I N S T I T U T D E

S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N

P A P E R

0037

INFERENCE ON MULTIVARIATE M-ESTIMATORS BASED ON BIVARIATE CENSORED DATA

I. VAN KEILEGOM and T.P. HETTMANSPERGER

http://www.stat.ucl.ac.be

Inference on multivariate M-estimators based on bivariate censored data

I. Van Keilegom[1] T. P. Hettmansperger

Université Catholique de Louvain *Penn State University*

December 5, 2000

Abstract

Consider two random variables subject to random right censoring, like the survival times of twins or two recurrence times of a certain disease recorded on the same individual. Of interest is the estimation of a multivariate M-functional of these two random variables.

Asymptotic results for the proposed estimator are established under general conditions on the criterion function of the M-estimator. The obtained results include the asymptotic normality of the estimator, its local power and the construction of a confidence region. Our primary example of interest is the bivariate L_1 median. We compare the small sample behavior of this estimator with that of the vector of marginal medians by means of a simulation study. Finally, we consider a data set on kidney dialysis patients and estimate the median time to two different infections for these individuals.

KEY WORDS: Asymptotic normality; Bootstrap; Confidence region; L_1 median; M-estimation; Right censoring.

[1] This research was supported by 'Projet d'Actions de Recherche Concertées', No. 98/03–217, from the Belgian government.

1 Introduction

In this paper we consider the concept of multivariate M-estimators when the data consist of pairs of possibly right censored observations. These kind of data arise naturally in e.g. twins studies, studies of the eyes, studies where two recurrence times of a certain disease are recorded, or where the onset of a disease and death are of interest. Since for censored observations, one will never have complete data in the tail of the distribution, it is impossible to estimate the bivariate mean in this context. This motivates the study of M-estimators, since they include, among others, alternative location estimators, like e.g. the L_1 median which will be our primary example.

Before defining M-estimators for censored data, let us consider the case where the data are completely observed. The d-dimensional vector $\boldsymbol{\theta}$ is defined implicitly by

$$\int\int \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})\, dF(\boldsymbol{y}) = \mathbf{0}, \tag{1.1}$$

where $F(\boldsymbol{y})$ is a bivariate distribution function. It is assumed that $\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}) \in I\!R^d$ satisfies the regularity conditions given by Huber (1981) in Case B on pages 130-132. These conditions imply the consistency and asymptotic normality of the estimator defined next. The corresponding M-estimator $\hat{\boldsymbol{\theta}}$ is the solution of

$$\int\int \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})\, d\hat{F}(\boldsymbol{y}) = \mathbf{0}, \tag{1.2}$$

where $\hat{F}(\boldsymbol{y})$ is the bivariate empirical distribution function. Our primary example and application is based on

$$\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}) = \frac{\boldsymbol{y} - \boldsymbol{\theta}}{\|\boldsymbol{y} - \boldsymbol{\theta}\|}.$$

The resulting parameter is the L_1 median described in detail by Small (1990) in a survey of multivariate medians. Hence, for uncensored data, $\hat{\boldsymbol{\theta}}$ solves

$$\sum_{i=1}^{n} \frac{\boldsymbol{y}_i - \boldsymbol{\theta}}{\|\boldsymbol{y}_i - \boldsymbol{\theta}\|} = \mathbf{0}$$

and minimizes the average distance from the data points to $\boldsymbol{\theta}$, i.e. $\hat{\boldsymbol{\theta}}$ minimizes $\sum \|\boldsymbol{y}_i - \boldsymbol{\theta}\|$. The L_1 median is equivariant under orthogonal transformations of the data but not under affine transformations. For uncensored data, a number of affine equivariant estimators of the multivariate median have been proposed. However, we prefer here to work with the

L_1 median, since in the present context of censored data, the variables of interest do have some physical interpretation (e.g. survival times of twins) and hence one is not interested in transforming these variables by some affine transformation. In addition, the L_1 median has the attractive property that it identifies an easy to interpret location vector even for asymmetric distributions, and in both the uncensored and the censored data cases, if the components are rescaled by the same factor, the L_1 median is unchanged. If we consider more general M-estimators then, similar to the univariate case, we must have a supplementary estimate of scale or assume the scale is known. See Maronna (1976) and Hampel et al. (1986) for a detailed development of multivariate M-estimators in the uncensored case. The L_1 median has 50 percent breakdown and bounded influence; see Lopuhaä and Rousseeuw (1991) and Kemperman (1987). It is unique in dimensions greater than 1. Huber (1981) developed the asymptotic normality in Example 3.1 in Section 6.3. Brown (1983) showed that for spherical normal models, the efficiency increases as the dimension increases beginning with a bivariate efficiency of .785 already larger than the univariate normal efficiency of .687. Chaudhuri (1992) provided a Bahadur representation. Möttönen and Oja (1995) developed multivariate spatial sign and rank methods based on the L_1 criterion function, and Choi and Marden (1997) developed multivariate rank tests for analysis of variance.

It is the purpose of this paper to extend M-estimators to the case of censored data. Let $\boldsymbol{T}$ be a bivariate lifetime vector with bivariate distribution function $F(\boldsymbol{y}) = P(T_1 \leq y_1, T_2 \leq y_2)$ and let $\boldsymbol{C}$ be a bivariate censoring vector with distribution $G(\boldsymbol{y})$. Assume that $\boldsymbol{T}$ and $\boldsymbol{C}$ are independent. Let $(\boldsymbol{T}_i, \boldsymbol{C}_i)$ $(i = 1, \ldots, n)$ be independent copies of $(\boldsymbol{T}, \boldsymbol{C})$. We observe $(\tilde{\boldsymbol{T}}_i, \boldsymbol{\Delta}_i)$ $(i = 1, \ldots, n)$ where $\tilde{T}_{ij} = T_{ij} \wedge C_{ij}$, $\Delta_{ij} = I(T_{ij} \leq C_{ij})$ $(j = 1, 2)$ and $\tilde{\boldsymbol{T}}_i$ has bivariate distribution function $H(\boldsymbol{y})$. Note that the observations can be doubly uncensored, singly censored in either component, or doubly censored.

We extend the estimating equation (1.2) of $\boldsymbol{\theta}$ to the case of censored data, by replacing the bivariate empirical distribution function by a suitable estimator of the bivariate distribution when both components are subject to right censoring. Nonparametric estimators $\hat{F}(\boldsymbol{y})$ have been proposed by (among others) Dabrowska (1988), Pruitt (1991), Prentice and Cai (1992), van der Laan (1996) and Akritas and Van Keilegom (2000), while e.g. Oakes (1989) and Wang and Wells (1999) proposed semiparametric estimators for $F(\boldsymbol{y})$. In Section 4, we will consider a subset of these estimators in more detail. Many of these estimators $\hat{F}(\boldsymbol{y})$ are restricted to the case where the survival and censoring times are non-negative. If $\boldsymbol{T}$ and $\boldsymbol{C}$ are transformed survival and censoring times not necessarily on $[0,\infty) \times [0,\infty)$ (e.g. a log transformation of the raw times), then

$$F(\boldsymbol{y}) = P(T_1 \leq y_1, T_2 \leq y_2) = P(V_1 \leq v(y_1), V_2 \leq v(y_2)),$$

where $V_j = v(T_j)$ $(j = 1, 2)$ are the raw survival times (starting at zero) for some monotone, increasing transformation v. We define $\hat{F}(\boldsymbol{y})$ in this case as the estimator for the bivariate distribution of (V_1, V_2) and evaluated at $v(y_1)$ and $v(y_2)$. The results we will develop can be viewed as a generalization of James (1986) and Wang (1999), where M-estimation with univariate censored data is considered.

In Section 2, we state our main results on the limit distributions of the bivariate M-estimators. In Section 3, we discuss a bootstrap approach to the estimation of the asymptotic covariance structure of the M-estimators. Section 4 discusses a number of estimators of the bivariate distribution that can be used in the construction of the M-estimator. Simulation results are established in Section 5, while an example is analyzed in Section 6. Proofs and derivations are given in the Appendix.

