{y \rightarrow -\infty} F_1(y)/F_0(y) = 0$
- C2.  $\lim_{y \rightarrow +\infty} (1 - F_0(y))/(1 - F_1(y)) = 0$
- C3.  $\lim_{y \rightarrow +\infty} \tilde{F}_0(y)/\tilde{F}_1(y) = 0$

For  $i = 2, 3$ , let  $\mathcal{M}_i$  denote the class of mixtures of the form (2.2), the component cdfs of which satisfy the tail conditions C1 and Ci. HKS state identification under C1 and C2, for convenience, we reformulate their result.

**Theorem 3.** *If  $F \in \mathcal{M}_i$  for  $i = 2$  or  $i = 3$ , then, under A1,  $F_0, F_1$  and  $\pi$  are nonparametrically identifiable within this class  $\mathcal{M}_i$ . Moreover, by C1,*

$$\lim_{y \rightarrow -\infty} \frac{F(y|B_1)}{F(y|B_0)} = \zeta \tag{2.5}$$

and

$$\text{under C2: } \lim_{y \rightarrow \infty} \frac{1 - F(y|B_1)}{1 - F(y|B_0)} = \xi, \quad \text{under C3: } \lim_{y \rightarrow \infty} \frac{\tilde{F}(y|B_1)}{\tilde{F}(y|B_0)} = \xi. \tag{2.6}$$

Let us comment on and illustrate the conditions C1-C3 as well as the statement of the theorem.

Assumptions C1 and C2 from HKS mean that  $F_0$  dominates the left tail of the distribution  $F$ , while  $F_1$  dominates the right tail. This assumption is natural for location mixtures, where it is satisfied for (exponentially) light tails of the underlying distribution. A class of examples is the (skew) normal distribution with equal skewness and scale parameters.

For scale mixtures, Assumptions C1 and C2 are not appropriate. For example, for normal distributions, the component with higher variance dominates both tails. Specifically, consider a normal mixture with  $\sigma_0 > \sigma_1$  and  $\mu_0, \mu_1 \in \mathbb{R}$  arbitrary. Then Assumptions C1 is satisfied and further

$$\frac{\tilde{F}_0(y)}{\tilde{F}_1(y)} = \exp(i(\mu_0 - \mu_1)y) \exp(-(\sigma_0^2 - \sigma_1^2)x^2/2) \rightarrow 0, \quad |x| \rightarrow \infty,$$

thus Assumptions **C1** and **C3** provide identification. More generally, for a scale mixture of a supersmooth density (for which the characteristic function decays at an exponential rate), **C3** is satisfied. Thus, **C3** allows to smoothly separate scale-mixtures.

**Example 2 (continued).** We show in the context of Example 2 that Theorem 3 only states that the specific representation of the conditional mixtures for which the components satisfy the conditions **C1** and **C2** or **C3** is identified, there might be further representations of the form (2.2). Nevertheless, as argued above, these representations are quite natural: **C1** and **C2** are appropriate if  $F_0$  and  $F_1$  dominate distinct tails of the distribution, while **C1** and **C3** are natural for a scale-type mixture in a smooth density with light tails.

In terms of densities, the model is

$$f(\mathbf{x}) = (1 - \pi)f_{0,1}(x_1)f_{0,2}(x_2) + \pi f_{1,1}(x_1)f_{1,2}(x_2), \quad \mathbf{x} = (x_1, x_2)' \in \mathbb{R}^2. \quad (2.7)$$

Let  $f_1$  and  $f_2$  denote the one-dimensional marginals of  $f$ . Then theorem 4.1 in HZ provides the factorization

$$f(\mathbf{x}) - f_1(x_1)f_2(x_2) = g_1(x_1)g_2(x_2),$$

where the functions  $g_1$  and  $g_2$  are uniquely determined up to constant multiples. Now, theorem 4.2 in HZ states the partial identifiability of (2.7) as follows. If  $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2)$  is a vector of real constants unequal to zero such that  $\text{sgn } \alpha_j = -\text{sgn } \beta_j$ , the fraction  $|\beta_j|/(|\alpha_j| + |\beta_j|)$  does not depend on  $j$ , and such that

$$f_{0,j}^\theta := f_j + \alpha_j g_j, \quad f_{1,j}^\theta := f_j + \beta_j g_j \quad (2.8)$$

are non-negative, then these latter functions are probability densities which also fulfill (2.7), with mixture weight  $1 - \pi^\theta := |\beta_1|/(|\alpha_1| + |\beta_1|)$ . Denote by  $\Theta$  the set of corresponding shifting vectors  $\theta$ . We show that our tail assumptions identify a unique value in  $\Theta$ . Indeed, suppose that  $\theta^* \in \Theta$  is such that

$$0 = \lim_{x \rightarrow -\infty} \frac{f_{1,1}^{\theta^*}(x)}{f_{0,1}^{\theta^*}(x)} = \lim_{x \rightarrow -\infty} \frac{1 + \beta_1^* g_1(x)/f_1(x)}{1 + \alpha_1^* g_1(x)/f_1(x)}, \quad (2.9)$$

corresponding to condition **C1** of section 2 in terms of densities. Since  $\beta_1^* \neq 0$ , this implies the convergence

$$\lim_{x \rightarrow -\infty} g_1(x)/f_1(x) = -1/\beta_1^*,$$

which by (2.7) yields the identification of  $f_1$ . The arguments under **C2** and **C3** are similar.  $\square$

The nonidentifiability without additional assumptions on the components in model (2.2) holds more generally, see Example 1 in Hohmann and Holzmann (2013). Further, without the additional regressor  $Z$  even in case of a known weight and tail conditions on the components, the model is not identified, see Example 2 in Hohmann and Holzmann (2013).

### 3. Estimation

In this section, given i.i.d. observations  $(Y_1, Z_1), \dots, (Y_n, Z_n)$  such that the conditional distribution  $F(y|z)$  satisfies (2.2), following HKS we propose nonparametric estimators of the components  $F_0, F_1$  and of the weight function  $\pi$ . The essential step is to estimate the quantities  $\zeta$  and  $\xi$ , then, based on (2.3) and (2.4), plug-in estimates are easily devised. Given the tail conditions (C1) and (C2)/(C3), we estimate  $\zeta$  and  $\xi$  as the limits arising in (2.5) and (2.6). To this end, Section 4 contains the asymptotic distribution theory for limits of quotients of characteristic functions, our major contribution, as well as of distribution functions, which is essentially covered in HKS, of independent samples in their tails. Here, we apply this theory to obtain asymptotics for the estimators in model (2.2).

#### 3.1. Estimation of $\zeta$ and $\xi$ under C1 and C3

Consider the empirical conditional distribution and characteristic functions

$$F_n(y|B_j) = \frac{\sum_{k=1}^n \mathbf{1}_{\{Y_k \leq y\}} \mathbf{1}_{\{Z_k \in B_j\}}}{\sum_{k=1}^n \mathbf{1}_{\{Z_k \in B_j\}}}, \quad y \in \mathbb{R},$$

$$\tilde{F}_n(t|B_j) = \frac{\sum_{k=1}^n \exp(it Y_k) \mathbf{1}_{\{Z_k \in B_j\}}}{\sum_{k=1}^n \mathbf{1}_{\{Z_k \in B_j\}}}, \quad t \in \mathbb{R}.$$

Motivated by (2.5) and (2.6), following HKS we consider estimators for  $\zeta$  and  $\xi$  of the form

$$\zeta_n = \frac{F_n(L_n|B_1)}{F_n(L_n|B_0)}, \quad \xi_n = \operatorname{Re} \frac{\tilde{F}_n(R_n|B_1)}{\tilde{F}_n(R_n|B_0)},$$

where the levels  $L_n$  and  $R_n$  need to be chosen appropriately. To this end, assume

**A2.** The Borel sets  $B_0$  and  $B_1$  in A1 i. satisfy  $B_0 \cap B_1 = \emptyset$  and  $p_j := P(Z \in B_j) > 0, j = 0, 1$ .

