

## Inference on the bivariate L1 median with censored data

**Citation for published version (APA):**

Van Keilegom, I., & Hettmansperger, T. P. (1999). *Inference on the bivariate L1 median with censored data*. (Memorandum COSOR; Vol. 9916). Technische Universiteit Eindhoven.

**Document status and date:**

Published: 01/01/1999

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

Memorandum COSOR 99-16

**Inference on the bivariate  $L_1$  median  
with censored data**

I. Van Keilegom and  
T.P. Hettmansperger

Eindhoven, October 1999  
The Netherlands

# Inference on the bivariate $L_1$ median with censored data

I. Van Keilegom and T. P. Hettmansperger

*Technical University of Eindhoven and Penn State University*

September 30, 1999

## **Abstract**

Consider two random variables subject to random right censoring, like the time to two different diseases for individuals under study or the survival times of twins. Of interest is the bivariate median of these two random variables.

There are various ways that the univariate median has been extended to higher dimensions for completely observed data. We concentrate on the so-called bivariate  $L_1$  median and extend this estimator to the censored data situation. The estimator is based on van der Laan (1996)'s estimator of the bivariate distribution of two random variables that are subject to censoring. Asymptotic results for the proposed estimator are established. The obtained results include the asymptotic normality of the estimator, its local power and the construction of a confidence region for the true median. 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; Right censoring.

# 1 Introduction

In this paper we consider the concept of bivariate median and sign test for the censored data model. The univariate median and sign test, both based on the  $L_1$  norm, have an extensive literature; see Hettmansperger and McKean (1998) and references therein. These  $L_1$  methods are highly robust but suffer from reduced efficiency at the normal model.

There are various ways that the median can be extended to higher dimensions. Small (1990) provides a full account of the various multivariate medians. Here, we concentrate on what Small calls the multivariate  $L_1$  median, defined as the vector  $\hat{\theta}$  that minimizes

$$\sum_{i=1}^n \|\mathbf{y}_i - \theta\| \tag{1.1}$$

for a multivariate data set  $\mathbf{y}_1, \dots, \mathbf{y}_n$  and  $\|\cdot\|$  denotes the Euclidean norm. Hence, we seek a point that minimizes the average distance to the data points. The multivariate  $L_1$  median was termed the mediancenter by Gower (1974) who discusses some of its properties and also provides a computational algorithm in the bivariate case. Brown (1983) referred to it as the spatial median and the corresponding sign test based on

$$S_n(\theta) = \sum_{i=1}^n \frac{\mathbf{y}_i - \theta}{\|\mathbf{y}_i - \theta\|} \tag{1.2}$$

as the angle test, useful when testing directional hypotheses. Small (1990) traces the history of the multivariate  $L_1$  median back to Hayford (1902). Gini and Galvani (1929) introduced the median into the statistical literature and then it was rediscovered by Haldane (1948).

This median is attractive for several reasons. The multivariate  $L_1$  median, in contrast to the vector of marginal medians, provides a better characterization of the center of the joint distribution. It has 50% breakdown (same value as the univariate median) and bounded influence; see Lopuhaä and Rousseeuw (1987) and Kemperman (1987). It is unique in dimensions greater than 1. The multivariate  $L_1$  median is equivariant under orthogonal transformations of the data but not under scale changes or affine transformations. 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 .637. Finally, Chaudhuri (1992) provides the most

complete and rigorous analysis of the asymptotic properties. He developed a Bahadur type representation for the multivariate  $L_1$  median. Möttönen and Oja (1995) develop multivariate spatial sign and rank methods based on (1.2) and Choi and Marden (1997) develop multivariate rank tests for analysis of variance based on (1.2).

We view (1.2) as defining an estimating equation for  $\hat{\boldsymbol{\theta}}$ . If  $\hat{F}(\mathbf{y})$  is the bivariate empirical cdf based on the sample  $\mathbf{y}_1, \dots, \mathbf{y}_n$  of bivariate observations, then  $\hat{\boldsymbol{\theta}}$  solves

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} d\hat{F}(\mathbf{y}) = \mathbf{0}. \quad (1.3)$$

The asymptotic properties of  $\hat{\boldsymbol{\theta}}$  follow from a linearization of the estimating equation. The corresponding population median  $\boldsymbol{\theta}$  is defined by (1.3) with  $\hat{F}(\mathbf{y})$  replaced by  $F(\mathbf{y})$ , the underlying bivariate distribution.