2 Main results

Because the asymptotic theory for $\hat{\boldsymbol{\theta}}$ will be based on an i.i.d. representation for an estimator $\hat{F}(\boldsymbol{y})$ of $F(\boldsymbol{y})$ which will turn out to be valid for any $\boldsymbol{y} = (y_1, y_2)$ in a certain region Ω (where Ω depends on the choice of the estimator $\hat{F}(\boldsymbol{y})$ and will be discussed later), we need to work with a slightly modified version of (1.1) and (1.2) :

$$\boldsymbol{S}(\boldsymbol{\theta}) = n \int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})\, dF(\boldsymbol{y}) \tag{2.1}$$

$$\boldsymbol{S}_n(\boldsymbol{\theta}) = n \int_{\hat{\Omega}} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})\, d\hat{F}(\boldsymbol{y}), \tag{2.2}$$

where $\hat{\Omega}$ is an estimator for the region Ω. The true M-functional is denoted by $\boldsymbol{\theta}_0 \in I\!R^d$ and is the value of $\boldsymbol{\theta}$ for which $\boldsymbol{S}(\boldsymbol{\theta}) = \boldsymbol{0}$.

The main results require a number of regularity conditions which we mention below for convenient reference.

(A1) (i) The functions $f(\boldsymbol{y})$ and $\boldsymbol{\psi}(\boldsymbol{y},\boldsymbol{\theta}_0)$ are uniformly bounded for $\boldsymbol{y} \in I\!R^2$.

(ii) The function $\boldsymbol{\psi}(\boldsymbol{y},\boldsymbol{\theta})$ is differentiable with respect to the components of $\boldsymbol{\theta}$, for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$.

(iii) For $1 \leq i, j \leq d$,

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\delta_n} \int_{\Omega} \left| \frac{\partial}{\partial\theta_j}\psi_i(\boldsymbol{y},\boldsymbol{\theta}) - \frac{\partial}{\partial\theta_j}\psi_i(\boldsymbol{y},\boldsymbol{\theta}_0)\right| dF(\boldsymbol{y}) = o_P(1)$$

for any sequence $\delta_n \to 0$ as $n \to \infty$.

(A2) (i) The matrix $\boldsymbol{A}$, defined in Theorem 2.1, is positive definite.

(ii) The matrix $\boldsymbol{B}$, defined in Theorem 2.2, is positive definite.

(iii) The matrix of derivatives of $\boldsymbol{S}_n(\boldsymbol{\theta})$ is negative definite.

Further, the estimators $\hat{F}(\boldsymbol{y})$ and $\hat{\Omega}$ and the function $\boldsymbol{\psi}(\boldsymbol{y},\boldsymbol{\theta}_0)$ need to satisfy the following conditions :

(A3) (i) The estimator $\hat{F}(\boldsymbol{y})$ admits the following representation :

$$\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}) = n^{-1}\sum_{i=1}^{n} g(\tilde{\boldsymbol{T}}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}) + R_n(\boldsymbol{y}), \tag{2.3}$$

for a certain function $g : I\!R^2 \times \{0,1\}^2 \times I\!R^2 \to I\!R$ and where

$$\int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y},\boldsymbol{\theta}_0)\, dR_n(\boldsymbol{y}) = \boldsymbol{o}_P(n^{-1/2}) \tag{2.4}$$

(ii) The region $\hat{\Omega}$ satisfies

$$\begin{aligned}&\int_{\hat{\Omega}} \boldsymbol{\psi}(\boldsymbol{y},\boldsymbol{\theta}_0)\, dF(\boldsymbol{y}) - \int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y},\boldsymbol{\theta}_0)\, dF(\boldsymbol{y})\\ &= n^{-1}\sum_{i=1}^{n} \boldsymbol{k}(\tilde{\boldsymbol{T}}_i, \boldsymbol{\Delta}_i, \boldsymbol{\theta}_0) + \boldsymbol{o}_P(n^{-1/2})\end{aligned} \tag{2.5}$$

for a certain function $\boldsymbol{k} : I\!R^2 \times \{0,1\}^2 \times I\!R^d \to I\!R^2$.

(iii)

$$\begin{aligned}&\int_{\hat{\Omega}} \boldsymbol{\psi}(\boldsymbol{y},\boldsymbol{\theta}_0)\, d(\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}))\\ &\quad - \int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y},\boldsymbol{\theta}_0)\, d(\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y})) = \boldsymbol{o}_P(n^{-1/2}),\end{aligned} \tag{2.6}$$

and

$$\int_{\hat{\Omega}} [\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0 + \boldsymbol{b} n^{-1/2}) - \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)] \, d\hat{F}(\boldsymbol{y}) - \int_{\Omega} [\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0 + \boldsymbol{b} n^{-1/2}) - \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)] \, dF(\boldsymbol{y}) = \boldsymbol{o}_P(n^{-1/2}) \tag{2.7}$$

for any vector $\boldsymbol{b} \in I\!R^d$.

In the appendix we will verify condition $(A3)$ for a number of choices for $\hat{F}(\boldsymbol{y})$, $\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)$ and $\hat{\Omega}$. Assumption $(A3)(i)$ is a key-condition for proving the asymptotics of the proposed M-estimators. Using this condition, we will be able to linearize the estimating equation $\boldsymbol{S}_n(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$. This linearization can then be carried over to $\hat{\boldsymbol{\theta}}$, from which the main results will follow.

We start with the asymptotic normality of $\boldsymbol{S}_n(\boldsymbol{\theta}_0)$. This result is useful for testing hypotheses concerning $\boldsymbol{\theta}_0$.

Theorem 2.1 *Assume* $(A1)(i), (A2)(i), (A3)$*. Then,*

$$n^{-1/2} \boldsymbol{S}_n(\boldsymbol{\theta}_0) \xrightarrow{d} N_d(\boldsymbol{0}, \boldsymbol{A}),$$

where

$$\boldsymbol{A} = Var\left[\int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0) \, dg(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \boldsymbol{y}) + \boldsymbol{k}(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \boldsymbol{\theta}_0) \right].$$

The next result is the asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}$.

Theorem 2.2 *Assume* $(A1) - (A3)$*. Then,*

$$n^{1/2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_d(\boldsymbol{0}, \boldsymbol{B}^{-1} \boldsymbol{A} (\boldsymbol{B}^t)^{-1}),$$

where $\boldsymbol{B} = (B_{ij})$ *$(i, j = 1, \ldots, d)$ and*

$$B_{ij} = - \int_{\Omega} \frac{\partial}{\partial \theta_j} \psi_i(\boldsymbol{y}, \boldsymbol{\theta}_0) \, dF(\boldsymbol{y}).$$

In Section 3, we discuss the estimation of $\boldsymbol{A}$ and $\boldsymbol{B}$. These estimates are needed to estimate the covariance matrix of $\hat{\boldsymbol{\theta}}$ and, hence, estimate the standard errors. In addition, the results in Theorem 2.2 can be used to calculate the estimation efficiency of $\hat{\boldsymbol{\theta}}$.

The following result holds for the special case of a location functional (i.e. $\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{\psi}_{\boldsymbol{0}}(\boldsymbol{y} - \boldsymbol{\theta})$ for some function $\boldsymbol{\psi}_{\boldsymbol{0}}$) and will be stated separately for location and scale models. Location models are appropriate when modelling logarithm of survival time and scale models for raw survival time. It should be noted that there is no simple transformation that allows one to switch between the location estimator of the log and the raw survival times.

