Estimators based on ω-dependent generalized weighted Cramér-von Mises distances under censoring - with applications to mixture models

Eric Beutner
Department of Quantitative Economics, Maastricht University

Laurent Bordes
Laboratoire de Mathématiques et de leurs Applications, UMR CNRS 5142
Université de Pau et des Pays des l'Adour

December 7, 2008

Abstract

Estimators based on ω-dependent generalized weighted Cramér-von Mises distances are defined for data that are subject to a possible right censorship. The distance between the data, summarized by the Kaplan-Meier estimator, and the target model is allowed to depend on the sample size and, for example, on the number of censored items. It is shown that the estimators are consistent and asymptotically multivariate normal for every p dimensional parametric family fulfilling some mild regularity conditions. The results are applied to finite mixtures. Simulation results for finite mixtures indicate that the estimators are useful for moderate sample sizes. Furthermore, the simulation results reveal the usefulness of sample size dependent and censoring sensitive distance functions for moderate sample sizes. Moreover, the estimators for the mixing proportion seems to be fairly robust against a 'symmetric' contamination model even when censoring is present.

Keywords: Minimum distance estimators, censored data, finite mixture models, Generalized weighted Cramér-von Mises estimators

1 Introduction

One of the basic problems in statistics is to fit a parametric family $\mathcal{F} = \{F_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$ to data. One approach to achieve this for uncensored data is based on goodness-of-fit statistics leading to minimum distance estimators which were first discussed in detail by Wolfowitz (1957). A minimum distance estimator of θ is a value, $\hat{\theta}$, which minimizes the distance between the data, summarized by the empirical distribution function, and the model $\mathcal{F} = \{F_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$. Minimum distance estimators have been studied for several goodness-of-fit measures such as weighted Cramér-von Mises, Kolmogorov-Smirnov or Hellinger distances; see, for example, Beran (1977), Beran (1984), Hettmansperger et al. (1994), Öztürk & Hettmansperger (1997), and Parr & Schucany (1980).
In the context of finite mixture models and uncensored data, several authors investigated the use of minimum distance estimators as an alternative to maximum likelihood estimation for the estimation of the mixing proportions; see, for example, Choi & Bulgren (1968), Cutler & Cordero-Brara (1996), Pardo (1997), Woodward et al. (1984), and Woodward et al. (1995). For general accounts on finite mixture models one may refer to Titterington et al. (1985) and McLachlan & Peel (2000).

Under random censorship the maximum likelihood method is well adapted to classical parametric families including Weibull, log-normal, etc. A general maximum likelihood theory under right censoring based on counting processes can be found in Andersen et al. (1993). However, this method fails to be adapted to complicated parametric families including, for example, finite mixtures of parametric distributions. Many alternative methods under random censorship are based on minimum distance estimators. Yang (1991) considers minimum Hellinger distance estimators while Ying (1992) introduces a new Hellinger-type minimum distance estimator based on hazard functions incorporating random censorship in a natural way. However, although these methods are general and can be adapted to any parametric family regular enough, they require to specify some bandwidth and to approximate the integrals to be minimized. This last drawback is, in the complete i.i.d case, also observed for generalized weighted Cramér-von Mises minimum distance estimators (cf. Öztürk & Hettmansperger (1997)). Recently, Ferland et al. (2003) introduced a Cramér-von Mises type minimum distance estimator for the estimation of the weights in a finite mixture model under random censorship and studied its properties through a simulation study. Their estimator is a direct extension of the Cramér-von Mises type estimator used by Choi & Bulgren (1968) and does not require any approximation of the functional to be minimized.

In this article, we propose a class of estimators based on ω-dependent generalized weighted Cramér-von Mises distances under random censorship which do not require an approximation method to compute the functionals to be minimized. We examine the asymptotic properties of these estimators and present results from a simulation study on the estimation of the mixing proportion in a finite mixture model. The material is organized as follows. In Section 2 we provide background material on minimum distance estimators and motivate the class of estimators based on ω-dependent generalized weighted Cramér-von Mises distances considered hereafter. In Section 3 we present strong consistency and asymptotic normality results for our class of estimators based on ω-dependent generalized weighted Cramér-von Mises distances under random censorship and discuss the conditions imposed to obtain these results. The proofs of the main theorems are also presented in Section 3. Moreover, we derive the influence curves for our class of estimators based on ω-dependent generalized weighted Cramér-von Mises distances under random censorship. Finally, in Section 4 we present results from a simulation study to compare the moderate sample size behavior of different estimators contained in our class of estimators based on ω-dependent generalized weighted Cramér-von Mises distances. It will turn that good results are, in particular, obtained by estimators based on sample size dependent and censoring sensitive generalized weighted Cramér-von Mises distances. The proofs of all results used to show the main theorems are given in the Appendix.

2 The Cramér-von Mises type minimum distance estimators

To motivate our choice of Cramér-von Mises type minimum distance estimators under random censorship let us briefly recall the complete i.i.d case. Let $X_1, \ldots, X_n$ be a complete i.i.d sample with unknown distribution function, and let $\mathcal{F} = \{F_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$ be a parametric family. The Cramér-von Mises distance D^{CvM} between distribution functions G_1 and G_2 is given by

$$D^{CvM}(G_1, G_2) = \int (G_1(x) - G_2(x))^2 dG_2(x).$$

A minimum Cramér-von Mises distance estimator of θ is then any $\hat{\theta}_n$ such that

$$D^{CvM}(F_n, F_{\hat{\theta}_n}) \leq D^{CvM}(F_n, F_\theta) + 1/n, \, \forall \theta \in \Theta,$$

where F_n is the empirical distribution function based on $X_1, \ldots, X_n$ (cf. Woodward et al. (1984)). Please recall that

$$D^{CvM}(F_n, F_\theta) = 1/(12n) + \sum_{i=1}^{n} \left(F_\theta(X_{(i)}) - i/n + 1/(2n)\right)^2, \tag{1}$$

by the probability integral transform. Here $X_{(i)}$ denotes the ith order statistic in the sample $X_1, \ldots, X_n$. In the context of estimating the weights in a finite mixture model Choi & Bulgren (1968) proposed to estimate the parameter of interest as the argmin of

$$\sum_{i=1}^{n} \left(F_\theta(X_{(i)}) - i/n\right)^2. \tag{2}$$

MacDonald (1971) (see also Section 4) provides empirical evidence that the small sample bias of the estimator based on (1) is smaller than the small sample bias of the estimator based on (2), although their asymptotic properties are the same. Öztürk & Hettmansperger (1997) introduced generalized weighted Cramér-von Mises distance estimators to control the asymptotic distribution and the robustness. They defined their estimator by

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \int G(F_n(t) - F_\theta(t)) w(t, \theta) \, dt \tag{3}$$

where G is taken from a broad class of distance functions, and G and w fulfill certain regularity conditions.

Let us now turn to the extension of the above approaches to right censorship. Here, we shall mean that, in addition to the i.i.d random variables $X_1, \ldots, X_n$ with distribution function (d.f.) F_{θ_0}, there exist censoring variables $C_1, \ldots, C_n$ with d.f. H independent of each other and independent of $X_1, \ldots, X_n$ such that only

$$(T_i, \Delta_i) \equiv \left(X_i \wedge C_i, I_{\{X_i \le C_i\}}\right), \; i = 1, \ldots, n$$

are observed. Here, and in the following I denotes the indicator function. Throughout, we assume that F_{θ_0} belongs to a parametric family $\mathcal{F} = \{F_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$, which is supposed to be dominated by the Lebesgue measure. Then, to each F_θ in $\mathcal{F}$ we associate its cumulative hazard function Λ_θ. It is well-known that F_{θ_0} can be nonparametrically estimated by the Kaplan-Meier estimator (see Kaplan & Meier (1958)) defined on $[0, \tau]$ by

$$\hat{F}_n(t) = 1 - \prod_{\{i: T_i \le t\}} \left(1 - \frac{\Delta_i}{Y(T_i)}\right),$$

where $Y(s) = \sum_{i=1}^n I_{\{T_i \ge s\}}$.
We define the ω-dependent generalized weighted Cramér-von Mises distance under random censorship by

$$\int_0^\tau G_n\left(\hat{F}_n(t) - F_\theta(t), \omega\right) w_n(t) d\hat{F}_n(t), \; \theta \in \Theta, \tag{4}$$

where w_n is a weight function, G_n is a distance function, which is allowed to depend on n as well as on ω, and τ, w_n and G_n satisfy Conditions A, B and C below. In the following, we shall drop the integration variable t.

Definition 1. An estimator of θ_0 based on a ω-dependent generalized weighted Cramér-von Mises distance under random censorship is then any $\hat{\theta}_n$ such that

$$\int_0^\tau G_n\left(\hat{F}_n - F_{\hat{\theta}_n}, \omega\right) w_n d\hat{F}_n \le \int_0^\tau G_n\left(\hat{F}_n - F_\theta, \omega\right) w_n d\hat{F}_n, \text{ for all } \theta \in \Theta. \tag{5}$$

Besides from the facts that it accounts for right censoring and that the distance is allowed to depend on ω, one of the main advantages of the above estimation procedure is that, contrary to many minimum distance estimation method, it does not require an approximation method to compute the integrals. Indeed, integration holds with respect to the empirical measure, transforming the integral into a simple sum. Therefore, this integral may be minimized by using standard routines from differential optimization.

Remark 1. To avoid the problem of existence and attainability of a minimum, $\hat{\theta}_n$ can be defined as in equation (2.1) of Woodward et al. (1984)

$$\int_0^\tau G_n\left(\hat{F}_n - F_{\hat{\theta}_n}, \omega\right) w_n d\hat{F}_n \leq \int_0^\tau G_n\left(\hat{F}_n - F_\theta, \omega\right) w_n d\hat{F}_n + \frac{1}{n} \quad \text{for all } \theta \in \Theta.$$

Let us briefly comment on the class of distances given by (4). Firstly, allowing the distance function G_n to depend on n has several advantages. For example, if there is no censoring, we obtain, by choosing $G_n(\cdot, \omega) = (\cdot + 1/(2n))^2$ and $w_n \equiv 1$, that minimizing (4) is equivalent to minimizing the Cramér-von Mises distance (cf. (1)). If we did not allow the distance to depend on n, it would be impossible to include the distance obtained from the Cramér-von Mises goodness-of-fit statistic since, here, we are integrating with respect to the measure induced by the Kaplan-Meier estimator. Moreover, we can define censoring sensitive distance functions if we allow G_n to depend on ω. For example taking $G_n(\cdot, \omega) = (\cdot + 1/(2\sum_{i=1}^n I_{\{X_i \leq C_i, X_i \leq \tau\}}))^2$ and $w_n \equiv 1$, the small sample behavior can be considerable improved compared to $G_n(\cdot, \omega) = (\cdot)^2$ and $w_n \equiv 1$ (see Section 4), although their large sample behavior is the same. Secondly, by introducing a weight function w_n the tail probabilities can be emphasized or de-emphasized. As, under right censoring, we usually have less observations in the right tail this is desirable.