Under A2, we define disjoint subsamples  $Y_1^*, \dots, Y_{m_n}^*$  and  $Y_1^{**}, \dots, Y_{l_n}^{**}$  of  $Y_1, \dots, Y_n$ , where the  $Y_j^*$  correspond to the observations  $Y_k$  such that  $Z_k \in B_0$ , the  $Y_j^{**}$  correspond to  $Z_k \in B_1$ , and set  $m_n = \sum_{k=1}^n \mathbf{1}_{\{Z_k \in B_0\}}$  and  $l_n = \sum_{k=1}^n \mathbf{1}_{\{Z_k \in B_1\}}$ .

We choose  $L_n$  as an intermediate lower order statistic of the subsample  $Y_1^*, \dots, Y_{m_n}^*$ : Let  $L_n$  be the  $[r_n]$ -th largest order statistic of the  $Y_k^*$ ,  $L_n = Y_{m_n([r_n])}^*$ , where  $r_n \rightarrow \infty$  is such that

$$r_n/n \rightarrow 0, \quad r_n/\sqrt{n} \rightarrow \infty \quad \text{as } n \rightarrow \infty. \tag{3.1}$$

In order to choose  $R_n$ , we assume that the characteristic function satisfies  $\tilde{F}(t|B_0) \rightarrow 0$  as  $t \rightarrow \infty$ . Let  $s_n \rightarrow \infty$  be such that  $s_n/n \rightarrow 0$ . Then, for large  $n$ , by continuity of  $\tilde{F}(y|B_0)$  there is a (not necessarily unique) solution  $t_n$  of the equation  $\tilde{F}(t_n|B_0) = \sqrt{s_n/m_n}$ , and we choose  $R_n$  as a solution of

the corresponding empirical version of this equation:  $|\tilde{F}_n(R_n|B_0)| = \sqrt{s_n/m_n}$ . More precisely, we require

**A3.** There exists  $\gamma > 0$  and a non-random sequence  $t_n \rightarrow \infty$  such that (4.7)-(4.10) (see Section 4) hold true for  $F = F(\cdot|B_1)$ ,  $G = F(\cdot|B_0)$ , and  $h_n = R_n$ .

For further discussion see Section 4. Finally, assume that the rates  $r_n$  and  $s_n$  are chosen such that there exist constants  $\beta_\zeta$  and  $\beta_\xi$  in  $\mathbb{R}$  (possibly zero) for which

$$\sqrt{r_n} \left( \frac{F(L_n|B_1)}{F(L_n|B_0)} - \zeta \right) \rightarrow_p \beta_\zeta, \quad \sqrt{s_n} \left( \operatorname{Re} \frac{\tilde{F}(R_n|B_1)}{\tilde{F}(R_n|B_0)} - \xi \right) \rightarrow_p \beta_\xi \quad (3.2)$$

as  $n \rightarrow \infty$ .

**Proposition 4.** *If A1-A3 and (3.2) hold, then*

$$\sqrt{r_n}(\zeta_n - \zeta) \rightsquigarrow \mathcal{N}(\beta_\zeta, \tau\zeta + \zeta^2), \quad \sqrt{s_n}(\xi_n - \xi) \rightsquigarrow \mathcal{N}(\beta_\xi, (\tau + \xi^2)/2)$$

with  $\tau = p_0/p_1$ . Moreover, if  $s_n = r_n$ , then the estimators are asymptotically independent.

The first part of the proposition is as in HKS, the second (which is based on characteristic functions) as well as the asymptotic independence are our main contributions to estimation. Under the Assumptions C1 and C2, we obtain analogous results to those in HKS, see Hohmann and Holzmann (2013) for the details.

For further discussion of the tail assumption (3.2) see HKS, Lemmas 7 and 8. Note that it also applies to the characteristic functions if these (and not the distributions functions) satisfy the shape constraints made in HKS.

### 3.2. Estimating the component distributions and the weight function

We now turn to the estimation of the components distributions  $F_0$  and  $F_1$  and the mixture weight function  $\pi$ . We obtain similar, though slightly more refined results as HKS.

By (2.3), natural estimates of  $F_0$  and  $F_1$  are given by

$$\begin{aligned} \hat{F}_0(y) &= F_n(y|B_0) + \frac{F_n(y|B_1) - F_n(y|B_0)}{1 - \xi_n}, \\ \hat{F}_1(y) &= F_n(y|B_0) + \frac{F_n(y|B_1) - F_n(y|B_0)}{1 - \zeta_n}, \end{aligned}$$

where  $\zeta_n$  is obtained using (C1) and  $\xi_n$  either from (C2) or from (C3). As a consequence of Proposition 4,

$$\sqrt{r_n}(\zeta_n - \zeta) \rightsquigarrow \mathcal{N}(\beta_\zeta, \sigma_\zeta^2), \quad \sqrt{s_n}(\xi_n - \xi) \rightsquigarrow \mathcal{N}(\beta_\xi, \sigma_\xi^2), \quad (3.3)$$

where the variances  $\sigma_\zeta^2$  and  $\sigma_\xi^2$  are given as in Proposition 4, and possibly  $r_n = s_n$ , in which case the estimators are asymptotically independent. Then we have



**Theorem 5.** *Suppose that (3.3) holds, where  $s_n$  and  $r_n$  satisfy (3.1). Then*

$$\sqrt{s_n}(\hat{F}_0(y) - F_0(y)) \overset{\ell^\infty(\mathbb{R})}{\rightsquigarrow} \mathbf{G}_\xi(y), \quad \sqrt{r_n}(\hat{F}_1(y) - F_1(y)) \overset{\ell^\infty(\mathbb{R})}{\rightsquigarrow} \mathbf{G}_\zeta(y),$$

where  $\mathbf{G}_i$ ,  $i = \xi, \zeta$ , are tight Gaussian processes with mean and covariance functions

$$\mu_i(y) = D(F, y)(1 - i)^{-2}\beta_i, \quad \rho_i(y_1, y_2) = D(F, y_1)D(F, y_2)(1 - i)^{-4}\sigma_i^2,$$

$i = \xi, \zeta$ , where  $D(F, y) = F(y|B_1) - F(y|B_0)$ .

We note that relations analogous to (2.3) also hold true for underlying densities, and hence that similar estimators for the densities could be devised.

Finally, we consider estimation of the mixture weight function  $\pi(B)$  for sets  $B \in \mathcal{B}^p$  with  $P(Z \in B) > 0$ ,  $B = \mathbb{R}^p$  being of particular interest. Fix a  $y_0$  satisfying Assumption A1. From (2.4), a suitable estimator is given by

$$\pi_n(B) = L_1(\xi_n, \zeta_n) - L_2(\xi_n, \zeta_n) \frac{F_n(y_0|B) - F_n(y_0|B_0)}{F_n(y_0|B_1) - F_n(y_0|B_0)},$$

where  $L_1(x_1, x_2) = (1 - x_2)/(x_1 - x_2)$ ,  $L_2(x_1, x_2) = (1 - x_1)(1 - x_2)/(x_1 - x_2)$ ,  $x_1 \neq x_2$ .

**Theorem 6.** *Let  $B \in \mathcal{B}^p$  with  $P(Z \in B) > 0$ , and assume (3.3), where in case of equal rates we additionally assume asymptotic independence.*

1. *If  $s_n = r_n$ , we have*

$$\sqrt{r_n}(\pi_n(B) - \pi(B)) \rightsquigarrow \mathcal{N}((\beta_\xi, \beta_\zeta)\mathbf{J}, \mathbf{J}' \text{diag}(\sigma_\xi^2, \sigma_\zeta^2)\mathbf{J}),$$

where

$$\mathbf{J} = (\xi - \zeta)^{-2} \left( \begin{pmatrix} \zeta - 1 \\ 1 - \xi \end{pmatrix} - \Lambda(B) \begin{pmatrix} -(1 - \zeta)^2 \\ (1 - \xi)^2 \end{pmatrix} \right).$$

2. *If  $s_n = o(r_n)$ , then*

$$\sqrt{s_n}(\pi_n(B) - \pi(B)) \rightsquigarrow \mathcal{N}(\beta_\xi \mathbf{j}, \mathbf{j}^2 \sigma_\xi^2), \quad \mathbf{j} = (\xi - \zeta)^{-2}(\zeta - 1 + \Lambda(B)(1 - \zeta)^2).$$