Location model : $F(\boldsymbol{y}) = F_{\boldsymbol{0}}(y_1 - \theta_1, y_2 - \theta_2)$ for some prototype distribution $F_{\boldsymbol{0}}(y_1, y_2)$.
Scale model : $F(\boldsymbol{y}) = F_{\boldsymbol{1}}(\frac{y_1}{\sigma_1}, \frac{y_2}{\sigma_2})$ for $\sigma_1, \sigma_2 > 0$ and for some prototype distribution $F_{\boldsymbol{1}}(y_1, y_2)$ satisfying $F_{\boldsymbol{1}}(\boldsymbol{y}) = 0$ for $y_1 < 0$ or $y_2 < 0$.

Theorem 2.3 *Assume* $(A1) - (A3)$ *and let* $\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{\psi}_{\boldsymbol{0}}(\boldsymbol{y} - \boldsymbol{\theta})$ *for some function* $\boldsymbol{\psi}_{\boldsymbol{0}}$ *(i.e.* $\boldsymbol{\theta}_0$ *is a location functional).*

1. **Location model :** *Let* $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ *and* $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{\gamma} n^{-1/2}$ *for some fixed vector* $\boldsymbol{\gamma}$. *Then, under* H_1,

$$n^{-1} \boldsymbol{S}_n^t(\boldsymbol{\theta}_0) \boldsymbol{A}^{-1} \boldsymbol{S}_n(\boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(d, \boldsymbol{\gamma}^t \boldsymbol{B}^t \boldsymbol{A}^{-1} \boldsymbol{B} \boldsymbol{\gamma}),$$

where $\chi^2(a, b)$ *is a non-central chi-square distribution with* a *degrees of freedom and non-centrality parameter* b.

2. **Scale model :** *Let* $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ *and* $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 e^{\gamma n^{-1/2}}$ *for some fixed value* γ, *and assume that* $\boldsymbol{\psi}_{\boldsymbol{0}}(a\boldsymbol{y}) = \boldsymbol{\psi}_{\boldsymbol{0}}(\boldsymbol{y})$ *for all* $\boldsymbol{y}$ *and all* $a > 0$. *Then, under* H_1,

$$n^{-1} \boldsymbol{S}_n^t(\boldsymbol{\theta}_0) \boldsymbol{A}^{-1} \boldsymbol{S}_n(\boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(d, \gamma^2 \boldsymbol{\theta}_0^t \boldsymbol{B}^t \boldsymbol{A}^{-1} \boldsymbol{B} \boldsymbol{\theta}_0).$$

Theorem 2.3 can be used to estimate the local power of the test based on $\boldsymbol{S}_n(\boldsymbol{\theta}_0)$. The test efficiency of two tests concerning $\boldsymbol{\theta}$, is defined as the limiting ratio of the samples sizes needed for the same asymptotic level and same asymptotic power along the same sequence of alternatives, or equivalently as the ratio of the noncentrality parameters of the limiting χ^2-distributions of the respective test statistics under the alternative hypothesis (see Bickel (1965), Hettmansperger and McKean (1998)). Hence, the above result enables us to calculate the test efficiency of $\boldsymbol{S}_n$ relative to another test. The calculations, however, are computationally intensive, so in Section 5 we report on a finite sample simulation study comparing different estimates of location.

We end this section with the construction of a confidence region for $\boldsymbol{\theta}_0$.

Theorem 2.4 *Assume* (A1)−(A3)*. Further, let* $\hat{\boldsymbol{A}}$ *and* $\hat{\boldsymbol{B}}$ *be weakly consistent estimators of respectively* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*. Then, a* $(1-\alpha)100\%$ *confidence region for* $\boldsymbol{\theta}_0$ *is given by the values of* $\boldsymbol{\theta}_0$ *that satisfy*

$$n(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^t\hat{\boldsymbol{B}}^t\hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{B}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\leq\chi^2_\alpha(d),$$

where $\chi^2_\alpha(d)$ *satisfies* $P(X\geq\chi^2_\alpha(d))=\alpha$ *if* $X\sim\chi^2(d)$*.*

The proof is a straightforward application of Lemma A.1 and Theorem 2.2 and is therefore omitted.

The calculation of standard errors, local power and the construction of a confidence region for $\boldsymbol{\theta}_0$ require the estimation of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The estimation of the matrix $\boldsymbol{B}$ is straightforward (simply replace $F(\boldsymbol{y})$, Ω and $\boldsymbol{\theta}_0$ by consistent estimators $\hat{F}(\boldsymbol{y})$, $\hat{\Omega}$ and $\hat{\boldsymbol{\theta}}$ respectively). More attention needs to be paid to the estimation of $\boldsymbol{A}$. An estimator for $\boldsymbol{A}$ could be obtained by replacing the functions $g(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\boldsymbol{y})$ and $\boldsymbol{k}(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\boldsymbol{\theta}_0)$ by appropriate estimators. However, the estimation of these functions is often complicated. For this reason we define in the next section a bootstrap procedure which can be used to obtain a bootstrap estimate for $\boldsymbol{A}$.

3 Bootstrap for two dimensional censored data

The purpose of this section is to define an appropriate bootstrap scheme that will allow us to estimate the matrix $\boldsymbol{A}$ in a consistent way.

We start with extending the so-called 'obvious' bootstrap (see Efron (1981)) from one to two dimensions. Conditioning on the responses $\boldsymbol{T}_i$ and censoring times $\boldsymbol{C}_i$ $(i=1,\ldots,n)$ we define the random variables $\boldsymbol{T}_i^*$ and $\boldsymbol{C}_i^*$ (independently) as follows :

$$\begin{array}{l}\boldsymbol{T}_1^*,\ldots,\boldsymbol{T}_n^* \text{ are independent; } \boldsymbol{T}_i^*\sim\hat{F}\\ \boldsymbol{C}_1^*,\ldots,\boldsymbol{C}_n^* \text{ are independent; } \boldsymbol{C}_i^*\sim\hat{G},\end{array}$$

where $\hat{G}(\boldsymbol{y})$ is obtained by interchanging the role of the survival and censoring times in the definition of $\hat{F}(\boldsymbol{y})$. Then define, for $i=1,\ldots,n$ and $j=1,2$,

$$\tilde{T}^*_{ij}=\min(T^*_{ij},C^*_{ij}) \text{ and } \Delta^*_{ij}=I(T^*_{ij}\leq C^*_{ij}).$$

The natural extension of the 'simple' bootstrap for one dimensional data to two dimensions consists in drawing pairs $(\tilde{\boldsymbol{T}}_i^*, \boldsymbol{\Delta}_i^*)$ with replacement from $(\tilde{\boldsymbol{T}}_1, \boldsymbol{\Delta}_1), \ldots, (\tilde{\boldsymbol{T}}_n, \boldsymbol{\Delta}_n)$.

Unlike the one-dimensional case, we now show that, regardless of the choice of the estimator $\hat{F}$, these two bootstrap procedures cannot be equivalent in two dimensions. Indeed, it is easily seen that two necessary conditions for equivalence are that (1) $\hat{F}^c(\boldsymbol{y})\hat{G}^c(\boldsymbol{y}) = \hat{H}^c(\boldsymbol{y})$ (where $\hat{F}^c(\boldsymbol{y}) = 1 - \hat{F}(y_1, \infty) - \hat{F}(\infty, y_2) + \hat{F}(\boldsymbol{y})$ is the estimator for $P(T_1 > y_1, T_2 > y_2)$ and similarly for $\hat{G}^c$ and $\hat{H}^c$ (replace T_j $(j = 1, 2)$ respectively by C_j and $\tilde{T}_j$)) and (2) the support points of $\hat{F}$ (respectively $\hat{G}$) are the doubly uncensored (respectively doubly censored) observations. But, if condition (2) is satisfied, (1) cannot hold, since $\hat{F}^c(\boldsymbol{y})\hat{G}^c(\boldsymbol{y})$ would then jump at the doubly uncensored and doubly censored observations, while $\hat{H}^c(\boldsymbol{y})$ in addition jumps at the singly censored observations.