3 Asymptotic results and robustness

This section is devoted to studying the properties of the estimator defined by (5). We first discuss the conditions imposed to derive consistency and asymptotic normality (Subsection 3.1), then we give the proofs (Subsection 3.2), and finally we derive the influence curves (Subsection 3.3).

3.1 Discussion of the conditions

Let us introduce the following conditions:

A. Let $\tau > 0$ be a real number such that $\tau < \sup\{t > 0; (1 - F_{\theta_0}(t))(1 - H(t)) > 0\}$.

B. Let w_n be a sequence of random nonnegative functions on $[0, \tau]$ satisfying

$$\sup_{[0,\tau]} |w_n - w_0| \xrightarrow{P} 0,$$

where w_0 is a bounded deterministic function on $[0, \tau]$.

C. Let the set Υ consists of all functions G such that

(i) $G : [a, b] \to \mathbb{R}^+$, $a \leq -1, 1 \leq b$, is nonnegative,

(ii) the restriction of G to the interval $[-1, 1]$ is twice continuously differentiable,

(iii) $G(0) = G'(0) = 0$ and $G''(0) > 0$.

We then assume that $G_n : [a, b] \times \Omega_n \to \mathbb{R}^+$ is such that $G_n(\cdot, \omega) = G\left(\cdot + o_p(1/\sqrt{n})\right)$ for some $G \in \Upsilon$. Here Ω_n is the sample space.

D. If $\theta_n \in \Theta \subset \mathbb{R}^p$, $n \in \mathbb{N}^*$ then

$$\lim_{n\to+\infty} \int_0^\tau G\left(F_{\theta_0} - F_{\theta_n}\right) w_0 dF_{\theta_0} = 0$$

implies $\lim_{n\to+\infty} \theta_n = \theta_0$.

E. There exists a measurable function $\eta = (\eta_1, \ldots, \eta_p)^t : (0, q) \equiv (0, F_{\theta_0}(\tau)) \to \mathbb{R}^p$ such that $\Sigma(\tau) = \int_0^\tau \eta(F_{\theta_0}(s))\eta^t(F_{\theta_0}(s))w_0(s)dF_{\theta_0}(s)$ is positive definite, and

$$\sup_{0\leq s\leq\tau} |F_\theta(s) - F_{\theta_0}(s) - (\theta - \theta_0)^t \eta \circ F_{\theta_0}(s)| = o(\|\theta - \theta_0\|)$$

as $\|\theta - \theta_0\| \to 0$.

Let us briefly discuss the assumptions imposed. Condition A together with the assumptions on the sample ensures that the standard results for the Kaplan-Meier estimator hold true on the interval $[0, \tau]$.

The sequence w_n can be equal to $w(\hat{F}_n)$ with w Lipschitz on $[0, 1]$. In this case we have $w_0 = w(F_{\theta_0})$ and Condition B is fulfilled since $\hat{F}_n$ converges to F_{θ_0} uniformly on $[0, \tau]$. Taking for w the functions $w_1(x) = x^p$, $w_2(x) = (1 - x)^p$, or $w_3(x) = x^p + (1-x)^p$ with $p > 1$ leads to the empirical versions of the three weight functions considered in Öztürk & Hettmansperger (1997). However, many other choices adapted to right censoring are also possible; for example, we can take $w_n = n/Y$ which satisfies Condition B too.

The sequence of functions $G_n(\cdot, \omega) = G(\cdot + 1/(2n))$, where $G(\cdot) = (\cdot)^2$, fulfills Condition C so that our model contains the estimator defined by (1) as well as the one defined by (2) and therefore allows for a unique treatment. Furthermore, it is seen that under Condition A we have that $(1/n)\sum_{i=1}^n I_{\{X_i \le C_i, X_i \le \tau\}}$ converges to $\int_0^\tau (1-H)dF_{\theta_0} \ge (1-H(\tau))F_{\theta_0}(\tau) > 0$ (assuming that $F_{\theta_0}(\tau) > 0$). Therefore, $G_n(\cdot, \omega) = \left(\cdot + 1/\left(2\sum_{i=1}^n I_{\{X_i \le C_i, X_i \le \tau\}}\right)\right)^2$, also fulfills Condition C. The conditions imposed by C on the function G are stronger than the condition imposed by Öztürk & Hettmansperger (1997) on their distance functions. Our Condition C implies that G and its two first derivatives are bounded on $[-1, 1]$. If this condition is false, taking for example $G(x) = x^2/(x+1)$ as in Öztürk & Hettmansperger (1997), existence and asymptotics may be obtained by reducing the parameter space Θ to a compact subset (see Öztürk & Hettmansperger (1997)).

Condition D is similar to Condition C of Woodward et al. (1984), and if $w_0 = 1$ they are equal. As discussed by these authors, if it is not satisfied, the search for any consistent estimator seems hopeless (cf. Woodward et al. (1984, Note 2)). Notice that Condition D implies that the model restricted to the interval $[0, \tau]$ is identifiable. To see this suppose that $P_{\theta|[0,\tau]} = P_{\theta_0|[0,\tau]}$ (where $P_{\theta|[0,\tau]}$ is the restriction of P_θ to $[0, \tau]$), then taking $\theta_n = \theta$ for all n we have

$$\lim_{n \to +\infty} \int_0^\tau G\left(F_{\theta_0} - F_{\theta_n}\right) w_0 dF_{\theta_0} = 0,$$

which by Condition D implies $\theta = \lim_{n\to+\infty} \theta_n = \theta_0$. Sufficient conditions to obtain D will be discussed below.

Condition E is very similar to Condition D in Woodward et al. (1984) but we assume that η is measurable which insures that it belongs to any $L^r(0, q)$ for $1 \le r \le \infty$. As discussed by these authors, Condition E allows a first order Taylor expansion of $\theta \mapsto F_\theta(x)$ around θ_0 uniformly in x.

Let us now turn to some sufficient conditions to obtain Condition D. First, let us write

$$d(\theta) = \int_0^\tau G(F_\theta - F_{\theta_0}) w_0 dF_{\theta_0}.$$

Obviously, $d(\theta_0) = 0$ and if $\theta \mapsto d(\theta)$ is twice continuously differentiable on Θ, then we have

$$\begin{aligned} d(\theta_n) &= d(\theta_0) + (\theta_n - \theta_0)^t \dot{d}(\theta_0) + \frac{1}{2}(\theta_n - \theta_0)^t \ddot{d}(\theta_n^*)(\theta_n - \theta_0) \\ &= \frac{1}{2}(\theta_n - \theta_0)^t \ddot{d}(\theta_n^*)(\theta_n - \theta_0), \qquad (6) \end{aligned}$$

where $\dot{d}$ and $\ddot{d}$ denote the first and second order derivative of d with respect to θ, respectively, and θ_n^* belongs to the line segment connecting θ_n and θ_0. Let $\sigma(A)$ be the minimum eigenvalue of the $p \times p$ matrix A.

Sufficient condition (I) to obtain D:

$$\inf_{\theta \in \Theta} \sigma\left(\ddot{d}(\theta)\right) > 0.$$

If (I) holds then by (6) it is clear that Condition D holds.

For example if $F_\theta(t) = 1 - \exp(-\theta t)$ for $t > 0$ we obtain choosing $G(x) = x^2$, $\tau = +\infty$, and $w_0 \equiv 1$

$$\begin{aligned} d(\theta) &= \theta_0 \left(\frac{1}{2\theta + \theta_0} - \frac{2}{\theta + 2\theta_0} + \frac{1}{3\theta_0} \right) \quad \text{and} \\ \ddot{d}(\theta) &= 4\theta_0 \left(\frac{2}{(2\theta + \theta_0)^3} - \frac{1}{(\theta + 2\theta_0)^3} \right). \end{aligned}$$

Here we see that if $\Theta = (0, +\infty)$ we can have $\ddot{d}(\theta) < 0$ for large values of θ. In fact, it is generally difficult to obtain *Sufficient condition (I)* on the whole parameter space Θ, but if we reduce the parameter space Θ to a subset Θ_1 where the minimum eigenvalue of $\ddot{d}$ is bounded from above by a positive constant then Condition D is satisfied. Such a reduction of the parameter space to a (finite) subset is generally possible when d is twice continuously differentiable with $\ddot{d}(\theta_0)$ positive definite. In this case we obtain a local version of Condition D and it is easy to check that

$$\ddot{d}(\theta_0) = G''(0) \int_0^\tau \dot{F}_{\theta_0} \dot{F}_{\theta_0}^t w_0 dF_{\theta_0}.$$

In the exponential example we obtain:

$$\ddot{d}(\theta_0) = 4G''(0)/(27\theta_0^3) > 0.$$

Sufficient conditions (II) to obtain D:

(i) There exist $\varepsilon > 0$ and $\eta > 0$ such that $B(\theta_0, \eta) \equiv \{\theta \in \mathbb{R}^p, \|\theta - \theta_0\| \leq \eta\} \subset \Theta$, and if $\theta \in \Theta \backslash B(\theta_0, \eta)$ then $d(\theta) \geq \varepsilon$,

(ii) d is continuous on $B(\theta_0, \eta)$,

(iii) $d \geq 0$ and $d(\theta) = 0$ if and only if $\theta = \theta_0$.

Functions satisfying (iii) are generally called *contrast functions*. Now using (ii) and (iii) we have that for any $\alpha > 0$ there exists $\beta > 0$ such that if $\theta \in B(\theta_0, \eta) \backslash B(\theta_0, \alpha)$ then $d(\theta) \geq \beta$. This with (i) imply that if $\|\theta - \theta_0\| > \alpha$ we have $d(\theta) \geq \min(\beta, \varepsilon)$. As a consequence if $d(\theta_n) \to 0$ this means that for n large enough θ_n belongs to $B(\theta_0, \alpha)$, and since α is arbitrary this proves that θ_n tends to θ_0.

In the above exponential example with $\Theta = (0, +\infty)$ it is easy to check that (i)–(iii) of the *Sufficient conditions (II)* hold. Thus, Condition D is satisfied in that case. Indeed (ii) and (iii) are obvious and (i) holds because d is nonincreasing on $(0, \theta_0)$ and nondecreasing on $(\theta_0, +\infty)$.