**Remark.** When estimating the mixture weights  $\pi(B_0)$  and  $\pi(B_1)$  of the sets  $B_0$  and  $B_1$  upon which the estimation procedure is based, the asymptotic covariance matrix has a simpler form: Since  $\Lambda(B_0) = 0$  and  $\Lambda(B_1) = 1$ , we obtain that

$$\mathbf{J}(B_0, B_1) = \frac{1}{(\xi - \zeta)^2} \begin{pmatrix} -(1 - \zeta) & 1 - \xi \\ -\zeta(1 - \zeta) & \xi(1 - \xi) \end{pmatrix}.$$

#### 4. Estimating quotients in the tails

Let  $X_1, X_2, \dots$  and  $Y_1, Y_2, \dots$  be mutually independent sequences of i.i.d. observations with distribution functions  $F$  and  $G$ , respectively. Assume that

$$F(y)/G(y) \rightarrow \theta \quad \text{as } y \rightarrow -\infty \tag{4.1}$$

or / and

$$\tilde{F}(y), \tilde{G}(y) \rightarrow 0, \quad \tilde{F}(y)/\tilde{G}(y) \rightarrow \eta \quad \text{as } y \rightarrow \infty \quad (4.2)$$

hold for some  $\theta > 0$  and  $\eta \in \mathbb{C} \setminus \{0\}$ , where as above  $\tilde{F}$  and  $\tilde{G}$  denote the characteristic functions of  $F$  and  $G$ . We shall construct asymptotically normal estimators of  $\theta$  and  $\eta$ . In the following, suppose that  $l_n$  and  $m_n$  are sequences in  $\mathbb{N}$  such that  $l_n, m_n \asymp n$  as  $n \rightarrow \infty$ .

#### 4.1. Characteristic functions

To estimate  $\eta$  in (4.2), let

$$\eta_n = \tilde{F}_n(h_n)/\tilde{G}_n(h_n),$$

where

$$\tilde{F}_n(y) = \frac{1}{l_n} \sum_{k=1}^{l_n} \exp(iyX_k), \quad \tilde{G}_n(y) = \frac{1}{m_n} \sum_{k=1}^{m_n} \exp(iyY_k),$$

with  $h_n$  a sequence tending to infinity. Decompose

$$\eta_n - \eta = (\eta_n - \bar{\eta}_n) + (\bar{\eta}_n - \eta), \quad \bar{\eta}_n = \tilde{F}(h_n)/\tilde{G}(h_n).$$

In order to handle the ‘‘variance term’’, write

$$\sqrt{s_n}(\eta_n - \bar{\eta}_n) = \frac{\sqrt{s_n/m_n}}{\tilde{G}_n(h_n)} \left( \sqrt{m_n/l_n} \tilde{\mathbf{F}}_n(h_n) - \bar{\eta}_n \tilde{\mathbf{G}}_n(h_n) \right), \quad (4.3)$$

where  $\tilde{\mathbf{F}}_n = \sqrt{l_n}(\tilde{F}_n - \tilde{F})$  and  $\tilde{\mathbf{G}}_n = \sqrt{m_n}(\tilde{G}_n - \tilde{G})$  are the characteristic processes and  $s_n \rightarrow \infty$ . Assume that  $s_n$  satisfies

$$s_n/n \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad (4.4)$$

and that  $h_n \rightarrow_p \infty$  is chosen such that

$$|\tilde{G}_n(h_n)| = \sqrt{s_n/m_n}(1 + o_P(1)). \quad (4.5)$$

We shall use strong approximations of the characteristic processes by

$$\mathbf{C}(y) = \int \exp(iyx) \mathbf{B}(F(dx)) \quad (4.6)$$

for  $\tilde{\mathbf{F}}_n(y)$ , and similarly for  $\tilde{\mathbf{G}}_n$ . In order that these processes are sample-continuous and that strong approximations work, some conditions on  $F$  and  $G$  are required, see Csörgő (1981). We shall adopt the following sufficient condition: Assume that there exists  $\gamma > 0$  such that

$$y^\gamma H(-y) + y^\gamma(1 - H(y)) = O(1) \quad \text{as } y \rightarrow \infty, \quad H = F, G. \quad (4.7)$$

Finally, we assume that there also exists a non-random sequences  $t_n \rightarrow \infty$  such that

$$t_n = o(n^{\gamma/(2\gamma+4)}(\log n)^{-(\gamma+1)/(\gamma+2)}), \tag{4.8}$$

$$|h_n - t_n| = o_P(1), \tag{4.9}$$

$$|\tilde{G}(h_n) - \tilde{G}(t_n)| = o_P(\sqrt{s_n/m_n}), \tag{4.10}$$

with  $\gamma$  determined by (4.7).

**Remark.** Given a non-random sequence  $t_n \rightarrow \infty$  of order (4.8), assume that the following separability criterion holds: There is a sequence  $a_n$ , either constant or tending to infinity, such that for all  $\varepsilon > 0$  there exists  $C_\varepsilon > 0$  fulfilling

$$\inf_{a_n |t_n - t_n| > \varepsilon} |\sqrt{m_n/s_n} |\tilde{G}(t)| - 1| > C_\varepsilon \tag{4.11}$$

for  $n$  sufficiently large. In case of a supersmooth density, one can show that (4.11) holds with  $a_n = t_n$ , and that this rate implies (4.10) if  $t_n$  is chosen such that  $|\tilde{G}(t_n)| = \sqrt{s_n/m_n}$ .

**Lemma 7.** Assume that (4.7) and (4.11) hold. If  $t_n$  is a non-random sequence of order (4.8), then there exists a sequence  $h_n$  such that  $\tilde{G}_n(h_n) = \sqrt{s_n/m_n}$  and  $|h_n - t_n| = o_P(a_n^{-1})$ .

**Example 8.** Consider exemplarily the Gaussian characteristic function

$$\tilde{\Phi}(x) = \tilde{\Phi}(x; \mu, \sigma^2) = \exp(i\mu x) \exp(-\sigma^2 x^2/2),$$

and let  $t_n$  be chosen such that  $|\tilde{\Phi}(t_n)| = \sqrt{s_n/m_n}$ . Given  $\varepsilon > 0$ , the infimum in (4.11) is attained at  $t_n^* = t_n + \varepsilon/a_n$ . A Taylor expansion then yields

$$|\sqrt{m_n/s_n} \tilde{\Phi}(t_n^*) - 1| = \sqrt{m_n/s_n} \sigma^2 \bar{t}_n \exp(-\sigma^2 \bar{t}_n^2/2) \varepsilon/a_n$$

with  $\bar{t}_n \in [t_n, t_n + \varepsilon/a_n]$ . Choosing  $a_n = t_n$ , it follows that  $\bar{t}_n/a_n \rightarrow 1$  and

$$\exp(-\sigma^2 \bar{t}_n^2/2) \geq \exp(-\sigma^2 (t_n + \varepsilon/a_n)^2/2) = \sqrt{s_n/m_n} \exp(-\sigma^2 (2\varepsilon + (\varepsilon/a_n)^2)/2),$$

and thus (4.11) holds. As a result, by Lemma 7 there exists a random sequence  $h_n$  such that  $|\tilde{\Phi}_n(h_n)| = \sqrt{s_n/m_n}$  and  $|h_n - t_n| = o_P(t_n^{-1})$ . Now, again by a Taylor expansion, similar arguments show that for some  $\bar{t}_n$  between  $h_n$  and  $t_n$ ,

$$\begin{aligned} \sqrt{m_n/s_n} |\operatorname{Re} \Phi(h_n) - \operatorname{Re} \Phi(t_n)| &= \sqrt{m_n/s_n} (\sigma^2 \bar{t}_n \cos(\mu \bar{t}_n) + \mu \sin(\mu \bar{t}_n)) \\ &\quad \exp(-\sigma^2 \bar{t}_n^2/2) |h_n - t_n| \\ &= o_P(1). \end{aligned}$$

Since the imaginary part of  $\tilde{\Phi}(h_n) - \tilde{\Phi}(t_n)$  can be handled likewise, we also see that (4.10) is fulfilled.  $\square$