In the case of right censoring, we need a bivariate estimate  $\hat{F}(\mathbf{y})$  of  $F(\mathbf{y})$ . Singly censored observations (i.e. observations in which only one component is censored) cause difficulties; for example, the nonparametric ML estimator is not consistent in this case. van der Laan (1996) reviews the various ways that have been proposed to circumvent these difficulties. In this paper, we consider only van der Laan's solution. His solution results in interval censoring for the singly censored data. Then van der Laan's  $\hat{F}(\mathbf{y})$  is asymptotically efficient under certain conditions; see van der Laan (1996) for details.

We follow van der Laan (1996) in modelling bivariate right censored data as follows:  $\mathbf{T}$  is a bivariate lifetime vector with bivariate cdf  $F(\mathbf{y}) = P(T_1 \leq y_1, T_2 \leq y_2)$ . Let  $\mathbf{C}$  be a bivariate censoring vector with cdf  $G(\mathbf{y})$ . Assume that  $\mathbf{T}$  and  $\mathbf{C}$  are independent. Let  $(\mathbf{T}_i, \mathbf{C}_i)$   $i = 1, \dots, n$  be independent copies of  $(\mathbf{T}, \mathbf{C})$ . We observe

$$\mathbf{Y}_i = \{\tilde{T}_{i1}, D_{i1}, \tilde{T}_{i2}, D_{i2}\} \quad (i = 1, \dots, n) \quad (1.4)$$

where  $\tilde{T}_{ij} = \min(T_{ij}, C_{ij})$  (with cdf  $H(\mathbf{y})$ ),  $D_{ij} = I(T_{ij} \leq C_{ij})$  ( $j = 1, 2$ ) and where  $I(\cdot)$  is the indicator of the event. Note that the observations can be doubly uncensored, singly censored in either component, or doubly censored.

For the construction of the van der Laan estimator  $\hat{F}(\mathbf{y})$ , we need to assume that the censoring times  $\mathbf{C}_i$  are observed or discrete. If this assumption is not met, one needs to simulate the unobserved censoring variables; see Section 4 for an example. van der Laan (1996) proposes a discretization of the censoring times which entails using a grid of width  $O(h_n)$ , where  $h_n$  tends to zero as  $n$  tends to infinity. The transformed data are

denoted by  $\mathbf{Y}_i^h = (\tilde{T}_i^h, \mathbf{D}_i^h)$  ( $i = 1, \dots, n$ ). For raw survival and censoring times defined on  $[0, \infty) \times [0, \infty)$ , van der Laan (1996) defines  $\hat{F}(\mathbf{y})$  as the nonparametric ML estimator computed from  $\mathbf{Y}_i^h$  ( $i = 1, \dots, n$ ); see van der Laan (1996, Section 2) for more details on the construction of  $\hat{F}(\mathbf{y})$ . More generally, if  $\mathbf{T}$  and  $\mathbf{C}$  are transformed survival and censoring times not necessarily on  $[0, \infty) \times [0, \infty)$  (e.g. a log transformation of the raw times), then

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

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  $\hat{F}(y_1, y_2)$  in this case as the nonparametric ML estimator of van der Laan (1996) for the bivariate distribution of  $(V_1, V_2)$  and evaluated at  $v(y_1)$  and  $v(y_2)$ . Note that this definition also applies to raw survival and censoring times by taking  $v$  to be the identity function. The estimator must be computed iteratively. We denote  $G_h(\mathbf{y})$  for the cdf of the transformed censoring times and  $H_h(\mathbf{y})$  for the cdf of the  $\tilde{T}_i^h$ 's.

In Section 2, we state our main results on the limit distributions of the bivariate  $L_1$  median and sign test. In Section 3, we discuss a bootstrap approach to the estimation of the asymptotic covariance structure of the bivariate  $L_1$  median and sign test. An example is analyzed in Section 4. Proofs and derivations are given in the Appendix.