More generally, if F_{θ_0} and the Lebesgue measure are contiguous on $(0, \tau)$, if G is strictly positive on $[-1, 1] \backslash \{0\}$, and if w_0 is strictly positive on $(0, \tau)$, then $d(\theta) = 0$ leads to $F_\theta = F_{\theta_0}$ on $[0, \tau]$. Moreover if τ is sufficiently large and if $\mathcal{F}$ is an identifiable parametric family we obtain (iii) of the *Sufficient conditions (II)* . Actually, $\mathcal{F}$ is identifiable for many parametric families like exponential, Weibull, gamma, lognormal, etc. Identifiability of mixtures of parametric families also holds for many parametric families. Teicher (1963) gave a sufficient condition for a finite mixture to be identifiable and applied it to the normal and gamma families. This result was extended to usual survival distributions, and recently Atienza et al. (2006) gave some new conditions for identifiability of finite mixture distributions and showed that the class of all finite mixture distributions generated by the union of lognormal, gamma and Weibull distributions is identifiable. For references concerning identifiability of finite mixture distributions we refer to the paper of Atienza et al. (2006).

Obtaining property (ii) of *Sufficient conditions (II)* is generally easy by Lebesgue's dominated convergence theorem.

Checking property (i) of *Sufficient conditions (II)* may be more fussy. For example, let us consider the following two-component mixture of exponential distributions:

$$F_\theta(t) = 1 - \lambda \exp(-\alpha t) - (1-\lambda)\exp(-\beta t), \quad t > 0,$$

with $\theta = (\lambda, \alpha, \beta) \in \Theta = (0,1) \times \Delta$ where $\Delta = \{(x,y) \in \mathbb{R}^2; 0 < x < y\}$. In this case $d(\theta)$ can be written in closed form and (i) is fulfilled.

3.2 Consistency and asymptotic normality

In the proof of the consistency we will need the following Lemma a proof of which is given in the appendix.

Lemma 1. *Let G fufill Condition C. Then, we have that the class of functions $\mathcal{Z} = \{G \circ (F_{\theta_0} - F_\theta)w_0; \theta \in \Theta\}$ is $\mathbb{P}$–Glivenko-Cantelli.*

We then have the following result.

Theorem 1. *Any sequence $(\hat\theta_n)_{n\geq 1}$ defined by (5) is consistent if Conditions A–D hold.*

Proof. Notice first that

$$\begin{aligned}\int_0^\tau G_n\left(\hat F_n - F_{\theta_0}, \omega\right) w_n d\hat F_n \quad &\leq \quad \sup_{[0,\tau]} G_n\left(\hat F_n - F_{\theta_0}, \omega\right) \times \sup_{[0,1]} |w_n| \\ &= \sup_{[0,\tau]} G\left(\hat F_n - F_{\theta_0} + o_p(1/\sqrt{n})\right) \times \sup_{[0,1]} |w_n|. \qquad (7)\end{aligned}$$

Moreover, by definition

$$\int_0^\tau G_n\left(\hat F_n - F_{\hat\theta_n}, \omega\right) w_n d\hat F_n \leq \int_0^\tau G_n\left(\hat F_n - F_{\theta_0}, \omega\right) w_n d\hat F_n. \qquad (8)$$

Finally, we have

$$
\begin{aligned}
& \sup_{\theta\in\Theta}\left|\int_0^\tau G_n\left(\hat{F}_n - F_\theta, \omega\right) w_n d\hat{F}_n - \int_0^\tau G\left(F_{\theta_0} - F_\theta\right) w_0 dF_{\theta_0}\right| \\
= \; & \sup_{\theta\in\Theta}\left|\int_0^\tau \left[G\left(\hat{F}_n - F_\theta + o_p(1/\sqrt{n})\right) - G\left(F_{\theta_0} - F_\theta\right)\right] w_n d\hat{F}_n \right. \\
& \left. + \int_0^\tau G(F_{\theta_0} - F_\theta)(w_n - w_0) d\hat{F}_n + \int_0^\tau G\left(F_{\theta_0} - F_\theta\right) w_0 \left(d\hat{F}_n - dF_{\theta_0}\right)\right| \\
\leq \; & \sup_{\theta\in\Theta}\left|\int_0^\tau \left[G\left(\hat{F}_n - F_\theta + o_p(1/\sqrt{n})\right) - G\left(F_{\theta_0} - F_\theta\right)\right] w_n d\hat{F}_n\right| \\
& + \sup_{[-1,1]}|G| \times \sup_{[0,\tau]}|w_n - w_0| + \sup_{\theta\in\Theta}\left|\int_0^\tau G\left(F_{\theta_0} - F_\theta\right) w_0 d\left(\hat{F}_n - F_{\theta_0}\right)\right| \\
\leq \; & \sup_{\theta\in\Theta}\sup_{[0,\tau]}\left|G\left(\hat{F}_n - F_\theta + o_p(1/\sqrt{n})\right) - G\left(F_{\theta_0} - F_\theta\right)\right| \times \sup_{[0,\tau]}|w_n| \\
& + \sup_{[-1,1]}|G| \times \sup_{[0,\tau]}|w_n - w_0| + \sup_{\theta\in\Theta}\left|\int_0^\tau G\left(F_{\theta_0} - F_\theta\right) w_0 d\left(\hat{F}_n - F_{\theta_0}\right)\right|.
\end{aligned}
\tag{9}
$$

Following the proof of Theorem 2.4.1 in van der Vaart & Wellner (1996), the uniform convergence to zero of the third term on the right hand side follows from the fact that the class $\mathcal{Z} = \{(w_0 G \circ z : z = F_{\theta_0} - F_\theta, \theta \in \Theta\}$ is $\mathbb{P}$–Glivenko-Cantelli (cf. Lemma 1). The second term converges to zero in probability by Conditions B and C. The first term can be majorized by

$$
\begin{aligned}
& \sup_{\theta\in\Theta}\sup_{[0,\tau]}\left|G\left(\hat{F}_n - F_\theta + o_p(1/\sqrt{n})\right) - G\left(\hat{F}_n - F_\theta\right)\right| \times \sup_{[0,\tau]}|w_n| \\
& + \sup_{\theta\in\Theta}\sup_{[0,\tau]}\left|G\left(\hat{F}_n - F_\theta\right) - G\left(F_{\theta_0} - F_\theta\right)\right| \times \sup_{[0,\tau]}|w_n|.
\end{aligned}
\tag{10}
$$

Using Condition C the first summand in (10) can, for every θ, be majorized by

$$
\sup_{[-1,1]}|G'| \times |o_P(1/\sqrt{n})| \times \sup_{[0,\tau]}|w_n|.
$$

Using Conditions B and C this turns to zero in probability which proves that the first summand in (10) converges to zero in probability. To obtain the convergence of the second summand in (10) notice that by Condition C we again obtain for every $\theta \in \Theta$

$$
\left|G\left(\hat{F}_n - F_\theta\right) - G\left(F_{\theta_0} - F_\theta\right)\right| \quad \leq \quad \sup_{[-1,1]}|G'| \times \sup_{[0,\tau]}\left|\hat{F}_n - F_{\theta_0}\right|.
$$

From Fleming & Harrington (2005, p. 115) we have

$$\sup_{[0,\tau]} \left| \hat{F}_n - F_{\theta_0} \right| \xrightarrow{P} 0, \tag{11}$$

because under Condition A and the strong law of large numbers we have, for all $t \in [0,\tau]$, $Y(t) \to +\infty$ as $n \to +\infty$. Hence, using the boundedness of G' we obtain

$$\sup_{\theta\in\Theta} \sup_{[0,\tau]} \left| G\left(\hat{F}_n - F_\theta\right) - G\left(F_{\theta_0} - F_\theta\right) \right| \xrightarrow{P} 0.$$

Using again that by Condition B we have $\sup_{[0,\tau]} |w_n| = O_P(1)$, we obtain that the second summand in (10) is also $o_P(1)$. Hence, as a consequence the remaining term in (9) is also a $o_P(1)$.

Putting the above results all together we obtain for n large enough

$$\begin{aligned}
\int_0^\tau G\left(F_{\theta_0} - F_{\hat{\theta}_n}\right) w_0 dF_{\theta_0} &\overset{(9)}{\leq} \int_0^\tau G_n\left(\hat{F}_n - F_{\hat{\theta}_n}, \omega\right) w_n d\hat{F}_n + o_P(1) \\
&\overset{(8)}{\leq} \int_0^\tau G_n\left(\hat{F}_n - F_{\theta_0}, \omega\right) w_n d\hat{F}_n + o_P(1) \\
&\overset{(7)}{\leq} \sup_{[0,\tau]} G\left(\hat{F}_n - F_{\theta_0} + o_p(1/\sqrt{n})\right) \sup_{[0,\tau]} |w_n| + o_P(1) \\
&\leq \sup_{[-1,1]} |G'| \left(\sup_{[0,\tau]} \left| \hat{F}_n - F_{\theta_0} + o_p(1/\sqrt{n}) \right| \right) \sup_{[0,\tau]} |w_n| \\
&\quad + o_P(1).
\end{aligned} \tag{12}$$

Hence, using (11) we obtain than under Conditions B and C the right hand side in (12) converges to zero in probability. We conclude by Condition D that $(\hat{\theta}_n)_{n\geq 1}$ converges to θ_0 in probability. □

Remark 2. (i) It can be seen from the above proof that the consistency result also holds if we only require $G_n(\cdot,\omega) = G(\cdot + o_P(1))$.
(ii) The above uniform consistency result also holds with probability one if the $o_p(1/\sqrt{n})$ term (or the $o_P(1)$ term) converges almost surely since, from Stute & Wang (1993), we have that the uniform convergence of the Kaplan-Meier estimator holds almost surely on $[0,\tau]$. For example, the $o_p(1/\sqrt{n})$ term converges almost surely for $G_n(\cdot,\omega) = (\cdot + 1/(2n))^2$ and
$G_n(\cdot,\omega) = \left(\cdot + 1/\left(2\sum_{i=1}^n I_{\{X_i\leq C_i, X_i\leq\tau\}}\right)\right)^2$.

Before continuing with the asymptotic distribution we introduce the following covariance function for $(s,t) \in [0,\tau]^2$

$$\rho(s,t) = (1 - F_{\theta_0}(s))(1 - F_{\theta_0}(t)) \int_0^{s\wedge t} \frac{dF_{\theta_0}(u)}{(1 - F_{\theta_0}(u))^2(1 - H(u))},$$

and we state the next result.

Lemma 2. *Under Conditions A–E,* $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ *is bounded in probability.*

The proof is presented in the Appendix. We now proceed with the asymptotic distribution.

Theorem 2. *Let* $\mathcal{B}$ *be a centered Gaussian process on* $[0,\tau]$ *with covariance function* ρ*. If Conditions A–E hold, then* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *converges to a centered normal distribution with variance* $\Sigma^{-1}(\tau)C(\tau)\Sigma^{-1}(\tau)$ *where* $C(\tau) = \mathrm{Var}\left(\int_0^\tau \mathcal{B}\eta \circ F_{\theta_0} w_0 dF_{\theta_0}\right)$.