For most existing estimators $\hat{F}$, we prefer to work with the simple bootstrap, since the estimators $\hat{F}$ and $\hat{G}$ needed for the obvious bootstrap are in many cases non-efficient. Only for van der Laan's (1996) estimator (see next section), we recommend the obvious bootstrap, since it is known that this estimator is efficient and since the obvious bootstrap provides us with the bootstrapped censoring times, which are needed for the calculation of the bootstrapped van der Laan estimator.

We are now able to construct a bootstrap estimate for the matrix $\boldsymbol{A}$, which is the (asymptotic) covariance matrix of $n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0)$. Define

$$n^{-1/2}\boldsymbol{S}_n^*(\hat{\boldsymbol{\theta}}) = n^{1/2} \int_{\hat{\Omega}^*} \boldsymbol{\psi}(\boldsymbol{y}, \hat{\boldsymbol{\theta}})\, d\hat{F}^*(\boldsymbol{y}),$$

where $\hat{F}^*(\boldsymbol{y})$ and $\hat{\Omega}^*$ are obtained from the bootstrap sample. Calculate this expression for B bootstrap samples, where B is a prespecified number, and let $\hat{\boldsymbol{A}}$ be the variance-covariance matrix of these B vectors.

4 Estimation of $F(\boldsymbol{y})$

In this section we discuss a number of estimators $\hat{F}(\boldsymbol{y})$, studied in the literature, that could be used in the proposed procedure. The verification of condition $(A3)$ for these estimators is postponed to the appendix.

- *van der Laan (1996).* Since the nonparametric maximum likelihood estimator (NPMLE) for bivariate censored data is inconsistent, the idea of van der Laan

(1996) is to slightly modify the data, and apply the NPMLE on the new data. More precisely, redefine the bivariate right censored data as the bivariate right censored data one would have seen if the observed vector of censoring times is pulled back to the left lower grid point of the grid rectangle it is in. Next, interval censor the observed failure times of the singly censored observations. Finally, compute the nonparametric maximum likelihood estimator for this reduced data structure by iterating the self-consistency equation

$$\hat{F}(\boldsymbol{y}) = n^{-1} \sum_{i=1}^{n} P_{\hat{F}}(T_1 \le y_1, T_2 \le y_2 | \boldsymbol{T} \in \boldsymbol{B}_i),$$

where $\boldsymbol{B}_i$ is a point, horizontal strip, vertical strip or an upper quadrant, depending on whether $\boldsymbol{\Delta}_i = (1,1), (0,1), (1,0)$ or (0,0) respectively. Note that this method requires the censoring times to be known. If not, they need to be simulated.

- *Akritas and Van Keilegom (2000).* The idea in this paper is based on the relationship

$$F(\boldsymbol{y}) = w(\boldsymbol{y}) \int_{-\infty}^{y_2} F_{1|2}(y_1|z_2)\, dF_2(z_2) + (1 - w(\boldsymbol{y})) \int_{-\infty}^{y_1} F_{2|1}(y_2|z_1)\, dF_1(z_1), \quad (4.1)$$

where $w(\boldsymbol{y})$ is any real number, F_1 and F_2 are the marginal distributions of respectively T_1 and T_2, $F_{1|2}$ is the conditional distribution of T_1 given T_2 and similarly for $F_{2|1}$. In order to estimate $F(\boldsymbol{y})$, we first of all replace F_j $(j = 1, 2)$ by the usual Kaplan-Meier (1958) estimator. An estimator for $F_{1|2}(y_1|y_2)$ is obtained by a Beran (1981)-type estimator, except that the averaging is done only over the uncensored observations in a window around y_2, since for a censored value of T_2, it is not clear whether or not it belongs to the window.

Finally, we like to note that if the data set of interest is quite small, it might be preferable to use a semiparametric estimator, instead of one of the two nonparametric estimators described above. See e.g. Wang and Wells (1999), where the bivariate distribution is estimated by choosing an appropriate copula function and estimating the marginals by the Kaplan-Meier estimator.

5 Simulations

To illustrate the finite sample performance of the proposed M-estimators, we carry out a small simulation study, in which we compare the performance of the vector of

Table 1: *Bias, variance and mean squared error of the L_1 median based on the van der Laan-estimator ($\hat{\boldsymbol{\theta}}_{VDL}$) and the Akritas-Van Keilegom estimator ($\hat{\boldsymbol{\theta}}_{AVK}$), and the vector of marginal medians ($\hat{\boldsymbol{\theta}}_{MAR}$). The first line shows the results for the first component of $\hat{\boldsymbol{\theta}}$, the second line for the second component.*

p	ρ_T	n	Bias			Variance			MSE		
			$\hat{\boldsymbol{\theta}}_{VDL}$	$\hat{\boldsymbol{\theta}}_{AVK}$	$\hat{\boldsymbol{\theta}}_{MAR}$	$\hat{\boldsymbol{\theta}}_{VDL}$	$\hat{\boldsymbol{\theta}}_{AVK}$	$\hat{\boldsymbol{\theta}}_{MAR}$	$\hat{\boldsymbol{\theta}}_{VDL}$	$\hat{\boldsymbol{\theta}}_{AVK}$	$\hat{\boldsymbol{\theta}}_{MAR}$
0.2	0	25	-.0129	-.0112	-.0039	.0580	.0553	.0658	.0582	.0554	.0658
			-.0107	-.0107	.0054	.0534	.0547	.0714	.0535	.0548	.0714
		50	-.0109	.0035	.0096	.0304	.0288	.0391	.0305	.0288	.0392
			-.0212	-.0084	-.0073	.0274	.0261	.0349	.0278	.0262	.0350
		100	-.0166	-.0050	-.0049	.0130	.0126	.0147	.0133	.0126	.0147
			-.0123	-.0042	-.0033	.0133	.0134	.0171	.0135	.0134	.0171
0.2	0.3	25	-.0009	-.0435	-.0039	.0601	.0559	.0658	.0601	.0578	.0658
			-.0065	-.0457	.0049	.0585	.0541	.0714	.0585	.0562	.0714
		50	-.0071	-.0361	.0096	.0312	.0284	.0391	.0313	.0297	.0392
			-.0156	-.0446	-.0021	.0280	.0267	.0368	.0282	.0287	.0368
		100	-.0075	-.0412	-.0049	.0135	.0126	.0147	.0136	.0143	.0147
			-.0074	-.0414	-.0033	.0138	.0135	.0160	.0139	.0152	.0160
0.3	0	25	-.0220	-.0192	-.0016	.0662	.0622	.0791	.0667	.0626	.0791
			-.0181	-.0124	-.0034	.0621	.0639	.0764	.0624	.0641	.0764
		50	-.0242	-.0030	.0158	.0332	.0307	.0417	.0338	.0307	.0419
			-.0222	-.0082	.0023	.0299	.0297	.0396	.0304	.0298	.0396
		100	-.0206	-.0038	-.0039	.0154	.0142	.0162	.0158	.0142	.0162
			-.0165	-.0055	-.0058	.0160	.0148	.0183	.0163	.0148	.0183
0.3	0.3	25	-.0097	-.0654	-.0016	.0689	.0586	.0791	.0690	.0629	.0791
			-.0133	-.0539	.0040	.0608	.0657	.0770	.0610	.0686	.0770
		50	.0000	-.0546	.0158	.0346	.0299	.0417	.0346	.0329	.0419
			-.0050	-.0608	.0011	.0316	.0293	.0390	.0316	.0330	.0390
		100	-.0063	-.0554	-.0039	.0161	.0142	.0162	.0161	.0173	.0162
			-.0039	-.0564	-.0012	.0162	.0145	.0176	.0162	.0177	.0176

marginal medians with the L_1 median for two estimators of the bivariate distribution : the estimator of van der Laan (1996) and the one of Akritas and Van Keilegom (2000). The data are generated in the following way. The vector of (transformed) survival times follows a bivariate normal distribution with mean vector zero and variances equal to one. The values of the correlation (denoted by ρ_T) considered in the simulation are 0 and 0.3. This means that the true vector of marginal medians, as well as the L_1 median, equals $(0, 0)$. The vector of (transformed) censoring times also follows a bivariate normal distribution, with variance-covariance matrix equal to the identity matrix and with mean given by the vector $(m, m)'$, where m is specified in such a way that the probability of censoring in each of the marginals (denoted by p) is given by 0.2 for the first simulation and 0.3 for the second one. For the situation where $\rho_T = 0$, this means that 36%, respectively 51% of the data have at least one censored component. The sample sizes considered are 25, 50 and 100. The results are based on 500 simulation runs and are shown in Table 1. To select the bandwidth required for the calculation of both $\hat{F}_{VDL}$ and $\hat{F}_{AVK}$ we adopted the approach of selecting the one for which the sum of the mean squared error of the two components of $\hat{\boldsymbol{\theta}}$ is minimal. The results in Table 1 show that in most situations the MSE of the L_1 median based on the estimator of Akritas and Van Keilegom is the smallest, closely followed by the one of van der Laan. The marginal median has considerably larger MSE. Note also that both estimators of the L_1 median tend to underestimate the true L_1 median and that for all three estimators the bias has only little effect on the MSE.