**Theorem 9.** Assume that (4.4), (4.5) and (4.7)-(4.10) hold. If there exists  $\tau > 0$  such that  $m_n/l_n \rightarrow \tau$ , then

$$\sqrt{s_n}(\eta_n - \bar{\eta}_n) \rightsquigarrow \mathcal{CN}(0, \tau + |\eta|^2, 0),$$

where  $\mathcal{CN}$  denotes the complex-normal distribution. More explicitly,

$$\sqrt{s_n} \begin{pmatrix} \operatorname{Re}(\eta_n - \bar{\eta}_n) \\ \operatorname{Im}(\eta_n - \bar{\eta}_n) \end{pmatrix} \rightsquigarrow \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} \tau + |\eta|^2 & 0 \\ 0 & \tau + |\eta|^2 \end{pmatrix} \right).$$

#### 4.2. Distribution functions

To estimate  $\theta$  in (4.1), let

$$\theta_n = F_n(h_n)/G_n(h_n), \quad F_n(y) = \frac{1}{l_n} \sum_{k=1}^{l_n} \mathbf{1}_{\{X_k \leq y\}}, \quad G_n(y) = \frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{1}_{\{Y_k \leq y\}},$$

where the level  $h_n$  is specified below. Write

$$\theta_n - \theta = (\theta_n - \bar{\theta}_n) + (\bar{\theta}_n - \theta), \quad \bar{\theta}_n = F(h_n)/G(h_n).$$

Assume that  $r_n \rightarrow \infty$  satisfies (3.1), and that  $h_n \rightarrow_p -\infty$  is chosen such that

$$G_n(h_n) = r_n/m_n + o_P(r_n/n) = r_n/m_n (1 + o_P(1)). \quad (4.12)$$

(4.12) is satisfied if we choose in particular  $h_n = Y_{m_n(\lfloor r_n \rfloor)}$ , where  $\lfloor r_n \rfloor$  is the largest integer smaller than  $r_n$ , and where  $Y_{m_n(\lfloor r_n \rfloor)}$  denotes the  $\lfloor r_n \rfloor$ -th largest order statistic of the sample  $Y_1, \dots, Y_{m_n}$ , since  $G_n(h_n) = \lfloor r_n \rfloor / m_n = r_n/m_n(1 + o(1))$ .

**Theorem 10.** Suppose that the assumptions of Theorem 9 for  $s_n = r_n$  as well as (3.1) and (4.12) hold. If there exists  $\tau > 0$  such that  $m_n/l_n \rightarrow \tau$ , then

$$\sqrt{r_n} \begin{pmatrix} \theta_n - \bar{\theta}_n \\ \operatorname{Re}(\eta_n - \bar{\eta}_n) \\ \operatorname{Im}(\eta_n - \bar{\eta}_n) \end{pmatrix} \rightsquigarrow \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 2(\tau\theta + \theta^2) & 0 & 0 \\ 0 & \tau + |\eta|^2 & 0 \\ 0 & 0 & \tau + |\eta|^2 \end{pmatrix} \right).$$

The asymptotic distribution of  $r_n(\theta_n - \bar{\theta}_n)$  follows from arguments along the lines of HKS, however, the asymptotic independence requires some additional work. Further applications, discussed in Hohmann and Holzmann (2013), are testing for tail dominance, as well as estimating the exponent of regular variation.

### 5. Simulation study

We investigate the finite-sample performance of the estimators in a simulation study in the HZ-model. Consider a random vector  $(Y, Z)'$  distributed according to

$$F_{Y,Z}(y, z) = (1 - p) \Psi(y; \mu_0, \sigma_0^2, \lambda_0) \Phi(z; 0, 1) + p \Psi(y; \mu_1, \sigma_1^2, \lambda_1) \Phi(z; 3, 1),$$

where  $\Phi$  and  $\Psi$  denote the normal and skew-normal distribution functions, resp. The distribution of  $Y$  conditional on  $Z$  is then given by (cf. (2.1))

$$F(y|B) = (1 - \pi(B)) \Psi(y; \mu_0, \sigma_0^2, \lambda_0) + \pi(B) \Psi(y; \mu_1, \sigma_1^2, \lambda_1). \tag{5.1}$$

As true values we choose  $(\mu_0, \sigma_0^2, \lambda_0) = (0, 1, 2)$ ,  $(\mu_1, \sigma_1^2, \lambda_1) = (0.5, 0.5, 3)$ , and  $p = 0.6$ . Let us show that the component distributions in (5.1) then fulfill the tail dominance assumptions C1 and C3.

The density of the skew-normal distribution is given by

$$\psi(x; \mu, \sigma^2, \lambda) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\lambda \frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R},$$

where  $\phi$  denotes the standard normal density, and its characteristic function by

$$\tilde{\Psi}(x; \mu, \sigma^2, \lambda) = e^{i\mu x - \frac{1}{2}\sigma^2 x^2} \left(1 + i \mathfrak{J}\left(\sigma x \lambda / \sqrt{1 + \lambda^2}\right)\right),$$

where  $\mathfrak{J}(x) = \int_0^x \sqrt{2/\pi} e^{u^2/2} du$ . Hence,

$$\Psi(x; \mu, \sigma^2, \lambda) \sim \frac{1}{\lambda \pi(\mu - x)} e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2 (1 + \lambda^2)}, \quad x \rightarrow -\infty,$$

and, as  $x \rightarrow \infty$ ,

$$\begin{aligned} |\tilde{\Psi}(x; \mu, \sigma^2, \lambda)| &= e^{-\sigma^2 x^2/2} \sqrt{1 + \mathfrak{J}\left(\sigma x \lambda / \sqrt{1 + \lambda^2}\right)^2} \\ &\sim \sqrt{\frac{2}{\pi} \frac{\sqrt{1 + \lambda^2}}{\lambda \sigma x}} e^{-\frac{1}{2}\sigma^2 x^2 (1 + \lambda^2)^{-1}}, \end{aligned}$$

and thus the condition

$$\sigma_0^2/\sigma_1^2 > (1 + \lambda_0^2)/(1 + \lambda_1^2)$$

is sufficient and necessary for both C1 and C3 to hold true.

For the estimation we further chose  $B_0 = (0, \infty)$  and  $B_1 = (-\infty, 0]$ , inducing true values  $\zeta = 2.3876$  and  $\xi = 0.4502$ , and set  $y_0 = 1$ . The estimation results for  $\xi$ ,  $\zeta$ , and  $p$ , using different sample sizes  $n$  and different rates  $r_n = s_n = n^\delta$ , are presented in Table 1 (a)-(c).

The choice of  $r_n$  turns out to highly affect the estimates' variance and bias properties. A small  $r_n$  leads to small bias, it however increases the variance, as should be expected from the theory. Therefore, in a second step we use a cross-validation scheme to choose  $r_n$ . This can for example be done by a repeated random sub-sampling validation, i. e., one randomly splits up the sample of observations into two sub-samples of equal size, uses these sub-samples to estimate both the mixture

$$(1 - \hat{p}) \hat{F}_0(x) + \hat{p} \hat{F}_1(x),$$

for the given  $r_n$ , where  $\hat{p} = \pi_n(\mathbb{R})$ , and the ordinary empirical distribution function of  $Y$ . One estimates the  $L_1$ -distance between these estimates, the cross-validated  $r_n$  is then the minimizer (on some fixed grid). The estimates for  $\xi$ ,  $\zeta$ , and  $p$  using cross-validation can be found in Table 1 (d). Also, Figure 1 shows estimates of the distribution functions  $F_0$  and  $F_1$  using cross-validation.