Proof. First let us define the two following sequences of stochastic processes $\bar{X}_n$ and X_n by

$$\bar{X}_n(\xi) = \frac{G''(0)}{2} \int_0^\tau \left(\mathcal{B}_n - \xi^t \eta \circ F_{\theta_0}\right)^2 w_0 dF_{\theta_0}, \quad \xi \in \mathbb{R}^p,$$

where $\mathcal{B}_n = \sqrt{n}(\hat{F}_n - F_{\theta_0})$ and

$$X_n(\xi) = n \int_0^\tau G_n\left(\hat{F}_n - F_{\theta_0 + \xi/\sqrt{n}}, \omega\right) w_n d\hat{F}_n, \quad \xi \in \mathbb{R}^p.$$

Then for any $A = \{\xi \in \mathbb{R}^p; \|\xi\| < c\}$ we have

$$\sup_{\xi \in A} \left|X_n(\xi) - \bar{X}_n(\xi)\right| \xrightarrow{P} 0. \tag{13}$$

Indeed first note that

$$\begin{aligned} \sup_{[0,\tau]} |\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}}| &\leq \sup_{[0,\tau]} |\hat{F}_n - F_{\theta_0}| + \sup_{[0,\tau]} |F_{\theta_0} - F_{\theta_0+\xi/\sqrt{n}}| \\ &\leq \sup_{[0,\tau]} |\hat{F}_n - F_{\theta_0}| + \sup_{[0,\tau]} \left|\frac{1}{\sqrt{n}} \xi^t \eta \circ F_{\theta_0}\right| + \frac{1}{\sqrt{n}} o(\|\xi\|), \quad (14) \end{aligned}$$

where the right hand side converges to 0 in probability since $\xi \in A$ and by using Condition E and properties of the Kaplan-Meier estimator. Multiplying the above inequality by $\sqrt{n}$ we also obtain that

$$\sup_{\xi \in A} \sup_{[0,\tau]} \sqrt{n} |\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}}| = O_P(1).$$

Now by using Condition C and a second order expansion of G around 0 we have

$$
\begin{aligned}
X_n(\xi) &= n\int_0^\tau G\left(\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}} + o_p(1/\sqrt{n})\right) w_n d\hat{F}_n, \\
&= \frac{n}{2}\int_0^\tau G''(H_n)\left(\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}} + o_P(1/\sqrt{n})\right)^2 w_n d\hat{F}_n, \qquad (15)
\end{aligned}
$$

where H_n belongs to the line segment connecting 0 and $\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}} + o_p(1/\sqrt{n})$. Since $|H_n| \leq |\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}} + o_p(1/\sqrt{n})|$ we have from (14) that $\sup_{[0,\tau]}|H_n| = o_P(1)$ and obtain

$$
\begin{aligned}
X_n(\xi) &= \frac{n}{2}G''(0)\int_0^\tau \left(\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}} + o_p(1/\sqrt{n})\right)^2 w_n d\hat{F}_n + o_P(1) \\
&= \frac{n}{2}G''(0)\int_0^\tau \left(\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}} + o_p(1/\sqrt{n})\right)^2 w_0 d\hat{F}_n + o_P(1),
\end{aligned}
$$

where the last equality follows from Condition B and where the $o_P(1)$ is uniform in $\xi \in A$. Then we can write

$$
\begin{aligned}
X_n(\xi) &= \frac{n}{2}G''(0)\int_0^\tau \left(\hat{F}_n - F_{\theta_0} + F_{\theta_0} - F_{\theta_0+\xi/\sqrt{n}} + o_p(1/\sqrt{n})\right)^2 w_0 dF_{\theta_0} \\
&\quad + \frac{n}{2}G''(0)\int_0^\tau \left(\hat{F}_n - F_{\theta_0} + F_{\theta_0} - F_{\theta_0+\xi/\sqrt{n}} + o_p(1/\sqrt{n})\right)^2 w_0 d(\hat{F}_n - F_{\theta_0}) \\
&= \frac{1}{2}G''(0)\int_0^\tau \left(\sqrt{n}(\hat{F}_n - F_{\theta_0}) - \xi^t\eta\circ F_{\theta_0} + R_n + o_P(1)\right)^2 w_0 dF_{\theta_0} \\
&\quad + \frac{1}{2}G''(0)\int_0^\tau \left(\sqrt{n}(\hat{F}_n - F_{\theta_0}) - \xi^t\eta\circ F_{\theta_0} + R_n + o_P(1)\right)^2 w_0 d(\hat{F}_n - F_{\theta_0}) \\
&= \bar{X}_n(\xi) + \frac{1}{2}G''(0)\int_0^\tau (R_n + o_P(1))[R_n + o_P(1) + 2(\mathcal{B}_n - \xi^t\eta\circ F_{\theta_0})] w_0 dF_{\theta_0} \\
&\quad + \frac{1}{2}G''(0)\int_0^\tau \left(\mathcal{B}_n - \xi^t\eta\circ F_{\theta_0} + R_n + o_P(1)\right)^2 w_0 d(\hat{F}_n - F_{\theta_0}).
\end{aligned}
$$

Since R_n converges uniformly to 0 on $[0,\tau]$ by Condition E, since $\sup_{[0,\tau]}|\mathcal{B}_n|$ is bounded in probability, since $\xi \in A$ and because the component functions of $\eta\circ F_{\theta_0}$ are in $L^2(0,q)$ the two last terms of the right hand side of the last equation are $o_P(1)$ uniformly in ξ, and (13) holds.

Let $\bar{\xi}_n$ be the maximizer of $\bar{X}_n$ it is straightforward to see that

$$
\bar{\xi}_n = \Sigma^{-1}(\tau)\int_0^\tau \mathcal{B}_n \eta\circ F_{\theta_0} w_0 dF_{\theta_0}
$$

converges weakly to

$$\xi_0 = \Sigma^{-1}(\tau) \int_0^\tau \mathcal{B}\eta \circ F_{\theta_0} w_0 dF_{\theta_0}.$$

Let $\hat{\xi}_n$ be the maximizer of X_n it easy to see that $\hat{\xi}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$. Now, to prove the theorem, let us show that $\hat{\xi}_n - \bar{\xi}_n = o_P(1)$.

Let $\varepsilon > 0$ be a real number. Notice that from Lemma 2 and the weak convergence of $\bar{\xi}_n$ to the normal random variable ξ_0 it is possible to chose $c > 0$ such that the probability of $E_n = \{\bar{\xi}_n \in A, \hat{\xi}_n \in A\}$ is as large as we want for n large enough.

Let us define $B_n = \{\xi \in \mathbb{R}^p; \|\xi - \bar{\xi}_n\| < \varepsilon\}$. We have

$$\{\hat{\xi}_n \in A \backslash B_n\} \subset \left\{ \inf_{\xi \in A\backslash B_n} X_n(\xi) \leq X_n(\bar{\xi}_n) \right\}$$

and if $\xi \in A\backslash B_n$ we can write

$$\begin{aligned} X_n(\xi) &= X_n(\xi) - \bar{X}_n(\xi) + \bar{X}_n(\xi) \\ &\geq X_n(\xi) - \bar{X}_n(\xi) + \inf_{\xi \in A\backslash B_n} \bar{X}_n(\xi) \\ &\geq -\sup_{\xi \in A} \left| X_n(\xi) - \bar{X}_n(\xi) \right| + \inf_{\xi \in A\backslash B_n} \bar{X}_n(\xi), \end{aligned}$$

thus

$$\inf_{\xi \in A\backslash B_n} X_n(\xi) \geq \inf_{\xi \in A\backslash B_n} \bar{X}_n(\xi) - \sup_{\xi \in A} \left| X_n(\xi) - \bar{X}_n(\xi) \right|$$

and therefore

$$\begin{aligned} & \{\hat{\xi}_n \in A\backslash B_n\} \\ \subset & \left\{ \inf_{\xi \in A\backslash B_n} X_n(\xi) \leq X_n(\bar{\xi}_n) \right\} \\ \subset & \left\{ \inf_{\xi \in A\backslash B_n} \bar{X}_n(\xi) - \sup_{\xi \in A} \left| X_n(\xi) - \bar{X}_n(\xi) \right| \leq X_n(\bar{\xi}_n) \right\} \\ \subset & \left\{ \inf_{\xi \in A\backslash B_n} \bar{X}_n(\xi) - \bar{X}_n(\bar{\xi}_n) \leq \sup_{\xi \in A} \left| X_n(\xi) - \bar{X}_n(\xi) \right| + X_n(\bar{\xi}_n) - \bar{X}_n(\bar{\xi}_n) \right\} \\ \subset & \left\{ \inf_{\xi \in A\backslash B_n} \bar{X}_n(\xi) - \bar{X}_n(\bar{\xi}_n) \leq 2 \sup_{\xi \in A} \left| X_n(\xi) - \bar{X}_n(\xi) \right| \right\} \\ \subset & \left\{ G''(0)\varepsilon^2 \sigma(\tau) \leq 2 \sup_{\xi \in A} \left| X_n(\xi) - \bar{X}_n(\xi) \right| \right\} \end{aligned} \tag{16}$$

with $\sigma(\tau) > 0$ the smallest eigenvalue of $\Sigma(\tau)$ and where the last inclusion holds because $\bar{X}_n(\xi)$ being quadratic in ξ we have

$$\begin{aligned}\bar{X}_n(\xi) - \bar{X}_n(\bar{\xi}_n) &= \left(\xi - \bar{\xi}_n\right)^t \underbrace{\bar{X}_n'(\bar{\xi}_n)}_{=0} + \frac{1}{2}\left(\xi - \bar{\xi}_n\right)^t G''(0)\Sigma(\tau)\left(\xi - \bar{\xi}_n\right) \\ &= \frac{1}{2} G''(0)\left(\xi - \bar{\xi}_n\right)^t \Sigma(\tau)\left(\xi - \bar{\xi}_n\right) \\ &\geq \frac{1}{2} G''(0)\varepsilon^2 \sigma(\tau) \quad \text{on } A \backslash B_n.\end{aligned}$$

Combining (13), (16) and the fact that the probability of E_n is as close to one as we want we conclude that $\hat{\xi}_n - \bar{\xi}_n = o_P(1)$. By the Slutsky lemma it follows that $\hat{\xi}_n$ converges weakly to ξ_0 which finishes the proof. □

Remark 3. If $w_0 \equiv 1$ the asymptotic variance of the estimator $\hat{\theta}_n$ is the same for all sample size and ω-dependent functions G_n fulfilling Condition C. This extends an observation by Öztürk & Hettmansperger (1997) who remarked that taking $G(\hat{F}_n - F_\theta)$, where G fulfills some regularity conditions, instead of $(\hat{F}_n - F_\theta)^2$ as in Woodward et al. (1984) does not change the asymptotic variance of the estimator $\hat{\theta}_n$ when $w_0 \equiv 1$.