6 Data analysis

We will illustrate the calculations of the multivariate L_1 median and an estimate of its covariance matrix on data provided in an example discussed by McGilchrist and Aisbett (1991). The data, given in the reference, consist in the recurrence times to infection at point of insertion of the catheter for 38 kidney patients using portable dialysis equipment. For each patient two such times are recorded and censoring has taken place in the data. Of the 38 patients, 23 are doubly uncensored, 12 are singly censored, and 3 are doubly censored. The data values range from under 10 to over 500 days in each component and are shown in Figure 1.

We will calculate the L_1 median by using both the van der Laan (1996) estimator

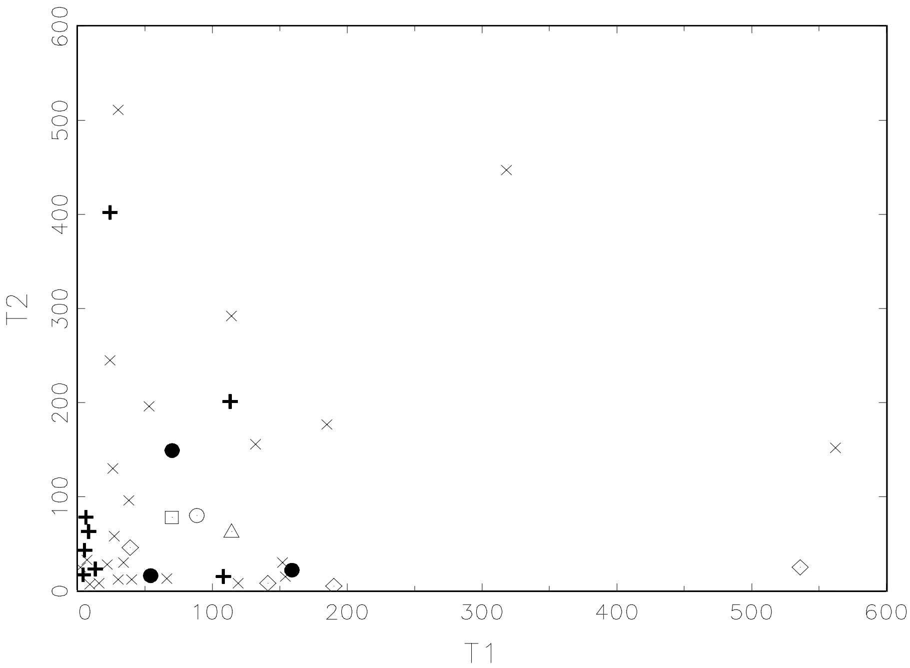

Figure 1: *Recurrence times to infection (in days) of 38 kidney patients. Uncensored observations are indicated by* $\times$*, observations censored in the first component by* $+$*, observations censored in the second component by* $\diamond$ *and doubly censored observations by* $\bullet$*.* $\hat{\boldsymbol{\theta}}_{VDL}$ *is indicated by* $\circ$*,* $\hat{\boldsymbol{\theta}}_{AVK}$ *by* $\square$ *and* $\hat{\boldsymbol{\theta}}_{MAR}$ *by* $\triangle$*.*

($\hat{\boldsymbol{\theta}}_{VDL}$) and the estimator proposed by Akritas and Van Keilegom (2000) ($\hat{\boldsymbol{\theta}}_{AVK}$) (see Section 4 for a discussion on these estimators). We start with computing the van der Laan estimator. Since the censoring times corresponding to uncensored observations are not observed in this example, we follow the suggestion of van der Laan and simulate them using a Kaplan-Meier estimator for the conditional censoring distribution. The last observation is changed to a censored observation so the Kaplan-Meier estimate is a proper distribution. The simulated censoring times are given in Table 2. The size of the grid rectangles is determined such that the average number of uncensored observations in each cell is some fixed number k. To select k, we estimate the mean squared error (MSE) of $\hat{\theta}_{VDL,j}$ ($j = 1, 2$) by means of the obvious bootstrap procedure described in Section 3 and for a grid of k values and select the value of k for which the sum of the MSE of $\hat{\theta}_{VDL,1}$ and

$\hat{\theta}_{VDL,2}$ is minimal. Table 3 shows the sum of the MSE's for a number of values of k. The value $k = 0.5$ yields the lowest MSE and is therefore selected. This leads to a grid size of 6.37. (To calculate the MSE of $\hat{\theta}_j$, note that $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = n^{-1/2}\boldsymbol{B}^{-1}\boldsymbol{S}_n(\boldsymbol{\theta}_0) + o_P(1)$ (see proof of Theorem 2.2) and hence $\text{AMSE}(\hat{\theta}_j) = n^{-2}[\boldsymbol{B}^{-1}\text{Var}\{\boldsymbol{S}_n(\boldsymbol{\theta}_0)\}\boldsymbol{B}^{-1}]_{jj} + n^{-2}[\boldsymbol{B}^{-1}$ $\text{Bias}\{\boldsymbol{S}_n(\boldsymbol{\theta}_0)\}]_j^2$, which can be estimated by means of the bootstrap procedure of Section 3).

Since the calculation of the estimator of van der Laan requires the data to live on a rectangle $[0, \tau_1] \times [0, \tau_2]$ such that $F(\tau_1, \tau_2) = 1$ and $S(\tau_1-, \tau_2-) > 0$, we project all data points outside $[0, 200] \times [0, 300]$ (5 data points in total) onto this rectangle, since it seems reasonable to believe that the distribution of the projected data approximately satisfies the above conditions. Once the van der Laan estimator is calculated, the L_1 median is found by using a bivariate Newton-Raphson algorithm. The values of the L_1 median and the marginal median (obtained from the marginal Kaplan Meier estimators) can be found in Table 4.