## 2 Main results

Because the asymptotic theory for  $\hat{\boldsymbol{\theta}}$  is based on an i.i.d. representation for  $\hat{F}(\mathbf{y})$  which is valid (under assumption (A1) and (A2) given below) for any  $\mathbf{y} = (y_1, y_2)$  such that  $y_j \leq \tau_j$  ( $j = 1, 2$ ) (where  $\tau_j$  is strictly less than the upper bound of the support of  $H_j(\mathbf{y}) = P(\tilde{T}_j \leq y)$ ), we need to work with a slightly modified version of (1.3) :

$$\mathbf{S}(\boldsymbol{\theta}) = n \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} dF(\mathbf{y}) \quad (2.1)$$

$$\mathbf{S}_n(\boldsymbol{\theta}) = n \int_{-\infty}^{\hat{\tau}_1} \int_{-\infty}^{\hat{\tau}_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} d\hat{F}(\mathbf{y}), \quad (2.2)$$

where we take  $\tau_j = H_j^{-1}(1 - \varepsilon)$  and  $\hat{\tau}_j = \hat{H}_j^{-1}(1 - \varepsilon)$  for  $\varepsilon > 0$  arbitrarily small and  $\hat{H}_j(\mathbf{y}) = n^{-1} \sum_{i=1}^n I(\tilde{T}_{ij} \leq y)$  ( $j = 1, 2$ ). Note that  $\mathbf{S}(\boldsymbol{\theta})$  can be made arbitrarily close to

$n \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\mathbf{y}-\boldsymbol{\theta}}{\|\mathbf{y}-\boldsymbol{\theta}\|} dF(\mathbf{y})$  if the upper bound of the support of  $H_j$  equals the upper bound of the support of  $F_j(\mathbf{y}) = P(T_j \leq \mathbf{y})$  for  $j = 1, 2$ . The true median is denoted by  $\boldsymbol{\theta}_0$  and is the value of  $\boldsymbol{\theta}$  for which  $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{0}$ .

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 bivariate  $L_1$  median of the log and the raw survival times.

**Location model :**  $F(\mathbf{y}) = F_0(y_1 - \theta_1, y_2 - \theta_2)$  for some prototype distribution  $F_0(y_1, y_2)$ .

**Scale model :**  $F(\mathbf{y}) = F_1(\frac{y_1}{\sigma_1}, \frac{y_2}{\sigma_2})$  for  $\sigma_1, \sigma_2 > 0$  and for some prototype distribution  $F_1(y_1, y_2)$  satisfying  $F_1(\mathbf{y}) = 0$  for  $y_1 < 0$  or  $y_2 < 0$ .

The main results require a number of regularity conditions which we mention below for convenient reference. The conditions will be expressed in terms of the raw survival times  $(V_1, V_2) = (v(T_1), v(T_2))$ . We use the notation  $F_V(\mathbf{y}) = P(V_1 \leq y_1, V_2 \leq y_2)$  and similar notations will be used for other distribution functions.

(A1) The bandwidth  $h_n$  tends to zero, but slower than  $n^{-1/18}$ .

(A2) (i) For all  $\varepsilon > 0$ ,  $S_{F_V}(F_{V,1}^{-1}(1 - \varepsilon), F_{V,2}^{-1}(1 - \varepsilon)) > 0$ , where  $S_{F_V}(\mathbf{y}) = P(V_1 > y_1, V_2 > y_2)$  and (with analogous notation)  $S_{G_V}(G_{V,1}^{-1}(1 - \varepsilon), G_{V,2}^{-1}(1 - \varepsilon)) > 0$ .

(ii)  $F_V(v(\tau_1), v(\tau_2)) = 1$  (data reduced to  $[0, v(\tau_1)] \times [0, v(\tau_2)]$ ).

(iii) Denote the grid points for the construction of  $\hat{F}_V(\mathbf{y})$  by  $u_k$  ( $k \geq 1$ ) for  $V_1$  and by  $w_\ell$  ( $\ell \geq 1$ ) for  $V_2$  depending on the choice of  $h_n > 0$ . Then,  $P(u_k < V_1 \leq u_{k+1}, V_2 \geq w_\ell) > \delta h_n$  and  $P(V_1 \geq u_k, w_\ell < V_2 \leq w_{\ell+1}) > \delta h_n$  for some  $\delta > 0$ .