3.3 Robustness

One of the main advantages of using minimum distance estimators is their stability in the neighborhood of the model. Öztürk & Hettmansperger (1997) showed that the influence curve at the model of their estimator defined by (3) does not depend on the choice of G; see also Lindsay (1994). Therefore, the robustness of their estimator can be obtained by choosing an appropriate weight function. Here, in contrast to the approach of Öztürk & Hettmansperger (1997), we are integrating with respect to the empirical measure if there is no censoring, i.e. $P(C_i = \infty) = 1$, and with respect to the measure induced by the Kaplan-Meier estimator if censoring is present. The influence curves in the latter case are readily obtained by using the results of Reid (1981) once we have obtained the influence curve in the former case. Therefore, if there is no censoring we obtain the next theorem.

Theorem 3. *Let H be an arbitrary distribution function, Δ_x the degenerate distribution function at x, and $T(H) = \arg\min_{\theta\in\Theta} \int_0^\tau G_n\left(H - F_\theta, \omega\right) w_n dH$. Sup-*

pose that

$$M_n(H) = \int_0^\tau \ddot{G}_n(H - F_\theta)_{|\theta=T(H)} w_n dH$$

is invertible. Then

(i) If w_n does not depend on H the influence curve of H at Δ_x, denoted by $IC_{\Delta_x}(H)$, is given by

$$\begin{aligned} IC_{\Delta_x}(H) &= M_n^{-1}(H) \left[\int_0^\tau \dot{G}_n(H - F_\theta)_{|\theta=T(H)} \times w_n d(\Delta_x - H) \right. \\ &\quad \left. + \int_0^\tau G_n'' \left(H - F_{T(H)}\right) \dot{F}_{\theta|_{\theta=T(H)}} [\Delta_x - H] w_n dH \right]. \end{aligned}$$

(ii) If w_n depends on H the influence curve of H at Δ_x is given by

$$IC_{\Delta_x}(H) + M_n^{-1}(H) \int_0^\tau G_n' \left(H - F_{T(H)}\right) \dot{F}_{\theta|_{\theta=T(H)}} \times [\Delta_x - H] w_n'(H)\, dH.$$

If H belongs to the parametric family $\mathcal{F}$ and if $G_n(\cdot, \omega) = G(\cdot)$, the influence curve simplifies as follows.

Lemma 3. *If $H \in \mathcal{F}$ and $G_n(\cdot, \omega) = G(\cdot)$, then (i) and (ii) in Theorem 3 are equal and $IC_{\Delta_x}(H)$ is given by*

$$IC_{\Delta_x}(H) = \tilde{M}_n^{-1}(H) \int_0^\tau \dot{F}_{\theta|_{\theta=T(H)}} \times [\Delta_x - H] w_n dH,$$

where

$$\tilde{M}_n(H) = \int_0^\tau \left(\dot{F}_\theta \dot{F}_\theta^t\right)_{|\theta=T(H)} \times w_n dH.$$

The proofs are given in the Appendix.
If $G_n(\cdot, \omega) = G(\cdot)$ as in Öztürk & Hettmansperger (1997) we see from Lemma 3 that their above mentioned observation is still true if we integrate with respect to the empirical measure instead of Lebesgue measure, i.e. by choosing an appropriate weight function the influence curves are bounded with respect to x. Moreover, we already observed in Theorem 2 that the asymptotic distribution of the estimator does not depend on the choice of $G_n(\cdot, \omega) = G(\cdot)$; a result which is also suggested by Theorem 3 since we have $G'(0) = 0$ and $o_P(1/\sqrt{n}) \to 0$.

Finally, if censoring is present we obtain the influence curves from Theorem 3, Lemma 3 and by using the results of Reid (1981). Under the conditions of Lemma 3 they are given by:

$$
\begin{aligned}
IC^1_{\Delta_x}(H) &= -\tilde{M}_n^{-1}(H)\int_0^\tau H(t)\left(\int_0^{x\wedge t}\frac{dS^u(t)}{(S^u(t)+S^c(t))^2}+\frac{I_{\{x\le t\}}}{S^u(t)+S^c(t)}\right)\\
&\qquad \times \dot{F}_\theta(t)_{|\theta=T(H)}w_n(t)H'_+(t)dt,\\
IC^2_{\Delta_x}(H) &= -\tilde{M}_n^{-1}(H)\int_0^\tau H(t)\left(\int_0^{x\wedge t}\frac{dS^u(t)}{(S^u(t)+S^c(t))^2}\dot{F}_\theta(t)_{|\theta=T(H)}\right)\\
&\qquad \times w_n(t)H'_+(t)dt
\end{aligned}
$$

where H'_+ is the right hand derivative of H, $S^u(t) = P(T_i > t, \Delta_i = 1)$, and $S^c(t) = P(T_i > t, \Delta_i = 0)$.

4 Simulation study

In this section, we present some findings from extensive simulations to compare several estimators contained in our class of estimators based on ω-dependent generalized weighted Cramér-von Mises distances and to illustrate the usefulness of the estimators for moderate sample sizes (Subsection 4.1). Furthermore, we study the behavior of the estimators under three different contamination models (Subsection 4.2).

4.1 Illustrative examples - part I

For the simulation results presented here we took a two component Weibull mixture with d.f. $F_\theta(x) = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$, and we assumed that only the mixing proportion $\pi = 0.3$ is unknown. We assumed that censoring might be present and we simulated data from the above mixture model with 0%, 20%, 40% and 60% censoring. The censoring times were assumed to be exponentially distributed. The sample size were taken to be equal to 20, 40, and 80. For each amount of censoring, and each sample size (12 cases in all), we simulated $N = 10,000$ samples and we calculated the mean $\hat{\pi}$ and the mean square error (MSE) of different estimators. We took two groups of estimators. For the first group, which we shall call G_1 type estimators, the basis was the function $G(x) = x^2$, and we looked at several modifications of this function, for example, sample size dependent modifications, i.e. $G_n(x) = (x + 1/(2n))^2$ or $G_n(x) = (x + 1/n)^2$. These functions were then combined with different weight functions. The results are given in Tables 1, 2 and 3. In the first column the different estimators can be found and the rows give the effect of censoring. For the second group called G_3 type estimators we took the function $G_3(x) = (\sqrt{x+1} - 1)^2$ as the basis (cf. Öztürk & Hettmansperger (1997)). We looked again at modifications of this function, for example $G_n(x) = (\sqrt{x + 1 + 1/(2n)} - 1)^2$ or $G_n(x) = (\sqrt{x + 1 + 1/n} - 1)^2$. Again, we combined these functions with different weight functions. The results are given in Tables 4, 5 and 6. The values given in Tables 1–6 were all calculated with the same (simulated) data. Throughout, c_n stands for $(2n - 2\sum_{i=1}^{n} \delta_i)$ where $\delta_i = I_{\{C_i < T_i\}}, i = 1, \ldots, n$.

Censoring		0%		20%		40%		60%	
$G_n(\cdot)$	w_n	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
$(\cdot)^2$	1	0.254	0.132	0.241	0.142	0.218	0.148	0.172	0.153
$(\cdot + 1/(2n))^2$	1	0.297	0.136	0.283	0.145	0.258	0.154	0.206	0.163
$(\cdot + 1/n)^2$	1	0.341	0.135	0.326	0.147	0.299	0.158	0.243	0.171
$(\cdot + 1/c_n)^2$	1	0.297	0.135	0.294	0.147	0.288	0.161	0.273	0.191
$(\cdot)^2$	$(1 - \hat{F}_n)^2$	0.264	0.158	0.258	0.167	0.250	0.175	0.223	0.185
$(\cdot)^2$	$(1 - \hat{F}_n)^4$	0.268	0.180	0.265	0.186	0.260	0.195	0.237	0.203
$(\cdot + 1/(2n))^2$	$(1 - \hat{F}_n)^{1.1}$	0.307	0.149	0.300	0.159	0.286	0.169	0.253	0.185
$(\cdot + 1/(2n))^2$	$(1 - \hat{F}_n)^2$	0.314	0.161	0.308	0.171	0.298	0.180	0.267	0.194
$(\cdot + 1/c_n)^2$	$(1 - \hat{F}_n)^{1.1}$	0.307	0.149	0.311	0.161	0.319	0.176	0.331	0.211

Table 1: Estimation of the mixture parameter $\pi = 0.3$ *based on distance functions of the type* G_1 *and on* $n = 20$ *observations from a Weibull mixture with cdf* $F_\theta(x) = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3))$. *The values given in the table are based on 10,000 simulations.*

Censoring		0%		20%		40%		60%	
$G_n(\cdot)$	w_n	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
$(\cdot)^2$	1	0.276	0.096	0.270	0.106	0.257	0.115	0.221	0.129
$(\cdot+1/(2n))^2$	1	0.298	0.096	0.291	0.106	0.278	0.115	0.242	0.131
$(\cdot+1/n)^2$	1	0.320	0.095	0.313	0.106	0.300	0.116	0.263	0.133
$(\cdot+1/c_n)^2$	1	0.298	0.098	0.297	0.106	0.293	0.117	0.276	0.138
$(\cdot)^2$	$(1-\hat{F}_n)^2$	0.279	0.121	0.275	0.130	0.271	0.137	0.256	0.153
$(\cdot)^2$	$(1-\hat{F}_n)^4$	0.279	0.144	0.277	0.152	0.274	0.157	0.264	0.171
$(\cdot+1/(2n))^2$	$(1-\hat{F}_n)^{1.1}$	0.303	0.109	0.298	0.119	0.292	0.127	0.272	0.145
$(\cdot+1/(2n))^2$	$(1-\hat{F}_n)^2$	0.306	0.120	0.302	0.130	0.297	0.137	0.281	0.154
$(\cdot+1/c_n)^2$	$(1-\hat{F}_n)^{1.1}$	0.303	0.109	0.304	0.119	0.307	0.129	0.310	0.151

Table 2: Estimation of the mixture parameter $\pi = 0.3$ based on distance functions of the type G_1 and on $n = 40$ observations from a Weibull mixture with cdf $F_\theta(x) = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$. The values given in the table are based on 10,000 simulations.