Next, we calculate

$$\hat{\boldsymbol{B}} = \begin{pmatrix} 0.0045 & -0.0005 \\ -0.0005 & 0.0058 \end{pmatrix}$$

and using the obvious bootstrap as described in Section 3, we find, based on B=1000 bootstrap samples,

$$\hat{\boldsymbol{A}} = \begin{pmatrix} 0.5890 & 0.1003 \\ 0.1003 & 0.6049 \end{pmatrix}$$

C_1	562	13	562	562	562	562	13	562	113	562	113	562	113
C_2	511	149	511	511	511	511	511	511	511	22	511	8	511
C_1	70	562	4	562	159	159	108	562	24	70	562	159	113
C_2	149	25	511	511	511	22	511	511	511	511	46	511	511
C_1	562	70	562	159	562	5	562	562	562	54	6	8	
C_2	511	511	511	149	511	511	511	5	511	16	511	511	

Table 2: *Simulated censoring times*

k	MSE	h	MSE
0.3	1482.1	10	2588.6
0.4	1682.9	15	2263.4
0.5	1451.2	20	1664.8
0.6	1716.4	25	1765.9
1	3033.8	30	2120.4

Table 3: *Sum of the MSE of $\hat{\theta}_1$ and $\hat{\theta}_2$ for different values of k (for van der Laan estimator) and different values of the bandwidth h (for Akritas-Van Keilegom estimator).*

	$\hat{\boldsymbol{\theta}}_{VDL}$	$\hat{\boldsymbol{\theta}}_{AVK}$	$\hat{\boldsymbol{\theta}}_{MAR}$
T_1	88.34	69.89	114
T_2	80.05	78.03	63

Table 4: *L_1 median based on the van der Laan estimator ($\hat{\boldsymbol{\theta}}_{VDL}$), L_1 median based on the Akritas-Van Keilegom estimator ($\hat{\boldsymbol{\theta}}_{AVK}$) and the vector of marginal medians ($\hat{\boldsymbol{\theta}}_{MAR}$).*

and, the estimated asymptotic covariance matrix of $n^{1/2}(\hat{\boldsymbol{\theta}}_{VDL} - \boldsymbol{\theta}_0)$ is

$$\hat{\boldsymbol{B}}^{-1}\hat{\boldsymbol{A}}\hat{\boldsymbol{B}}^{-1} = \begin{pmatrix} 30915.9 & 8169.5 \\ 8169.5 & 19022.2 \end{pmatrix}.$$

Hence, the estimated standard error of $\hat{\theta}_{VDL,1}$ is $\sqrt{30915.9/38} = 28.52$ and similarly the estimated standard error of $\hat{\theta}_{VDL,2}$ is 22.37. A simple comparison of the components of the L_1 median, taking into account the estimated standard errors, suggests that there is no statistical evidence that they differ.

Next, we calculate the L_1 median by means of the estimator of Akritas and Van Keilegom (2000). We select the bandwidth h_n (which is required for the calculations of this estimator) in the same way as for the grid size of the van der Laan estimator (i.e. by minimizing the sum of the MSE of $\hat{\theta}_{AVK,1}$ and $\hat{\theta}_{AVK,2}$). Table 3 shows that $h_n = 20$ minimizes the sum of the MSE's and is therefore selected. The weights $w(\boldsymbol{y})$ in relation (4.1) are chosen to be 1/2, since there is no a priori reason not to assign the same weights to the two integrals in (4.1). The L_1 median $\hat{\boldsymbol{\theta}}_{AVK}$ is shown in Table 4. In a similar way as before, we calculate the variance covariance matrix of $\hat{\boldsymbol{\theta}}_{AVK}$ (except that we now use the simple bootstrap instead of the obvious bootstrap, as was explained in Section 3) and find

$$\hat{\boldsymbol{B}}^{-1}\hat{\boldsymbol{A}}\hat{\boldsymbol{B}}^{-1} = \begin{pmatrix} 23402.5 & 15021.0 \\ 15021.0 & 28933.7 \end{pmatrix}.$$

Hence, the standard error of $\hat{\theta}_{AVK,1}$ is 24.82 and for $\hat{\theta}_{AVK,2}$ we have 27.52. So, also for this estimator, the standard errors indicate that the two components of the L_1 median do not differ significantly.

Appendix : Proof of main results

Proof of Theorem 2.1. Consider

$$\begin{aligned} n^{-1/2}[\boldsymbol{S}_n(\boldsymbol{\theta}_0) - \boldsymbol{S}(\boldsymbol{\theta}_0)] &= n^{1/2}\int_{\hat{\Omega}} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, d(\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y})) \\ &\quad + \left[n^{1/2}\int_{\hat{\Omega}} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y}) - n^{1/2}\int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y})\right] \\ &= \sum_{j=1}^{2} T_j. \end{aligned}$$

From condition $(A3)$ we have that

$$\begin{aligned} T_1 &= n^{1/2}\int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, d(\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y})) + \boldsymbol{o}_P(1) \\ &= n^{-1/2}\sum_{i=1}^{n}\int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dg(\tilde{\boldsymbol{T}}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}) + \boldsymbol{o}_P(1). \end{aligned}$$

The term T_2 can be written as $n^{-1/2}\sum_{i=1}^{n} \boldsymbol{k}(\tilde{\boldsymbol{T}}_i, \boldsymbol{\Delta}_i, \boldsymbol{\theta}_0) + \boldsymbol{o}_P(1)$, again by condition $(A3)$. The result now follows from the central limit theorem.

The following is a technical lemma, which will be needed in the proofs of Theorems 2.2 and 2.3.

Lemma A.1 *Assume $(A1) - (A3)$. Then, for all $B > 0$,*

$$\sup_{\|\boldsymbol{b}\| \le B} |n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0 + \boldsymbol{b}n^{-1/2}) - n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0) + \boldsymbol{B}\boldsymbol{b}| \xrightarrow{P} \boldsymbol{0}. \tag{A.1}$$

and

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \boldsymbol{O}_P(1). \tag{A.2}$$

Proof. To prove this result, we will make use of Theorem 2 in Brown (1985) (applied to $-\boldsymbol{S}_n(\boldsymbol{\theta}_0)$ instead of $\boldsymbol{S}_n(\boldsymbol{\theta}_0)$ however), which states that the two above results follow from the conditions (2), (3), (10) and (11) in his paper. The verification of conditions (2), (3) and (11) is straightforward, since they require (respectively) that $n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0)$ is asymptotically normal (see Theorem 2.1), that $\boldsymbol{B}$ is positive definite (see condition $(A2)(ii)$) and that the matrix of derivatives of $\boldsymbol{S}_n(\boldsymbol{\theta})$ is negative definite (see condition $(A2)(iii)$). It remains to verify his condition (10), which states that the expression within absolute values in (A.1) is $\boldsymbol{o}_P(1)$ for every $\boldsymbol{b}$ such that $\|\boldsymbol{b}\| \leq B$. From condition $(A3)$, we have that $n^{-1/2}[\boldsymbol{S}_n(\boldsymbol{\theta}_0 + \boldsymbol{b}n^{-1/2}) - \boldsymbol{S}(\boldsymbol{\theta}_0 + \boldsymbol{b}n^{-1/2}) - \boldsymbol{S}_n(\boldsymbol{\theta}_0) + \boldsymbol{S}(\boldsymbol{\theta}_0)] = \boldsymbol{o}_P(1)$ and hence it remains to show that for each $\|\boldsymbol{b}\| \leq B$, $n^{-1/2}[\boldsymbol{S}(\boldsymbol{\theta}_0 + \boldsymbol{b}n^{-1/2}) - \boldsymbol{S}(\boldsymbol{\theta}_0)] + \boldsymbol{B}\boldsymbol{b} \xrightarrow{P} \mathbf{0}$. This follows from the mean value theorem together with condition $(A1)(iii)$.

We are now ready to prove the remaining results of Section 2.

Proof of Theorem 2.2. Since $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \boldsymbol{O}_P(1)$, we can apply (A.1) on $\boldsymbol{b} = n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ yielding

$$n^{-1/2}\boldsymbol{S}_n(\hat{\boldsymbol{\theta}}) - n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0) + n^{1/2}\boldsymbol{B}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \boldsymbol{o}_P(1).$$

Because $\boldsymbol{S}_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, it follows that

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = n^{-1/2}\boldsymbol{B}^{-1}\boldsymbol{S}_n(\boldsymbol{\theta}_0) + \boldsymbol{o}_P(1),$$

which tends to a normal distribution with zero mean and covariance matrix $\boldsymbol{B}^{-1}\boldsymbol{A}(\boldsymbol{B}^t)^{-1}$ (see Theorem 2.1).