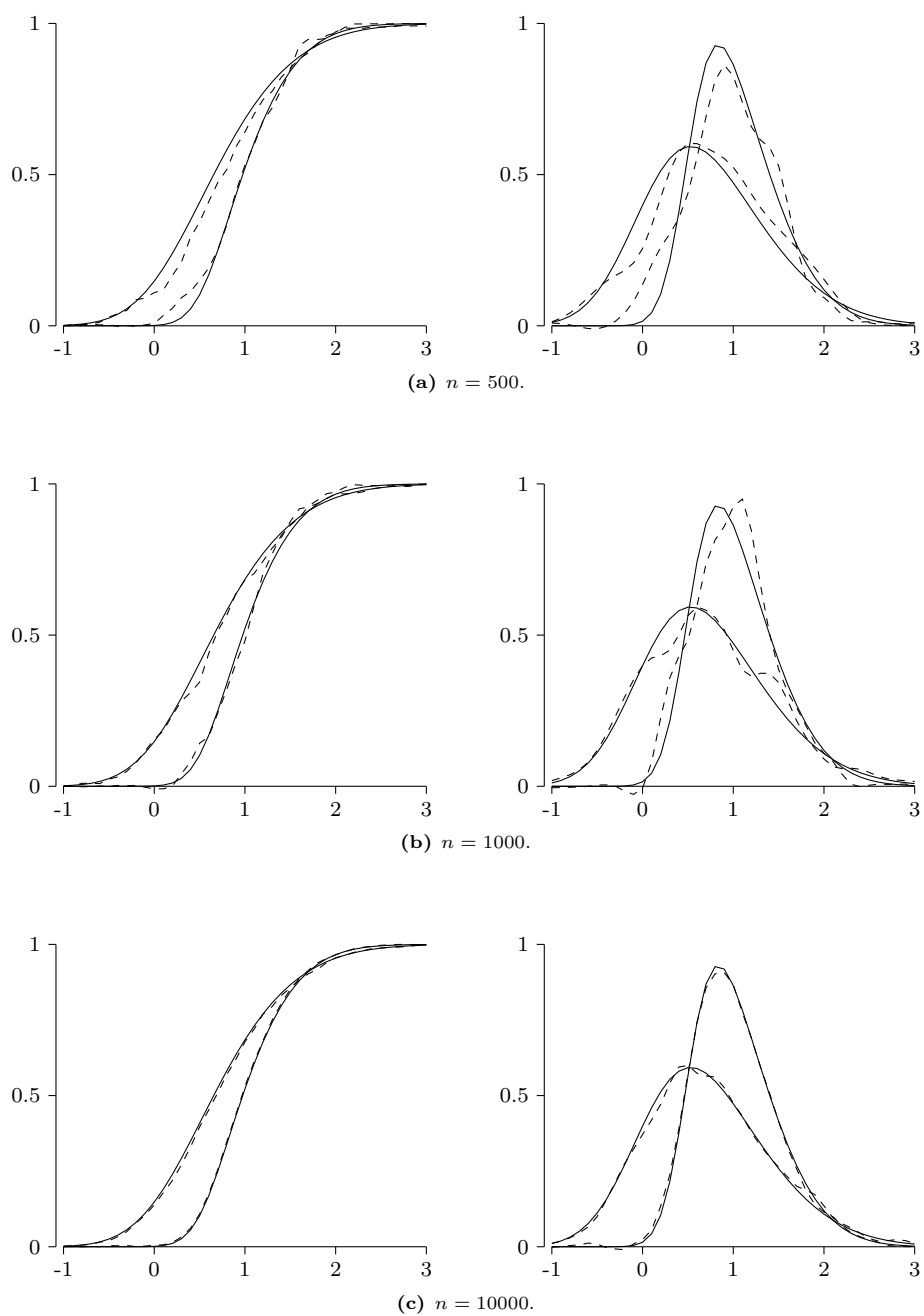


FIG 1. Estimates  $\hat{F}_0$  and  $\hat{F}_1$  (dashed) of the component distribution functions  $F_0$  and  $F_1$  (solid) for different sample sizes  $n$ . The right column shows estimates of the corresponding densities.

TABLE 1

Estimates of  $\zeta$ ,  $\xi$ , and  $\pi$  in the model of Hall and Zhou (2003) for true values  $\zeta = 2.3876$ ,  $\xi = 0.4502$ , and  $p = 0.6$ . The table shows mean values and standard deviations (in brackets) of  $10^4$  repetitions, based on different sample sizes  $n$  and different rates  $r_n = s_n = n^\delta$

(a) Estimates using  $\delta = 0.4$ .

$n$	$\zeta_n$	$\xi_n$	$p_n$
500	2.3346 (0.9088)	0.3634 (0.3974)	0.5266 (0.2544)
1000	2.3736 (0.8536)	0.3626 (0.3433)	0.5308 (0.2311)
10000	2.3826 (0.5371)	0.3995 (0.2089)	0.5649 (0.1321)

(b) Estimates using  $\delta = 0.6$ .

$n$	$\zeta_n$	$\xi_n$	$p_n$
500	1.9299 (0.3493)	0.5129 (0.2249)	0.5685 (0.1319)
1000	2.0911 (0.3201)	0.4755 (0.1825)	0.5734 (0.0976)
10000	2.3657 (0.2009)	0.3967 (0.0853)	0.5693 (0.0495)

(c) Estimates using  $\delta = 0.8$ .

$n$	$\zeta_n$	$\xi_n$	$p_n$
500	1.2328 (0.1147)	0.7408 (0.1128)	0.4356 (0.1725)
1000	1.2935 (0.0935)	0.7163 (0.0876)	0.4663 (0.1029)
10000	1.5335 (0.0440)	0.6305 (0.0359)	0.5266 (0.0241)

(d) Estimates using cross validation for  $\delta$ .

$n$	$\zeta_n$	$\xi_n$	$p_n$	$\delta_n$
500	2.2586 (0.8189)	0.3827 (0.3668)	0.5301 (0.2027)	0.4675 (0.0942)
1000	2.2757 (0.6341)	0.3816 (0.2631)	0.5392 (0.1341)	0.5181 (0.0863)
10000	2.3657 (0.2652)	0.3894 (0.0983)	0.5649 (0.0528)	0.5772 (0.0552)

### Acknowledgements

The authors gratefully acknowledge financial support from the DFG, grant Ho 3260/3-1. Comments from the associate editor and a referee helped to improve the presentation of the paper.

### Appendix A: Proofs

#### A.1. Proofs of Section 4

*Proof of Lemma 7.* Let  $\delta > 0$ , and without loss of generality assume that  $|\tilde{G}(x)|$  is non-increasing for  $x$  sufficiently large. Then, by (4.11) there exists  $0 < C_\delta < 1$

such that

$$|\tilde{G}(t_n + \delta)| \leq (1 - C_\delta) \sqrt{s_n/m_n}.$$

From Lemma 11 and (A.3) it follows that  $|\tilde{G}_n(t_n + \delta) - \tilde{G}(t_n + \delta)| = o_P(\sqrt{s_n/m_n})$ , yielding

$$||\tilde{G}_n(t_n + \delta)| - |\tilde{G}(t_n + \delta)|| < C_\delta/2 + o_P(1)$$

and thus

$$|\tilde{G}_n(t_n + \delta)| < \sqrt{s_n/m_n} + o_P(1).$$

As a result, there exists a sequence  $h_n$  such that  $\tilde{G}_n(h_n) = \sqrt{s_n/m_n}$ , where  $h_n$  fulfills the order condition (4.8), too. Hence, it also holds that  $|\tilde{G}_n(h_n) - \tilde{G}(h_n)| = o_P(\sqrt{s_n/m_n})$ , yielding

$$|\sqrt{m_n/s_n}|\tilde{G}(h_n)| - 1| = \sqrt{m_n/s_n}||\tilde{G}_n(h_n)| - |\tilde{G}(h_n)|| = o_P(1).$$

Since, for all  $\varepsilon > 0$ , (4.11) implies

$$\{a_n|h_n - t_n| > \varepsilon\} \subset \{|\sqrt{m_n/s_n}|\tilde{G}(h_n)| - 1| > C_\varepsilon\},$$

we conclude that  $|h_n - t_n| = o_P(a_n^{-1})$ .  $\square$

#### Proof of Theorem 9

The proof of Theorem 9 proceeds in several steps.