Censoring		0%		20%		40%		60%	
$G_n(\cdot)$	$w_n = 1$	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
$(\cdot)^2$	1	0.289	0.069	0.284	0.074	0.278	0.082	0.258	0.098
$(\cdot+1/(2n))^2$	1	0.300	0.069	0.295	0.074	0.289	0.082	0.269	0.098
$(\cdot+1/n)^2$	1	0.311	0.069	0.306	0.074	0.300	0.082	0.279	0.098
$(\cdot+1/c_n)^2$	1	0.300	0.069	0.297	0.074	0.296	0.083	0.285	0.099
$(\cdot)^2$	$(1-\hat{F}_n)^2$	0.289	0.088	0.286	0.093	0.284	0.101	0.277	0.116
$(\cdot)^2$	$(1-\hat{F}_n)^4$	0.287	0.108	0.285	0.112	0.284	0.120	0.280	0.133
$(\cdot+1/(2n))^2$	$(1-\hat{F}_n)^{1.1}$	0.301	0.078	0.298	0.083	0.295	0.092	0.285	0.107
$(\cdot+1/(2n))^2$	$(1-\hat{F}_n)^2$	0.303	0.087	0.300	0.092	0.297	0.101	0.290	0.116
$(\cdot+1/c_n)^2$	$(1-\hat{F}_n)^{1.1}$	0.301	0.078	0.301	0.083	0.303	0.092	0.304	0.109

Table 3: Estimation of the mixture parameter $\pi = 0.3$ based on distance functions of the type G_1 and on $n = 80$ observations from a Weibull mixture with cdf $F_\theta(x) = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$. The values given in the table are based on 10,000 simulations.

In what follows we discuss how the use of sample size dependent and censoring sensitive estimators as well as the use of weight functions affect the quality of the estimation. A general observation, as expected, is that as the sample size increases, the bias due to censoring decreases. More explicitly, consider the first

three rows of Tables 1, 2 and 3. If there is no censoring the use of the correction term $1/(2n)$, i.e. $G_n(\cdot,\omega) = (\cdot + 1/(2n))^2$ leads to an estimator with a negligible bias, whereas the uncorrected version, i.e. $G_n(\cdot,\omega) = (\cdot)^2$, leads to underestimation, and the correction term $1/n$, i.e. $G_n(\cdot,\omega) = (\cdot + (1/n))^2$, leads to overestimation. As censoring increases from 0% to 60% the mean of all three estimators also decreases. This decrease in the mean is for all three estimators of approximately the same value. As already mentioned increasing the sample size leads to a smaller bias for this three estimators. Comparing this three estimators with the sample size dependent and censoring sensitive version, i.e. $G_n(\cdot,\omega) = (\cdot + 1/c_n)^2$, it is directly seen from Tables 1, 2 and 3 that the overall performance of the sample size dependent and censoring sensitive version is much better. Even for a censoring of 60% the mean of the estimate for $n = 20$, $n = 40$, and $n = 80$ is 0.273, 0.276, and 0.285. The reason for this good performance seems to be as follows: In the complete i.i.d case based on n failures the correction term $1/(2n)$ leads to nearly unbiased estimates. Under censoring we only observe $n - \sum_{i=1}^n \delta_i$ failures. Therefore, using the correction term $1/(2(n - \sum_{i=1}^n \delta_i))$ seems to be appropriate in the case of right censoring.
Comparing the first row of Tables 1, 2, and 3 with the fourth and fifth line of these Tables, clarifies the use of a weight function in the case where G_n does neither depend on the sample size nor on the censoring. As censoring increases the weighted versions of $G_n(\cdot) = (\cdot)^2$ performs much better than the unweighted version. A similar behavior is seen if we take $G_n(\cdot) = (\cdot + 1/(2n))^2$ although the unweighted version already leads to considerable good results. Thus, when censoring is present, the small sample bias can be further reduced by using a weight function which de-emphasize the largest observations. (cf. lines 2, 7 and 8 of Tables 1, 2 and 3).

Remark 4. It should be mentioned that using $G_n(\cdot,\omega) = (\cdot)^2$ with $w_n = 1$ and $G_n(\cdot,\omega) = (\cdot + 1/n)^2$ with $w_n = 1$ does not always lead to underestimation and overestimation, respectively. For example, if we take $F_\theta(x) = 1 - 0.3\exp(-(x/2)^3) - 0.7\exp(-(x/5)^3)$ instead of $F_\theta(x) = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$, then $G_n(\cdot,\omega) = (\cdot)^2$ with $w_n = 1$ leads to overestimation and $G_n(\cdot,\omega) = (\cdot + 1/n)^2$ with $w_n = 1$ to underestimation, whereas the behavior of $G_n(\cdot,\omega) = (\cdot + 1/(2n))^2$ with $w_n = 1$, and $G_n(\cdot,\omega) = (\cdot + 1/c_n)^2$ with $w_n = 1$, respectively, is not changed.

Censoring		0%		20%		40%		60%	
$G_n(\cdot)$	w_n	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
$(\sqrt{\cdot+1}-1)^2$	1	0.250	0.132	0.236	0.141	0.212	0.147	0.165	0.150
$(\sqrt{\cdot+1+1/(2n)}-1)^2$	1	0.293	0.135	0.278	0.145	0.252	0.153	0.199	0.161
$(\sqrt{\cdot+1+1/n}-1)^2$	1	0.336	0.136	0.321	0.147	0.293	0.157	0.235	0.170
$(\sqrt{\cdot+1+1/c_n}-1)^2$	1	0.293	0.135	0.289	0.147	0.281	0.160	0.264	0.189
$(\sqrt{\cdot+1}-1)^2$	$(1-\hat{F}_n)^2$	0.259	0.158	0.253	0.166	0.244	0.174	0.217	0.183
$(\sqrt{\cdot+1}-1)^2$	$(1-\hat{F}_n)^4$	0.263	0.180	0.260	0.186	0.254	0.194	0.231	0.202
$(\sqrt{\cdot+1+1/(2n)}-1)^2$	$(1-\hat{F}_n)^{1.1}$	0.302	0.149	0.294	0.159	0.280	0.168	0.247	0.183
$(\sqrt{\cdot+1+1/(2n)}-1)^2$	$(1-\hat{F}_n)^2$	0.309	0.161	0.302	0.170	0.291	0.179	0.260	0.193
$(\sqrt{\cdot+1+1/c_n}-1)^2$	$(1-\hat{F}_n)^{1.1}$	0.302	0.149	0.305	0.161	0.312	0.175	0.323	0.209

Table 4: Estimation of the mixture parameter $\pi = 0.3$ based on distance functions of the type G_3 and on $n = 20$ observations from a Weibull mixture with cdf $F_\theta(x) = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$. The values given in the table are based on 10,000 simulations.

Censoring		0%		20%		40%		60%	
$G_n(\cdot)$	w_n	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
$(\sqrt{\cdot+1}-1)^2$	1	0.274	0.096	0.267	0.106	0.253	0.115	0.216	0.128
$(\sqrt{\cdot+1+1/(2n)}-1)^2$	1	0.296	0.096	0.289	0.106	0.275	0.115	0.236	0.130
$(\sqrt{\cdot+1+1/n}-1)^2$	1	0.318	0.096	0.311	0.106	0.297	0.116	0.257	0.132
$(\sqrt{\cdot+1+1/c_n}-1)^2$	1	0.296	0.096	0.294	0.106	0.289	0.117	0.270	0.137
$(\sqrt{\cdot+1}-1)^2$	$(1-\hat{F}_n)^2$	0.276	0.122	0.272	0.131	0.267	0.137	0.251	0.152
$(\sqrt{\cdot+1}-1)^2$	$(1-\hat{F}_n)^4$	0.276	0.144	0.273	0.152	0.270	0.157	0.259	0.170
$(\sqrt{\cdot+1+1/(2n)}-1)^2$	$(1-\hat{F}_n)^{1.1}$	0.300	0.109	0.295	0.119	0.288	0.127	0.267	0.145
$(\sqrt{\cdot+1+1/(2n)}-1)^2$	$(1-\hat{F}_n)^2$	0.303	0.121	0.299	0.130	0.293	0.137	0.276	0.154
$(\sqrt{\cdot+1+1/c_n}-1)^2$	$(1-\hat{F}_n)^{1.1}$	0.300	0.109	0.301	0.119	0.307	0.129	0.305	0.151

Table 5: Estimation of the mixture parameter $\pi = 0.3$ based on distance functions of the type G_3 and on $n = 40$ observations from a Weibull mixture with cdf $F_\theta(x) = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$. The values given in the table are based on 10,000 simulations.

Censoring		0%		20%		40%		60%	
$G_n(\cdot)$	w_n	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
$(\sqrt{\cdot+1}-1)^2$	1	0.287	0.069	0.283	0.074	0.276	0.082	0.254	0.097
$(\sqrt{\cdot+1+1/(2n)}-1)^2$	1	0.298	0.069	0.294	0.074	0.287	0.082	0.265	0.098
$(\sqrt{\cdot+1+1/n}-1)^2$	1	0.309	0.069	0.305	0.074	0.298	0.082	0.276	0.098
$(\sqrt{\cdot+1+1/c_n}-1)^2$	1	0.298	0.069	0.296	0.074	0.294	0.083	0.282	0.099
$(\sqrt{\cdot+1}-1)^2$	$(1-\hat{F}_n)^2$	0.287	0.088	0.284	0.093	0.281	0.102	0.274	0.116
$(\sqrt{\cdot+1}-1)^2$	$(1-\hat{F}_n)^4$	0.285	0.108	0.283	0.113	0.281	0.120	0.277	0.134
$(\sqrt{\cdot+1+1/(2n)}-1)^2$	$(1-\hat{F}_n)^{1.1}$	0.300	0.079	0.296	0.083	0.293	0.092	0.282	0.108
$(\sqrt{\cdot+1+1/(2n)}-1)^2$	$(1-\hat{F}_n)^2$	0.301	0.088	0.298	0.092	0.295	0.101	0.287	0.116
$(\sqrt{\cdot+1+1/c_n}-1)^2$	$(1-\hat{F}_n)^{1.1}$	0.300	0.079	0.299	0.084	0.301	0.092	0.301	0.109

Table 6: Estimation of the mixture parameter $\pi = 0.3$ based on distance functions of the type G_3 and on $n = 80$ observations from a Weibull mixture with cdf $F_\theta(x) = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$. The values given in the table are based on 10,000 simulations.

The qualitative behavior of the estimators based on versions of G_3 seems to be the same as for G_1. It is interesting to see, that the correction term $1/(2n)$, which was suggested by MacDonald to improve the small sample behavior of $(\cdot)^2$, seems also to improve the small sample behavior of the G_3 based estimator. Moreover, comparing the results in Tables 1–6 it seems that $G_n(\cdot) = (\sqrt{\cdot+1+1/c_n}-1)^2$ with $w_n = (1-\hat{F}_n)^{1.1}$ has the best performance in terms of small sample bias without decreasing too much the MSE performances.