Proof of Theorem 2.3. For the location model, the asymptotic distribution of

$$n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0) = n^{1/2}\int_{\hat{\Omega}} \boldsymbol{\psi}_{\mathbf{0}}(\boldsymbol{y} - \boldsymbol{\theta}_0)\, d\hat{F}(\boldsymbol{y})$$

under H_1 equals the asymptotic distribution under H_0 of

$$\begin{aligned} & n^{1/2}\int_{\hat{\Omega}_\ell} \boldsymbol{\psi}_{\mathbf{0}}(\boldsymbol{y} - \boldsymbol{\theta}_0)\, d\hat{F}(\boldsymbol{y} - n^{-1/2}\boldsymbol{\gamma}) \\ &= n^{1/2}\int_{\hat{\Omega}} \boldsymbol{\psi}_{\mathbf{0}}(\boldsymbol{y} - \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\gamma})\, d\hat{F}(\boldsymbol{y}) \\ &= n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0 - n^{-1/2}\boldsymbol{\gamma}) \\ &= n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0) + \boldsymbol{B}\boldsymbol{\gamma} + \boldsymbol{o}_P(1) \end{aligned}$$

by (A.1), where $\hat{\Omega}_\ell = \{\boldsymbol{y} : (y_1 - \gamma_1 n^{-1/2}, y_2 - \gamma_2 n^{-1/2}) \in \hat{\Omega}\}$. Since the latter converges to a d-variate normal distribution with mean $\boldsymbol{B\gamma}$ and variance matrix $\boldsymbol{A}$, the result follows. For the scale model the proof is similar : the asymptotic distribution of $n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0)$ under H_1 equals the asymptotic distribution under H_0 of

$$
\begin{aligned}
&n^{1/2} \int_{\hat{\Omega}_s} \boldsymbol{\psi}_{\mathbf{0}}(\boldsymbol{y} - \boldsymbol{\theta}_0)\, d\hat{F}(\boldsymbol{y} e^{-\gamma n^{-1/2}}) \\
&= n^{1/2} \int_{\hat{\Omega}} \boldsymbol{\psi}_{\mathbf{0}}(\boldsymbol{y} e^{\gamma n^{-1/2}} - \boldsymbol{\theta}_0)\, d\hat{F}(\boldsymbol{y}) \\
&= n^{1/2} \int_{\hat{\Omega}} \boldsymbol{\psi}_{\mathbf{0}}(\boldsymbol{y} - \boldsymbol{\theta}_0 e^{-\gamma n^{-1/2}})\, d\hat{F}(\boldsymbol{y}) \\
&= n^{-1/2} \boldsymbol{S}_n(\boldsymbol{\theta}_0 e^{-\gamma n^{-1/2}}) \\
&= n^{-1/2} \boldsymbol{S}_n(\boldsymbol{\theta}_0 - \gamma \boldsymbol{\theta}_0 n^{-1/2}) + \boldsymbol{o}_P(1) \\
&= n^{-1/2} \boldsymbol{S}_n(\boldsymbol{\theta}_0) + \gamma \boldsymbol{B} \boldsymbol{\theta}_0 + \boldsymbol{o}_P(1),
\end{aligned}
$$

where $\hat{\Omega}_s = \{\boldsymbol{y} : (y_1 e^{-\gamma n^{-1/2}}, y_2 e^{-\gamma n^{-1/2}}) \in \hat{\Omega}\}$, from which the result follows.

Verification of (A3) for van der Laan (1996)-estimator. In his Theorem 5.1 he proves an asymptotic representation for his estimator. One of the conditions in this result requires that Ω equals $(-\infty, \tau_1] \times (-\infty, \tau_2]$, where τ_1 and τ_2 satisfy $S_F(\tau_1-, \tau_2-) > 0$ and $S_G(\tau_1, \tau_2) > 0$, and where $S_F(\boldsymbol{y}) = P(T_1 > y_1, T_2 > y_2)$ is the survival distribution of $\boldsymbol{T}$ and similarly for S_G. Choosing $\tau_j = H_j^{-1}(1-\varepsilon)$ (where H_j is the marginal distribution of $\tilde{T}_j$ and $\varepsilon > 0$), it is easily seen that the required condition is satisfied provided that $S_F(F_1^{-1}(1-\varepsilon)-, F_2^{-1}(1-\varepsilon)-) > 0$ and $S_G(G_1^{-1}(1-\varepsilon), G_2^{-1}(1-\varepsilon)) > 0$. Natural estimators for τ_j $(j = 1, 2)$ are $\hat{\tau}_j = \hat{H}_j^{-1}(1-\varepsilon)$ (where $\hat{H}_j(y) = n^{-1}\sum_{i=1}^n I(\tilde{T}_{ij} \leq y)$). Since the remainder term in his Theorem 5.1 satisfies $\sup_{\boldsymbol{y} \in \Omega} |R_n(\boldsymbol{y})| = o_P(n^{-1/2})$, condition (2.4) is satisfied e.g. for $\boldsymbol{\psi}$'s that are continuously differentiable and that are defined on a compact region of $I\!R^2$. Next, we need to show (2.5). Write

$$
\begin{aligned}
&\int_{\hat{\Omega}} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y}) - \int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y}) \qquad \text{(A.3)} \\
&= \int_{\tau_1}^{\hat{\tau}_1} \int_{-\infty}^{\hat{\tau}_2} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y}) + \int_{-\infty}^{\tau_1} \int_{\tau_2}^{\hat{\tau}_2} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y}).
\end{aligned}
$$

If we use the notation

$$L_{1,\boldsymbol{\theta}}(z) = \int_{-\infty}^{z} \int_{-\infty}^{\tau_2} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})\, dF(\boldsymbol{y}),$$

the first term of this sum equals

$$\int_{\tau_1}^{\hat{\tau}_1} \int_{-\infty}^{\tau_2} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y}) + \boldsymbol{o}_P(n^{-1/2})$$
$$= L_{1,\boldsymbol{\theta}_0}(\hat{\tau}_1) - L_{1,\boldsymbol{\theta}_0}(\tau_1) + \boldsymbol{o}_P(n^{-1/2})$$
$$= L'_{1,\boldsymbol{\theta}_0}(\tau_1)(\hat{\tau}_1 - \tau_1) + \boldsymbol{o}_P(n^{-1/2})$$
$$= -\frac{L'_{1,\boldsymbol{\theta}_0}(\tau_1)}{h_1(\tau_1)} \left[n^{-1} \sum_{i=1}^{n} I(\tilde{T}_{i1} \leq \tau_1) - (1-\varepsilon) \right] + \boldsymbol{o}_P(n^{-1/2}),$$

provided H_1 is differentiable in τ_1 and $h_1(\tau_1) > 0$ (h_1 being the density of H_1). The derivation for the second term of (A.3) is similar. This shows that (2.5) is satisfied. Finally, conditions (2.6) and (2.7) hold (among others) for the $\boldsymbol{\psi}$ functions described above. If in addition the conditions in van der Laan (1996) are satisfied, the results of Section 2 can be applied.

Verification of (A3) for Akritas and Van Keilegom (2000)-estimator. The representation in (2.3) is given by Theorem 3.4 of Akritas and Van Keilegom (2000). Condition (2.4) holds for the same functions $\boldsymbol{\psi}$ as the ones that can be used for van der Laan. The verification of equation (2.5) requires more work. First note that

$$\Omega = \left\{ \boldsymbol{y} : y_1 \leq \inf_{y \leq y_2} \tau_1(y) \text{ and } y_2 \leq \inf_{y \leq y_1} \tau_2(y) \right\},$$

where we choose $\tau_1(y) = H^{-1}_{1|2u}(1-\varepsilon|y)$ for some $\varepsilon > 0$, where $H_{1|2u}(y_1|y_2) = P(\tilde{T}_1 \leq y_1|\tilde{T}_2 = y_2, \Delta_2 = 1)$, and similarly for $\tau_2(y)$. We estimate this region by replacing $\tau_1(y)$ by $\hat{\tau}_1(y) = \hat{H}^{-1}_{1|2u}(1-\varepsilon|y)$ (and similarly for $\tau_2(y)$), where

$$\hat{H}_{1|2u}(y_1|y_2) = \sum_{i=1}^{n} W_{n2i}(y_2, h_n) I(\tilde{T}_{i1} \leq y_1),$$

$$W_{n2i}(y, h_n) = \begin{cases} \dfrac{K\left(\frac{y - T_{i2}}{h_n}\right)}{\sum_{\Delta_{j2}=1} K\left(\frac{y-T_{j2}}{h_n}\right)} & \text{if } \Delta_{i2} = 1 \\ 0 & \text{if } \Delta_{i2} = 0, \end{cases}$$