**Lemma 11.** *Assume that (4.7) holds. On a sufficiently rich probability space there exist versions of the  $X_k$  and  $Y_k$ , and independent sequences  $\mathbf{B}_{1,n}$  and  $\mathbf{B}_{2,n}$  of standard Brownian bridges on  $[0, 1]$  such that, defining*

$$\mathbf{C}_{1,n}(y) = \int \exp(iyx) \mathbf{B}_{1,n}(F(dx)), \quad \mathbf{C}_{2,n}(y) = \int \exp(iyx) \mathbf{B}_{2,n}(G(dx)), \quad (\text{A.1})$$

for all sequences  $T_n = o(n^{\gamma/(2\gamma+4)}(\log n)^{-(\gamma+1)/(\gamma+2)})$  it holds that

$$\sup_{0 \leq y \leq T_n} |\tilde{\mathbf{F}}_n(y) - \mathbf{C}_{1,l_n}(y)| \rightarrow_p 0, \quad \sup_{0 \leq y \leq T_n} |\tilde{\mathbf{G}}_n(y) - \mathbf{C}_{2,m_n}(y)| \rightarrow_p 0.$$

*Proof.* According to Theorem 4 and Corollary 2 in Csörgő (1981), for all  $T, \delta > 0$  and  $n \in \mathbb{N}$  there exists a constant  $C$  which depends only on  $\delta$  and  $F$  such that

$$\mathbb{P}\left(\sup_{0 \leq y \leq T} |\tilde{\mathbf{F}}_n(y) - \mathbf{C}_{1,l_n}(y)| > Cq_n T\right) \lesssim Tn^{-(1+\delta)},$$

where the sequence  $q_n$  satisfies  $q_n \sim n^{-\gamma/(2\gamma+4)}(\log n)^{(\gamma+1)/(\gamma+2)}$ . Hence, for all sequences  $T_n = o(q_n^{-1})$  and all  $\varepsilon > 0$  we find that

$$\mathbb{P}\left(\sup_{0 \leq y \leq T_n} |\tilde{\mathbf{F}}_n(y) - \mathbf{C}_{1,l_n}(y)| > \varepsilon\right) \lesssim T_n n^{-(1+\delta)}$$

eventually, where the right side converges to zero as  $n \rightarrow \infty$ .  $\square$



**Remark.** The processes  $\mathbf{C}_{i,n}$  are zero mean complex Gaussian. In particular, with  $\mathbf{C}$  being defined as in (4.6) and  $\mathbf{W}$  a standard Brownian motion on  $[0, 1]$ ,

$$\mathbf{C}(t) =_d \int_0^1 \exp(itF^{-1}(x)) \mathbf{W}(dx) - \mathbf{W}(1) \int_0^1 \exp(itF^{-1}(x)) dx.$$

With this, basic properties of the Itô integral give

$$\begin{aligned} E(\mathbf{C}(s)\overline{\mathbf{C}(t)}) &= \int \exp(isx)\overline{\exp(itx)} F(dx) - \int \exp(isx) F(dx) \int \overline{\exp(itx)} F(dx) \\ &= \tilde{F}(s-t) - \tilde{F}(s)\tilde{F}(-t), \end{aligned} \tag{A.2}$$

$$\begin{aligned} E(\mathbf{C}(s)\mathbf{C}(t)) &= \int \exp(isx)\exp(itx) F(dx) - \int \exp(isx) F(dx) \int \exp(itx) F(dx) \\ &= \tilde{F}(s+t) - \tilde{F}(s)\tilde{F}(t), \end{aligned}$$

and hence  $\mathbf{C}(t)$  has variance  $\sigma^2(t) = E(|\mathbf{C}(t)|^2) = 1 - |\tilde{F}(t)|^2$  and relation  $\rho(t) = E(\mathbf{C}(t)^2) = \tilde{F}(2t) - \tilde{F}(t)^2$ . In particular, if  $t_n \rightarrow \infty$ , then

$$\mathbf{C}(t_n) \rightsquigarrow \mathcal{CN}(0, 1, 0) \quad \text{as } n \rightarrow \infty. \tag{A.3}$$

**Lemma 12.** Let  $\mathbf{B}_n$  be a sequence of standard Brownian bridges, not necessarily independent of  $h_n$ , and define  $\mathbf{C}_n(y) = \int \exp(iyx) \mathbf{B}_n(F(dx))$ . If (4.7) and (4.9) hold, then  $|\mathbf{C}_n(h_n) - \mathbf{C}_n(t_n)| = o_P(1)$ .

*Proof.* For any  $\varepsilon, \delta > 0$ , defining  $I_{n,\delta} = [t_n - \delta, t_n + \delta]$ ,

$$P(|\mathbf{C}_n(h_n) - \mathbf{C}_n(t_n)| > \varepsilon) \leq P\left(\sup_{t \in I_{n,\delta}} |\mathbf{C}_n(t) - \mathbf{C}_n(t_n)| > \varepsilon\right) + P(|h_n - t_n| > \delta).$$

The right probability tends to zero due to (4.9). The left probability can be made arbitrarily small by the choice of  $\delta$ . In fact, by the maximal inequality as given in Lemma 2.1 in Talagrand (1996), there exists a finite constant  $K$  such that, for all  $x > 0$ ,

$$P\left(\sup_{t \in I_{n,\delta}} |\mathbf{C}_n(t) - \mathbf{C}_n(t_n)| > Kx \int_0^\infty \sqrt{\log N(I_{n,\delta}, \eta)} d\eta\right) \leq \exp(-x^2),$$

where  $N(T, \eta)$  denotes the smallest number of open balls of radius  $\eta$  with respect to the distance  $d(s, t) = [E|\mathbf{C}_n(s) - \mathbf{C}_n(t)|^2]^{1/2}$  (which does not depend on  $n$ ) that are necessary in order to cover an index set  $T \subset \mathbb{R}$ . Hence, it remains to show that the entropy integral  $\int_0^\infty \sqrt{\log N(I_{n,\delta}, \eta)} d\eta$  is finite and can be made arbitrarily small by the choice of  $\delta$ .

As already mentioned, the process  $\mathbf{C}_n$  is zero mean complex Gaussian, and with (A.2) we find that

$$\begin{aligned} E|\mathbf{C}_n(s) - \mathbf{C}_n(t)|^2 &= E((\mathbf{C}_n(s) - \mathbf{C}_n(t))\overline{(\mathbf{C}_n(s) - \mathbf{C}_n(t))}) \\ &= 2 - \tilde{F}(s-t) - \tilde{F}(t-s) - |\tilde{F}(s) - \tilde{F}(t)|^2 \\ &\lesssim 1 - \operatorname{Re} \tilde{F}(|s-t|). \end{aligned}$$

By (4.7), there exists  $\gamma > 0$  such that  $y^\gamma F(-y) = O(1)$  and  $y^\gamma(1 - F(y)) = O(1)$  as  $y \rightarrow \infty$ . Without loss of generality, it can be assumed that  $\gamma < 2$ , so that from Theorem 11.3.2 in Kawata (1972) it follows that  $1 - \operatorname{Re} \tilde{F}(t) = O(t^\gamma)$  as  $t \downarrow 0$ . Conclude that

$$d(s, t) = O(|s - t|^{\gamma/2}) \quad \text{as } |s - t| \rightarrow 0.$$

As a result, there exists an absolute constant  $C$  such that  $d(s, t) \leq C|s - t|^{\gamma/2}$  for sufficiently small  $\delta$  and  $s, t \in I_{n,\delta}$ , so that each  $\eta^{2/\gamma}/C$ -cover of  $I_{n,\delta}$  with respect to the absolute value distance is a valid  $\eta$ -cover with respect to  $d$ , yielding

$$N(I_{n,\delta}, \eta) \leq 4C\delta\eta^{-2/\gamma}.$$

This and noting that  $N(I_{n,\delta}, \eta) = 1$  whenever  $\eta \geq C(2\delta)^{\gamma/2} \geq \operatorname{diam}(I_{n,\delta})$  gives

$$\begin{aligned} \int_0^\infty \sqrt{\log N(I_{n,\delta}, \eta)} \, d\eta &\leq \int_0^{C(2\delta)^{\gamma/2}} \sqrt{\log 4C\delta\eta^{-2/\gamma}} \, d\eta \\ &= C(2\delta)^{\gamma/2} \int_0^1 \sqrt{\log 2C^{1-2/\gamma}x^{-2/\gamma}} \, dx \end{aligned}$$

which is  $O(\delta^{\gamma/2})$  as  $\delta \rightarrow 0$ .  $\square$

*Proof of Theorem 9.* For all  $\varepsilon > 0$  and  $T_n \rightarrow \infty$ , with  $\mathbf{C}_{1,n}$  as given in (A.1),