4.2 Illustrative examples - part II

It is well known that minimum distance estimators, for example for a location parameter, are robust against symmetric contamination models. To study the behavior of the estimators defined by (5) under a contamination model when censoring might be present, we took three different contamination models, namely, $CM_i = \{\tilde{F} = (1-\epsilon)F + \epsilon H_i\}$, for $i = 1, 2, 3$ where $\epsilon = 0.05$, F is the d.f. of the Part-I Weibull mixture, and H_1, H_2 and H_3 are gamma mixtures df with respective density functions:

$$\begin{aligned} h_1(x) &= 0.1/5\exp(-x/5) + 0.9/(2^{1.4}\Gamma(1.4))x^{0.4}\exp(-x/2), \\ h_2(x) &= 0.1/(5^{0.5}\Gamma(0.5))x^{-0.5}\exp(-x/5) + 0.9/(2^{0.2}\Gamma(0.2))x^{-0.8}\exp(-x/2), \\ h_3(x) &= 0.1/(5^{1.5}\Gamma(1.5))x^{0.5}\exp(-x/5) + 0.9/(2^{4.5}\Gamma(4.5))x^{3.5}\exp(-x/2). \end{aligned}$$

We simulated data from the above contamination models with censoring equal to 0% and 40%. In the case of censoring, the censoring time was taken to be exponential with parameter 0.218. The sample size was taken to be equal to 40. Please note that h_1 corresponds to a 'symmetric' contamination model in the sense that $P(X < Y_1) \approx P(X > Y_1)$, where $X \sim F, Y_1 \sim H_1$, and that the probability of being censored is approximately equal for X and Y_1. Furthermore, h_2 corresponds to a 'left' contamination model, i.e. $P(X < Y_2) \approx 9\%$, $Y_2 \sim H_2$, and the probability for X to be censored is much larger than the probability for Y_2 to be censored, and h_3 corresponds to a 'right' contamination model, i.e. $P(X < Y_3) \approx 93\%$, $Y_3 \sim H_3$, the probability for Y_3 to be censored much larger than the probability for X to be censored. The results are given in Tables 7 (CM_1), 8 (CM_2), and 9 (CM_3). The qualitative behavior of the estimators based on G_1 and G_3, respectively, is the same. We therefore only present the results for the G_1-type estimators.

Now, consider Table 7. Comparing these values with the values in Table 2 it is easily seen that the values of $\hat{\pi}$ are hardly affected, both, when there is no censoring and when censoring is present. In particular, the size dependent and censoring sensitive estimator $G_n(\cdot) = (\cdot + 1/c_n)^2$ still performs rather well. To conclude, it seems that the estimators are fairly robust to 'symmetric' contamination models even when censoring is present. Let us briefly discuss the effects of CM_2 and CM_3. When there is no censoring we have under CM_2 and CM_3 that $\hat{\pi}$ is shifted to the left and to the right, respectively (cf. Tables 8 and 9). A simple heuristic seems to explain these observations. On average, we have under CM_2 two small observations. Therefore, by minimizing the distance between the Weibull mixture and the empirical distribution function one puts less weight on the stochastically larger component of the mixture, i.e. the Weibull distribution with d.f. $1 - \exp(-(x/5)^3)$. Under CM_3 we have, on average, two large observations. Thus, one puts less weight on the stochastically smaller component of the Weibull mixture, i.e. the Weibull distribution with d.f. $1 - \exp(-(x/2)^3)$.

This behavior can also be seen from Lemma 3. Since for $F_\theta(t) = 1 - \theta\exp(-(t/5)^3) - (1-\theta)\exp(-(t/2)^3)$ and $w_n \equiv 1$ the influence function of H at x is, up to a positive constant, given by

$$\int_0^\tau \big(-\exp(-(t/5)^3) + \exp(-(t/2)^3)\big) \cdot (I_{\{t\geq x\}} - H(t))dH(t).$$

This expression is negative for small x and positive for large x, $x \leq \tau$.
Given the above observations, one would expect that under right censoring $\hat{\pi}$ is further shifted to the left under CM_2 and that under CM_3 the shift to the right of $\hat{\pi}$ is reduced. That is exactly what can be seen from Tables 8 and 9.

Censoring		0%		40%	
$G_n(\cdot)$	w_n	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
$(\cdot)^2$	1	0.278	0.100	0.258	0.117
$(\cdot+1/(2n))^2$	1	0.300	0.100	0.280	0.117
$(\cdot+1/n)^2$	1	0.322	0.100	0.301	0.118
$(\cdot+1/c_n)^2$	1	0.300	0.100	0.295	0.119
$(\cdot)^2$	$(1-\hat{F}_n)^2$	0.274	0.125	0.267	0.141
$(\cdot)^2$	$(1-\hat{F}_n)^4$	0.268	0.148	0.265	0.162
$(\cdot+1/(2n))^2$	$(1-\hat{F}_n)^{1.1}$	0.300	0.113	0.290	0.131
$(\cdot+1/(2n))^2$	$(1-\hat{F}_n)^2$	0.300	0.125	0.293	0.142
$(\cdot+1/c_n)^2$	$(1-\hat{F}_n)^{1.1}$	0.300	0.113	0.307	0.132

Table 7: Estimation of the mixture parameter $\pi = 0.3$ based on distance functions of the type G_1 and on $n = 40$ observations from contamination model CM_1. The values given in the table are based on 10,000 simulations.

Censoring		0%		40%	
$G_n(\cdot)$	w_n	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
$(\cdot)^2$	1	0.245	0.099	0.224	0.115
$(\cdot+1/(2n))^2$	1	0.267	0.099	0.245	0.116
$(\cdot+1/n)^2$	1	0.289	0.099	0.267	0.117
$(\cdot+1/c_n)^2$	1	0.267	0.099	0.259	0.118
$(\cdot)^2$	$(1-\hat{F}_n)^2$	0.225	0.124	0.217	0.137
$(\cdot)^2$	$(1-\hat{F}_n)^4$	0.203	0.142	0.200	0.152
$(\cdot+1/(2n))^2$	$(1-\hat{F}_n)^{1.1}$	0.258	0.113	0.246	0.129
$(\cdot+1/(2n))^2$	$(1-\hat{F}_n)^2$	0.250	0.125	0.241	0.139
$(\cdot+1/c_n)^2$	$(1-\hat{F}_n)^{1.1}$	0.259	0.113	0.261	0.132

Table 8: Estimation of the mixture parameter $\pi = 0.3$ based on distance functions of the type G_1 and on $n = 40$ observations from contamination model CM_2. The

values given in the table are based on 10,000 simulations. The 40% censoring is with respect to the Weibull mixture. The overall censoring being less.

Censoring		0%		40%	
$G_n(\cdot)$	w_n	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
$(\cdot)^2$	1	0.321	0.097	0.299	0.117
$(\cdot + 1/(2n))^2$	1	0.343	0.097	0.320	0.117
$(\cdot + 1/n)^2$	1	0.365	0.096	0.342	0.117
$(\cdot + 1/c_n)^2$	1	0.343	0.097	0.337	0.119
$(\cdot)^2$	$(1 - \hat{F}_n)^2$	0.317	0.119	0.307	0.136
$(\cdot)^2$	$(1 - \hat{F}_n)^4$	0.314	0.142	0.308	0.157
$(\cdot + 1/(2n))^2$	$(1 - \hat{F}_n)^{1.1}$	0.342	0.107	0.329	0.126
$(\cdot + 1/(2n))^2$	$(1 - \hat{F}_n)^2$	0.344	0.118	0.333	0.136
$(\cdot + 1/c_n)^2$	$(1 - \hat{F}_n)^{1.1}$	0.343	0.107	0.347	0.128

Table 9: Estimation of the mixture parameter $\pi = 0.3$ *based on distance functions of the type* G_1 *and on* $n = 40$ *observations from contamination model* CM_3. *The values given in the table are based on 10,000 simulations. The 40% censoring is with respect to the Weibull mixture. The overall censoring being larger.*

5 Concluding remarks

In this paper we studied a new class of minimum distance estimators for parametric models. An advantage of these estimators is that the minimum distance function, which is an extended empirical version of the generalized weighted Cramér-von Mises distance, can be obtained in closed form and accounts for right censoring. Furthermore, as indicated by a simulation study, allowing the distance function to depend on the number of censored items reduces the bias considerably without deteriorating the standard deviation of the estimates. Robustness properties of our estimators are established and checked numerically.

References

Andersen, P. K., Borgan, O., Gill, R. & Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer, New York.

Atienza, N., Garcia-Heras, J. & Muñoz-Pichardo, J. M. (2006). A new condition for identifiability of finite mixture distributions. *Metrika* **63**, 215-221.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5**, 445–463.

Beran, R. (1984). Minimum distance procedure. In *Handbook of Statistics, Vol. 4* (eds Krishnaian, P. R., Sen, P. K.), 741–754. Elsevier Science. Amsterdam.

Choi, K. & Bulgren, W. B. (1968). An estimation procedure for mixtures of distributions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **30**, 444-460.

Cutler, A. & Cordero-Brara, O. I. (1996). Minimum Hellinger distance estimation for finite mixture models. *J. Amer. Statist. Assoc.* **91**, 1716–1723.

Ferland, R., Froda, S. & Lavigne, J. (2003). A simulation study of a minimum distance estimator for finite mixtures under censoring. In *Mathematical Statistics and Applications: Festschrift for Constance van Eeden* (eds M. Moore, S. Froda, and C. Léger), 483–495. Institute of Mathematical Statistics.

Fleming, T. R. & Harrington, D. P. (2005). *Counting Processes and Survival Analysis.* Wiley, New York.

Hettmansperger, T. P., Hueter, I. & Hüsler, J. (1994). Minimum distance estimators. *J. Statist. Plann. Inference* **41**, 291–302.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation for incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457–481.

Lindsay, B. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *Ann. Statist.* **22**, 1081–1114.

MacDonald, P. D. M. (1971). Comment on "An estimation procedure for mixtures of distributions" by Choi and Bulgren. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **33**, 326-329.

McLachlan, G. J. & Peel, D. (2000). *Finite Mixture Models.* Wiley-Interscience, New York.

Öztürk, Ö. & Hettmansperger, T. P. (1997). Generalised Weighted Cramér-von Mises Distance Estimators. *Biometrika* **84**, 283-294.

Pardo, M. C. (1997). A comparison of some estimators of the mixture proportion of mixed normal distributions. *J. Comput. Appl. Math.* **84**, 207–217.

Parr, W. C. & Schucany, W. R. (1980). Minimum distance and robust estimation. *J. Amer. Statist. Assoc.* **75**, 616–624.

Reid, N. (1981). Influence functions for censored data. *Ann. Statist.* **9**, 78–92.

Stute, W. & Wang, J.-L. (1993). The Strong Law under Random Censorship. *Ann. Statist.* **21**, 1591–1607.

Teicher, H. (1963). Identifiability of finite mixtures. *Ann. Math. Statist.* **34**, 1265-1269.

Titterington, D. M., Smith, A. F. M. & Makov, U. E. (1985). Statistical Analysis of Finite Mixture Distributions. John Wiley, Chichester.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, New York.

Wolfowitz, J. (1957). The minimum distance method. *Ann. Math. Statist.* **28**, 75–88.