K is a known density function (kernel) and $\{h_n\}$ a sequence of positive constants tending to zero as n tends to infinity (bandwidth sequence). It is easy to show that

$$\int_{\hat{\Omega}} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y}) - \int_{\Omega} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y})$$

$$= \int_{-\infty}^{e_1} \int_{\inf_{y \le y_1} \tau_2(y)}^{\inf_{y \le y_1} \hat{\tau}_2(y)} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y})$$

$$+ \int_{-\infty}^{e_2} \int_{\inf_{y \le y_2} \tau_1(y)}^{\inf_{y \le y_2} \hat{\tau}_1(y)} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta}_0)\, dF(\boldsymbol{y}) + \boldsymbol{o}_P(n^{-1/2}), \tag{A.4}$$

where $\boldsymbol{e} = (e_1, e_2)$ is the point in the plane where the curves $\tau_1(y_2)$ and $\tau_2(y_1)$ intersect (we assume for simplicity that this point is unique). Using the notation

$$N_{1,\boldsymbol{\theta}}(y_2, z) = \int_{-\infty}^{z} \boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})\, dF_{1|2}(y_1|y_2),$$

we can write the second term of (A.4) as

$$\int_{-\infty}^{e_2} \left[N_{1,\boldsymbol{\theta}_0}\left(y_2, \inf_{y \le y_2} \hat{\tau}_1(y)\right) - N_{1,\boldsymbol{\theta}_0}\left(y_2, \inf_{y \le y_2} \tau_1(y)\right) \right] dF_2(y_2)$$

$$= \int_{-\infty}^{e_2} N'_{1,\boldsymbol{\theta}_0}\left(y_2, \inf_{y \le y_2} \tau_1(y)\right) \left[\inf_{y \le y_2} \hat{\tau}_1(y) - \inf_{y \le y_2} \tau_1(y) \right] dF_2(y_2)$$

$$+ o_P\left(\left| \inf_{y \le y_2} \hat{\tau}_1(y) - \inf_{y \le y_2} \tau_1(y) \right| \right).$$

Hence, it suffices to construct an asymptotic representation for $\inf_{y \le y_2} \hat{\tau}_1(y) - \inf_{y \le y_2} \tau_1(y)$ (and similarly for $\tau_2(y)$). Let $z = \text{argmin}_{y \le y_2} \tau_1(y)$ and write

$$\inf_{y \le y_2} \hat{\tau}_1(y) - \tau_1(z) = \left[\inf_{y \le y_2, \hat{\tau}_1(y) \le \hat{\tau}_1(z)} \hat{\tau}_1(y) - \hat{\tau}_1(z) \right] + [\hat{\tau}_1(z) - \tau_1(z)]$$

$$= T_1 + T_2.$$

The term T_2 equals

$$T_2 = -[\hat{H}_{1|2u}(\tau_1(z)|z) - (1 - \varepsilon)]/h_{1|2u}(\tau_1(z)|z) + o_P(n^{-1/2}).$$

We will show that $T_1 = o_P(n^{-1/2})$. Fix $\varepsilon > 0$.

$$P(n^{1/2}|T_1| > \varepsilon)$$

$$= P\left(n^{1/2}|T_1| > \varepsilon \,\middle|\, \sup_y |\hat{\tau}_1(y) - \tau_1(y)| \le \frac{a_n}{2} \right) P\left(\sup_y |\hat{\tau}_1(y) - \tau_1(y)| \le \frac{a_n}{2} \right)$$

$$+ P\left(n^{1/2}|T_1| > \varepsilon \,\middle|\, \sup_y |\hat{\tau}_1(y) - \tau_1(y)| > \frac{a_n}{2} \right) P\left(\sup_y |\hat{\tau}_1(y) - \tau_1(y)| > \frac{a_n}{2} \right),$$

where $a_n = Kn^{-1/4}$ for some $K > 0$. Since $\sup_y |\hat{\tau}_1(y) - \tau_1(y)| = O_P(n^{-1/2})$, the second term above is $o(1)$. The first one is bounded by

$$P\left(n^{1/2} \inf_{y \le y_2, \tau_1(y) \le \tau_1(z) + a_n} |\hat{\tau}_1(y) - \hat{\tau}_1(z)| > \varepsilon \right)$$

$$= P\left(n^{1/2} \sup_{y \le y_2, \tau_1(y) \le \tau_1(z) + a_n} |\hat{\tau}_1(y) - \hat{\tau}_1(z) - \tau_1(y) + \tau_1(z)| > \frac{\varepsilon}{2}\right)$$
$$+P\left(n^{1/2} \inf_{y \le y_2, \tau_1(y) \le \tau_1(z) + a_n} |\tau_1(y) - \tau_1(z)| > \frac{\varepsilon}{2}\right).$$

The second term of this sum is $o(1)$, as well as the first one, using the modulus of continuity of the estimator $\hat{H}_{1|2u}(y_1|y_2)$. This shows that $T_1 = o_P(n^{-1/2})$, and hence we have proved condition (2.5). Finally, conditions (2.6) and (2.7) are satisfied for the same $\boldsymbol{\psi}$ functions as the ones that can be used for van der Laan (1996).

Acknowledgment. The authors would like to thank Mark van der Laan, Derick Peterson and Jyrki Möttönen for sending programs and for helpful discussions.

References

Akritas, M. G. and Van Keilegom, I. (2000). Estimation of the bivariate and marginal distributions with censored data (submitted).

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

Bickel, P. J. (1965). On some asymptotically nonparametric competitors of Hotelling's T^2. *Ann. Math. Statist.*, **36**, 160-173.

Brown, B. M. (1983). Statistical uses of the spatial median. *J. Royal Statist. Soc., Series B*, **45**, 25-30.

Brown, B. M. (1985). Multiparameter linearization theorems. *J. Royal Statist. Soc., Series B*, **47**, 323-331.

Chaudhuri, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics. *Ann. Statist.*, **20**, 897-916.

Choi, K. and Marden, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *J. Amer. Statist. Assoc.*, **92**, 1581-1590.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.*, **16**, 1475-1489.

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.*, **76**, 312-319.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. J. (1986). *Robust statistics, the approach based on influence functions.* Wiley, New York.

Hettmansperger, T. P. and McKean, J. W. (1998). *Robust nonparametric statistical methods*. Arnold, London.

Huber, P. J. (1981). *Robust statistics*. Wiley, New York.

James, I. R. (1986). On estimating equations with censored data. *Biometrika*, **73**, 35-42.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457-481.

Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In : *Statistical Data Analysis Based on the L_1-norm and related methods*, ed. Y. Dodge, Amsterdam: North-Holland, 217-230.

Lopuhaä, H. P. and Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.*, **19**, 229-248.

Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.*, **4**, 51-67.

McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, **47**, 461-466.

Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonpar. Statist.*, **5**, 201-213.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.*, **84**, 487-493.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, **79**, 495-512.

Pruitt, R. C. (1991). Strong consistency of self-consistent estimators: general theory and an application to bivariate survival analysis. Technical report 543, Univ. Minnesota.

Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Review*, **58**, 263-277.

van der Laan, M. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.*, **24**, 596-627.

Wang, J.-L. (1999). Asymptotic properties of M-estimators based on estimating equations and censored data. *Scand. J. Statist.*, **26**, 297-318.

Wang, W. and Wells, M.T. (1999). Semi-parametric estimation of the bivariate survival function. Technical report, Institute of statistics, Academia Sinica.