$$\mathbb{P}(|\tilde{\mathbf{F}}_n(h_n) - \mathbf{C}_{1,l_n}(h_n)| > \varepsilon) \leq \mathbb{P}\left(\sup_{0 \leq y \leq T_n} |\tilde{\mathbf{F}}_n(y) - \mathbf{C}_{1,l_n}(y)| > \varepsilon\right) + \mathbb{P}(h_n > T_n),$$

so that, choosing  $T_n = 2t_n$  which is  $o(n^{\gamma/(2\gamma+4)}(\log n)^{-(\gamma+1)/(\gamma+2)})$  by (4.8), from Lemma 11 and (4.9) it follows that  $\tilde{\mathbf{F}}_n(h_n) = \mathbf{C}_{1,l_n}(h_n) + o_P(1)$ . With this, Lemma 12 immediately shows that  $\tilde{\mathbf{F}}_n(h_n) = \mathbf{C}_{1,l_n}(t_n) + o_P(1)$ , and all the same we find that  $\tilde{\mathbf{G}}_n(h_n) = \mathbf{C}_{2,m_n}(t_n) + o_P(1)$ . Hence, in view of (4.3) and since  $\bar{\eta}_n \rightarrow_p \eta$ ,

$$\sqrt{s_n}(\eta_n - \bar{\eta}_n) = \frac{\sqrt{s_n/m_n}}{\tilde{G}_n(h_n)} \left( \sqrt{\tau} \mathbf{C}_{1,l_n}(t_n) - \eta \mathbf{C}_{2,m_n}(t_n) \right) + o_P(1).$$

Below we show that

$$\frac{\sqrt{s_n/m_n}}{\tilde{G}_n(h_n)} = (1 + o_P(1)) \frac{|\tilde{G}(t_n)|}{\tilde{G}(t_n)}, \quad (\text{A.4})$$

so that defining  $z_n = |\tilde{G}(t_n)|/\tilde{G}(t_n)$ ,

$$\sqrt{s_n}(\eta_n - \bar{\eta}_n) = \frac{z_n}{(1 + o_P(1))} \left( \sqrt{\tau} \mathbf{C}_{1,l_n}(t_n) - \eta \mathbf{C}_{2,m_n}(t_n) \right) + o_P(1). \quad (\text{A.5})$$

The conclusion follows by using (A.3), the fact that the  $\mathcal{CN}(0, 1, 0)$ -distribution is invariant under multiplication with the complex non-random numbers  $z_n$  with  $|z_n| = 1$ , and the independence of  $\mathbf{C}_{1,n}$  and  $\mathbf{C}_{2,n}$ .

To show (A.4), from (A.3) it follows that  $\tilde{G}_n(h_n) = O_P(1)$ , which implies  $|\tilde{G}_n(h_n) - \tilde{G}(h_n)| = o_P(\sqrt{s_n/m_n})$ . By (4.10) this gives

$$|\tilde{G}_n(h_n) - \tilde{G}(t_n)| = o_P(\sqrt{s_n/m_n}).$$

Applying (4.5) we therefore have

$$\begin{aligned} \frac{\tilde{G}(t_n)}{\tilde{G}_n(h_n)} &= \frac{\tilde{G}_n(h_n) + o_P(\sqrt{s_n/m_n})}{\tilde{G}_n(h_n)} = 1 + \frac{|\tilde{G}_n(h_n)|}{\tilde{G}_n(h_n)} \frac{o_P(\sqrt{s_n/m_n})}{\sqrt{s_n/m_n}(1 + o_P(1))} \\ &= 1 + o_P(1). \end{aligned}$$

Since  $||\tilde{G}_n(h_n)| - |\tilde{G}(t_n)|| \leq |\tilde{G}_n(h_n) - \tilde{G}(t_n)|$  by triangle inequality, we further have  $|\tilde{G}(t_n)| = \sqrt{s_n/m_n}(1 + o_P(1))$ , and thus

$$\frac{\sqrt{s_n/m_n}}{\tilde{G}_n(h_n)} = \frac{\tilde{G}(t_n)}{\tilde{G}_n(h_n)} \frac{\sqrt{s_n/m_n}}{\tilde{G}(t_n)} = (1 + o_P(1)) \frac{|\tilde{G}(t_n)|}{\tilde{G}(t_n)}.$$

□

The proof of Theorem 10 is given in Hohmann and Holzmann (2013).

### A.2. Proofs of Section 2

*Proof of Theorem 3.* From C1, for  $i = 0, 1$

$$\frac{F(y|B_i)}{(1 - \pi(B_i))F_0(y)} = 1 + \frac{\pi(B_i)F_1(y)}{(1 - \pi(B_i))F_0(y)} \rightarrow 1, \quad y \rightarrow -\infty,$$

and similarly under C2 and C3

$$\begin{aligned} 1 - F(y|B_i) &\sim \pi(B_i)(1 - F_1(y)) && \text{as } y \rightarrow +\infty, \\ \tilde{F}(y|B_i) &\sim \pi(B_i)\tilde{F}_1(y) && \text{as } y \rightarrow +\infty, \end{aligned}$$

Therefore, as  $y \rightarrow -\infty$ ,

$$\frac{F(y|B_1)}{F(y|B_0)} \sim \frac{(1 - \pi(B_1))F_0(y)}{(1 - \pi(B_1))F_0(y)} = \frac{1 - \pi(B_1)}{1 - \pi(B_1)} = \zeta,$$

which is (2.5), and (2.6) follows similarly. Then, given the identification of  $\zeta$  and  $\xi$ , one determines the component cdfs  $F_0$  and  $F_1$  by applying (2.3), and the mixture weight  $\pi$  by using (2.4).

Now let  $G(y|B) = (1 - \mu(B))G_0(y) + \mu(B)G_1(y)$  be another mixture in  $\mathcal{M}_i$  such that  $G = F$ . Then, by C1,

$$\begin{aligned} \zeta &= \lim_{y \rightarrow -\infty} \frac{F(y|B_1)}{F(y|B_0)} = \lim_{y \rightarrow -\infty} \frac{G(y|B_1)}{G(y|B_0)} \\ &= \lim_{y \rightarrow -\infty} \frac{1 + \mu(B_1)(G_1(y)/G_0(y) - 1)}{1 + \mu(B_0)(G_1(y)/G_0(y) - 1)} = \frac{1 - \mu(B_1)}{1 - \mu(B_0)}, \end{aligned}$$

from which we conclude that  $\mu(B_0) < 1$  and  $\mu(B_0) \neq \mu(B_1)$  since  $\zeta \neq 1$ . Similarly, we obtain  $\mu(B_0) > 0$  by C2 respectively C3. Thus, (2.3) applies to  $G$  as well, and since  $F = G$  and the values for  $\xi$  and  $\zeta$  coincide, (2.3) implies that  $F_i = G_i$ ,  $i = 0, 1$ . In particular,  $G_0(y_0) \neq G_1(y_0)$ , and by (2.4) it follows that  $\pi = \mu$ .  $\square$

### A.3. Proofs of Section 3

*Proof of Proposition 4.* Using our previous notation, set  $\bar{\zeta}_n = F(L_n|B_1)/F(L_n|B_0)$  and  $\bar{\xi}_n = \text{Re}(\bar{F}(R_n|B_1)/\bar{F}(R_n|B_0))$ . Let  $\mathcal{Z} = \sigma(\mathbf{1}_{\{Z_k \in B_j\}}, k \in \mathbb{N}, j = 0, 1)$  be the  $\sigma$ -field generated by the indicator variables  $\mathbf{1}_{\{Z_k \in B_j\}}$ . By A2, the distribution of  $Y_1, \dots, Y_n$  conditional on  $\mathcal{Z}$  is given by

$$\begin{aligned} & \mathbb{P}(Y_1 \leq y_1, \dots, Y_n \leq y_n | \mathcal{Z}) \\ &= \prod_{\substack{1 \leq i \leq n \\ Z_i \in B_0}} F(y_i | B_0) \prod_{\substack{1 \leq i \leq n \\ Z_i \in B_1}} F(y_i | B_1) \prod_{\substack{1 \leq i \leq n \\ Z_i \notin B_0 \cup B_1}} F(y_i | (B_0 \cup B_1)^c), \end{aligned}$$

and thus, evaluating the latter at a fixed  $\omega$ ,

$$\begin{aligned} & \mathbb{P}(Y_i^* \leq y_i^*, Y_j^{**} \leq y_j^{**}, i = 1, \dots, m_n, j = 1, \dots, l_n | \mathcal{Z})(\omega) \\ &= \prod_{i=1}^{m_n(\omega)} F(y_i^* | B_0) \prod_{j=1}^{l_n(\omega)} F(y_j^{**} | B_1). \end{aligned}$$