Woodward, W. A., Parr, W. C., Schucany, W. R. & Lindsey, H. (1984). A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *J. Amer. Statist. Assoc.* **79**, 590-598.

Woodward, W. A., Whitney, P. & Eslinger, P. W. (1995). Minimum Hellinger distance estimation for mixture proportions. *J. Statist. Plann. Inference* **48**, 303–319.

Yang, S. (1991). Minimum Hellinger distance estimation of parameter in the random censorship model. *Ann. Statist.* **19**, 579–602.

Ying, Z. (1992). Minimum Hellinger-Type Distance Estimation for Censored Data. *Ann. Statist.* **20**, 1361–1390.

Eric Beutner, Department of Quantitative Economics, Maastricht University, P.O.Box 616, NL-6200 MD Maastricht, The Netherlands
e-mail: e.beutner@ke.unimaas.nl

Laurent Bordes, Laboratoire de Mathématiques et de leurs Applications, Université de Pau et des Pays de l'Adour, Avenue de l'Université, Bât. IPRA, B.P 1155, F-61013 Pau Cedex, France
e-mail: laurent.bordes@univ-pau.fr

Appendix

Proof of Lemma 1. It follows from Condition C that G has finite variations on $[-1,1]$ and then is the difference of two bounded increasing functions G^+ and G^-. Now, let $\mathcal{W}=\{F_\theta;\theta\in\Theta\}$. Then $\mathcal{W}=\{F_\theta;\theta\in\Theta\}\subset\mathcal{M}$ where $\mathcal{M}$ is the set of monotone increasing functions. The class $\mathcal{M}$ has a finite bracketing number (see van der Vaart and Wellner (1996, Theorem 2.7.5)), and hence $\mathcal{W}'=\{F_\theta-F_{\theta_0};\theta\in\Theta\}$ has a finite bracketing number. Obviously, given ε–brackets $[l_i,u_i]$, $i=1,\dots,m$, covering $\mathcal{W}'$, the brackets $[w_0G^+\circ l_i, w_0G^+\circ u_i]$, $i=1,\dots,m$, cover $G^+\circ\mathcal{W}'=\{w_0G^+\circ z; z\in\mathcal{W}'\}$ and have L^1 size proportional to ε. Applying the same argument to $G^-\circ\mathcal{W}'=\{G^-\circ z; z\in\mathcal{W}'\}$ the result follows as $\mathcal{Z}\subset G^+\circ\mathcal{W}'-G^-\circ\mathcal{W}'$ and as $G^+\circ\mathcal{W}'-G^-\circ\mathcal{W}'$ is again $\mathbb{P}$–Glivenko-Cantelli (see van der Vaart and Wellner (1996, p. 125)). □

Proof of Lemma 2. From Condition E and Theorem 1 we have $\sup_{[0,\tau]}|F_{\hat\theta_n}-F_{\theta_0}|=o_P(1)$. By a second order expansion of G around 0 we then obtain

$$\begin{aligned}&n\int_0^\tau G\left(F_{\hat\theta_n}-F_{\theta_0}+o_p(1/\sqrt{n})\right)w_n d\hat F_n\\ =&\ \frac{n}{2}\int_0^\tau G''\left(H_{\hat\theta_n,\theta_0}\right)\left(F_{\hat\theta_n}-F_{\theta_0}+o_p(1/\sqrt{n})\right)^2 w_n d\hat F_n,\end{aligned}$$

where $H_{\hat\theta_n,\theta_0}$ satisfies $0\le|H_{\hat\theta_n,\theta_0}|\le|F_{\hat\theta_n}-F_{\theta_0}+o_p(1/\sqrt{n})|$. Notice that by using Condition E we now have

$$\begin{aligned}&\frac{n}{2}\int_0^\tau G_n''\left(H_{\hat\theta_n,\theta_0}\right)\left(F_{\hat\theta_n}-F_{\theta_0}+o_p(1/\sqrt{n})\right)^2 w_n d\hat F_n\\ =&\ \frac{n}{2}\int_0^\tau G_n''\left(H_{\hat\theta_n,\theta_0}\right)\left(\left(\hat\theta_n-\theta_0\right)^t\eta\circ F_{\theta_0}+o\left(\|\hat\theta_n-\theta_0\|\right)+o_p(1/\sqrt{n})\right)^2 w_n d\hat F_n,\end{aligned}$$

where the right hand side is equal to

$$\sqrt{\frac{n}{2}}\left(\hat{\theta}_n-\theta_0\right)^t \times \left[\int_0^\tau G_n''\left(H_{\hat{\theta}_n,\theta_0}\right)(\eta\circ F_{\theta_0})(\eta\circ F_{\theta_0})^t w_n d\hat{F}_n\right.$$
$$+2\int_0^\tau G_n''\left(H_{\hat{\theta}_n,\theta_0}\right)\eta\circ F_{\theta_0}w_n d\hat{F}_n\frac{(\hat{\theta}_n-\theta_0)^t o(\|\hat{\theta}_n-\theta_0\|)}{\|\hat{\theta}_n-\theta_0\|^2}$$
$$+\int_0^\tau G_n''\left(H_{\hat{\theta}_n,\theta_0}\right)w_n d\hat{F}_n\left(\frac{o(\|\hat{\theta}_n-\theta_0\|)}{\|\hat{\theta}_n-\theta_0\|}\right)^2$$
$$\left.+2\int_0^\tau G_n''\left(H_{\hat{\theta}_n,\theta_0}\right)o_p(1/\sqrt{n})\frac{\eta\circ F_{\theta_0}(\hat{\theta}_n-\theta_0)^t+o(\|\hat{\theta}_n-\theta_0\|)}{\|\hat{\theta}_n-\theta_0\|^2}w_n d\hat{F}_n\right]$$
$$\times\sqrt{\frac{n}{2}}\left(\hat{\theta}_n-\theta_0\right)+o_P(1)\int_0^\tau G_n''\left(H_{\hat{\theta}_n,\theta_0}\right)w_n d\hat{F}_n.$$

From Conditions B, C, and E, Theorem 1, and properties of the Kaplan-Meier estimator we obtain that the term within brackets converges to the positive definite matrix $G''(0)\Sigma(\tau)$ and that $o_P(1)\int_0^\tau G_n''\left(H_{\hat{\theta}_n,\theta_0}\right)w_n d\hat{F}_n$ converges to zero in probability. Then we have

$$\begin{aligned}&\sqrt{\frac{n}{2}}(\hat{\theta}_n-\theta_0)^t\left[G''(0)\Sigma(\tau)+o_P(1)\right]\sqrt{\frac{n}{2}}(\hat{\theta}_n-\theta_0)\\ =\ & n\int_0^\tau G\left(F_{\hat{\theta}_n}-F_{\theta_0}+o_p(1/\sqrt{n})\right)w_n d\hat{F}_n. \end{aligned}\tag{17}$$

Let us now show that $A_n\equiv n\int_0^\tau G\left(F_{\hat{\theta}_n}-F_{\theta_0}+o_p(1/\sqrt{n})\right)w_n d\hat{F}_n$ is bounded in probability from which, using (17), the assertion follows. Let $\varepsilon>0$ be a real number. By Condition C there exists a neighborhood V of 0 and constants $0<\alpha<\beta<+\infty$ such that $\alpha x^2\le G(x)\le\beta x^2$ on V. Hence, on V we have

$$G(x+y)\le 2\beta(G(x)/\alpha)+2\beta y^2. \tag{18}$$

From Condition E and Theorem 1 we have $\sup_{[0,\tau]}|F_{\hat{\theta}_n}-F_{\theta_0}|=o_P(1)$ and by properties of the Kaplan-Meier estimator we have for $n\ge n_0$ both $\sup_{[0,\tau]}|\hat{F}_n-F_{\theta_0}|$ and $\sup_{[0,\tau]}|F_{\hat{\theta}_n}-F_{\theta_0}+o_p(1/\sqrt{n})|$ in V with probability more than $1-\varepsilon$.

It follows that

$$
\begin{aligned}
& 1-\varepsilon \\
\overset{(18)}{\leq} \; & \mathbb{P}\left(A_n \leq 2n\beta \int_0^\tau \left[(1/\alpha)G(\hat{F}_n - F_{\hat{\theta}_n} + o_p(1/\sqrt{n})) + (\hat{F}_n - F_{\theta_0})^2\right] w_n d\hat{F}_n\right) \\
\overset{\text{def. } \hat{\theta}_n}{\leq} \; & \mathbb{P}\left(A_n \leq 2n\beta \int_0^\tau \left[(1/\alpha)G(\hat{F}_n - F_{\theta_0} + o_p(1/\sqrt{n})) + (\hat{F}_n - F_{\theta_0})^2\right] w_n d\hat{F}_n\right) \\
\leq \; & \mathbb{P}\left(A_n \leq 2n\beta \int_0^\tau \left[(\beta/\alpha)(\hat{F}_n - F_{\theta_0} + o_p(1/\sqrt{n}))^2 + (\hat{F}_n - F_{\theta_0})^2\right] w_n d\hat{F}_n\right) \\
\leq \; & \mathbb{P}\left(A_n \leq B_n\right),
\end{aligned}
$$

where, B_n is equal to

$$
2\beta\left[\left(\frac{\beta}{\alpha}+1\right)\left(\sup_{[0,\tau]}\left|\sqrt{n}\left(\hat{F}_n - F_{\theta_0}\right)\right|\right)^2 + 2\sup_{[0,\tau]}|o_P(1)\sqrt{n}(\hat{F}_n - F_{\theta_0})| + o_P(1)\right]
$$
$$
\times \sup_{[0,\tau]}|w_n|.
$$

Under Condition B, by the continuous mapping theorem and because $\sqrt{n}(\hat{F}_n - F_{\theta_0})$ converges weakly to a Gaussian process in $D[0,\tau]$, the sequence B_n is a $O_P(1)$. By this with the above inequality we obtain that $A_n = O_P(1)$ and since $\Sigma(\tau)$ is positive definite we have by (17) that $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_P(1)$. □

Proof of Theorem 3 and Lemma 3. All results follow by differentiating the right hand side in the following equation with respect to ε at $\varepsilon = 0$ and using that $G(0) = G'(0) = 0$.

$$
\begin{aligned}
0 \;=\; & \int_0^\tau G'\left((1-\varepsilon)H + \varepsilon\Delta_x - F_{T((1-\varepsilon)H+\varepsilon\Delta_x)} + o_P(1/\sqrt{n})\right) \\
& \times \dot{F}_{\theta|_{\theta=T((1-\varepsilon)H+\varepsilon\Delta_x)}} w_n((1-\varepsilon)H + \varepsilon\Delta_x)\, d((1-\varepsilon)H + \varepsilon\Delta_x).
\end{aligned}
$$

□