Hence, conditional on  $\mathcal{Z}$ ,  $Y_1^*, \dots, Y_{m_n}^*$  and  $Y_1^{**}, \dots, Y_{l_n}^{**}$  are independent i.i.d. samples from  $F(\cdot | B_0)$  and  $F(\cdot | B_1)$ , resp., where for almost all  $\omega$  it holds that

$$m_n(\omega)/l_n(\omega) \rightarrow p_0/p_1.$$

Defining  $\tau = p_0/p_1$ , it therefore follows from Theorem 10 and (3.2) that

$$\sqrt{r_n} \begin{pmatrix} \zeta_n - \zeta \\ \xi_n - \xi \end{pmatrix} | \mathcal{Z} \rightsquigarrow \mathcal{N} \left( \begin{pmatrix} \beta_\zeta \\ \beta_\xi \end{pmatrix}, \begin{pmatrix} \tau\zeta + \zeta^2 & 0 \\ 0 & (\tau + \xi^2)/2 \end{pmatrix} \right)$$

with probability one. Finally, since the weak limit does actually not depend on the specific realization of the sequence  $\mathbf{1}_{\{Z_k \in B_j\}}$ , the result follows.  $\square$

*Proof of Theorem 5.* We only consider the process  $\mathbf{F}_n = \sqrt{r_n}(\hat{F}_1 - F_1)$ . Let  $K(x) = (1-x)^{-1}$ ,  $x \neq 1$ . Since

$$\begin{aligned} \hat{F}_1(y) - F_1(y) &= F_n(y|B_0) - F(y|B_0) + (K(\zeta_n) - K(\zeta)) D(F_n, y) \\ &+ (D(F_n, y) - D(F, y)) K(\zeta), \end{aligned} \tag{A.6}$$

where  $F_n(y|B_i) - F(y|B_i) = O_P(n^{-1/2})$  and  $r_n/n \rightarrow 0$ , with  $K_n = \sqrt{r_n}(K(\zeta_n) - K(\zeta))$  we find that

$$\mathbf{F}_n(y) = D(F_n, y)K_n + o_P(1).$$

The function  $K$  being differentiable at  $\zeta$  with derivative  $(1 - \zeta)^{-2}$ , the Delta method and (3.3) yield

$$K_n \rightsquigarrow \mathcal{N}((1 - \zeta)^{-2}\beta_\zeta, (1 - \zeta)^{-4}\sigma_\zeta^2).$$

Since  $D(F_n, y) \rightarrow_{a.s.} D(F, y)$ , we therefore find that, for any  $y_1, y_2 \in \mathbb{R}$ , defining  $D(y_1, y_2) = (D(F, y_1), D(F, y_2))'$ ,

$$(\mathbf{F}_n(y_1), \mathbf{F}_n(y_2))' \rightsquigarrow \mathcal{N}(D(y_1, y_2)(1 - \zeta)^{-2}\beta_\zeta, D(y_1, y_2)D(y_1, y_2)'(1 - \zeta)^{-4}\sigma_\zeta^2),$$

which shows the weak convergence of the finite dimensional distributions of  $\mathbf{F}_n$  to those of the Gaussian process  $\mathbf{G}_\zeta$ .

To conclude weak convergence in  $\ell^\infty(\mathbb{R})$ , it remains to show that  $\mathbf{F}_n$  is asymptotically tight, a sufficient condition for which is to show that for any  $\varepsilon, \eta > 0$  we can find a finite partition  $\{I_1, \dots, I_d\}$  of  $\mathbb{R}$  such that

$$\limsup_{n \rightarrow \infty} P\left(\max_{k=1, \dots, d} \sup_{s, t \in I_k} |\mathbf{F}_n(t) - \mathbf{F}_n(s)| > \varepsilon\right) < \eta \tag{A.7}$$

holds, cf. van der Vaart and Wellner (2000, Theorem 1.5.6). Similar to (A.6), write

$$\mathbf{F}_n(s) - \mathbf{F}_n(t) = (F_n(s|B_1) - F_n(t|B_1) + F_n(t|B_0) - F_n(s|B_0)) K_n + o_P(1),$$

where the remainder is uniform in  $s, t$ . By the uniform convergence of  $F_n$ ,

$$\sup_{s, t \in \mathbb{R}} |F_n(s|B_i) - F_n(t|B_i)| \leq \sup_{s, t \in \mathbb{R}} |F(s|B_i) - F(t|B_i)| + o_P(1).$$

For all  $\gamma \in (0, 1)$  one can find a finite partition  $\{I_k\}$  of  $\mathbb{R}$  such that

$$\sup_{s, t \in I_k} (|F(s|B_1) - F(t|B_1)| + |F(s|B_0) - F(t|B_0)|) < \gamma,$$

implying

$$\max_k \sup_{s, t \in I_k} |\mathbf{F}_n(s) - \mathbf{F}_n(t)| \leq \gamma |K_n| + o_P(1).$$

By asymptotic normality of  $K_n$ , we find  $C > 0$  such that

$$\limsup_n P(|K_n| > C) < \eta.$$

Now choosing  $\gamma = C/\varepsilon$  gives (A.7) and hence the asymptotic tightness of  $\mathbf{F}_n$ .  $\square$

*Theorem 6* is proved by using the Delta method.

### References

BORDES, L., MOTTELET, S. and VANDEKERKHOVE, P. (2006). Semiparametric Estimation of a Two-component Mixture Model. *Ann. Statist.* **34** 1204-1232. [MR2278356](#)

- BORDES, L. and VANDEKERKHOVE, P. (2010). Semiparametric two-component mixture model when a component is known: an asymptotically normal estimator. *Mathematical Methods of Statistics* **19** 22-41. [MR2682853](#)
- CSÖRGŐ, S. (1981). Limit behaviour of the Empirical Characteristic Function. *Ann. Probab.* **9** 130-144. [MR0606802](#)
- HALL, P. and ZHOU, X. (2003). Nonparametric Estimation of Component Distributions in a Multivariate Mixture. *Ann. Statist.* **31** 201-224. [MR1962504](#)
- HENRY, M., KITAMURA, Y. and SALANIÉ, B. (2010). Identifying Finite Mixtures in Econometric Models. *Cowles Foundation Discussion Paper Series* **1767**.
- HOHMANN, D. and HOLZMANN, H. (2012). Semiparametric location mixtures with distinct components. *Statistics*.
- HOHMANN, D. and HOLZMANN, H. (2013). Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation. *Technical Report, Marburg University*.
- HUNTER, D., WANG, S. and HETTMANSPERGER, T. (2007). Inference for Mixtures of symmetric distribution. *Ann. Statist.* **35** 224-251. [MR2332275](#)
- KASAHARA, H. and SHIMOTSU, K. (2009). Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices. *Econometrica* **vol. 77** 135-175. [MR2477846](#)
- KAWATA, T. (1972). *Fourier Analysis in Probability Theory. Probability and Mathematical Statistics*. Academic Press, New York. [MR0464353](#)
- KITAMURA, Y. (2004). Nonparametric Identifiability of Finite Mixtures Unpublished manuscript, Yale University, Department of Economics.
- MC LACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models. Wiley series in probability and statistics*. John Wiley & Sons, New York. [MR1789474](#)
- TALAGRAND, M. (1996). Lower Classes for Fractional Brownian Motion. *J. Theor. Probab.* **9** 191-213. [MR1371076](#)
- VAN DER VAART, A. and WELLNER, J. (2000). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York.
- VANDEKERKHOVE, P. (2010). Estimation of a semiparametric contaminated regression model. *Preprint*.