(iv)  $F_V$  has a continuous density, uniformly bounded away from zero on  $[0, v(\tau_1)] \times [0, v(\tau_2)]$ .

(A3) The function  $H_{V,j}(y)$  ( $j = 1, 2$ ) is differentiable in  $H_{V,j}^{-1}(1 - \varepsilon)$  and  $h_{V,j}(H_{V,j}^{-1}(1 - \varepsilon)) > 0$ , where  $h_{V,j}(y)$  is the probability density function of  $H_{V,j}(y)$ .

(A4) The function  $v(y)$  has a bounded derivative on  $(-\infty, \tau_1 \vee \tau_2]$ .

(A5) (i) The matrix  $\mathbf{A}$ , defined in Theorem 2.1, is positive definite.

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

We start with the asymptotic normality of  $S_n(\boldsymbol{\theta}_0)$ . This result is useful for testing hypotheses concerning  $\boldsymbol{\theta}_0$ . The following notation is required :

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

$$L_{2,\boldsymbol{\theta}}(z) = \int_{-\infty}^{\tau_1} \int_{-\infty}^z \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} dF(\mathbf{y}).$$

**Theorem 2.1** Assume (A1) – (A4), (A5)(i). Then,

$$n^{-1/2} S_n(\boldsymbol{\theta}_0) \xrightarrow{d} N_2(\mathbf{0}, \mathbf{A}),$$

where

$$\mathbf{A} = \text{Var} \left[ \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} \frac{\mathbf{y} - \boldsymbol{\theta}_0}{\|\mathbf{y} - \boldsymbol{\theta}_0\|} dg(\mathbf{Y}^h, \mathbf{y}) - \sum_{j=1}^2 \frac{L'_{j,\boldsymbol{\theta}_0}(\tau_j)}{h_j(H_j^{-1}(1 - \varepsilon))} I(\tilde{T}_j \leq H_j^{-1}(1 - \varepsilon)) \right],$$

$h_j(y)$  is the probability density function of  $H_j(y)$  ( $j = 1, 2$ ) and the function  $g$  is discussed in Theorem A.1.

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

**Theorem 2.2** Assume (A1) – (A5). Then,

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_2(\mathbf{0}, \mathbf{B}^{-1} \mathbf{A} \mathbf{B}^{-1}),$$

where

$$\mathbf{B} = \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} \left[ \frac{\mathbf{I}}{\|\mathbf{y} - \boldsymbol{\theta}_0\|} - \frac{1}{\|\mathbf{y} - \boldsymbol{\theta}_0\|^3} (\mathbf{y} - \boldsymbol{\theta}_0)(\mathbf{y} - \boldsymbol{\theta}_0)^t \right] dF(\mathbf{y}).$$

Below we discuss the estimation of  $\mathbf{A}$  and  $\mathbf{B}$ . 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  $\hat{\boldsymbol{\theta}}$ .

**Theorem 2.3** Assume (A1) – (A5).

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

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



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  $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 e^{\gamma m^{-1/2}}$  for some fixed value  $\gamma$ . Then, under  $H_1$ ,

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

Theorem 2.3 can be used to estimate the local power of the test based on  $\mathbf{S}_n(\boldsymbol{\theta}_0)$ . 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  $\chi^2$ -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  $\mathbf{S}_n$  relative to another test.

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

**Theorem 2.4** Assume (A1)–(A5). Further, let  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{B}}$  be weakly consistent estimators of respectively  $\mathbf{A}$  and  $\mathbf{B}$ . Then, a  $(1 - \alpha)100\%$  confidence region for  $\boldsymbol{\theta}_0$  is given by the values of  $\boldsymbol{\theta}_0$  that satisfy

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

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

The calculation of standard errors, local power and the construction of a confidence region for  $\boldsymbol{\theta}_0$  requires the estimation of the matrices  $\mathbf{A}$  and  $\mathbf{B}$ . The estimation of the matrix  $\mathbf{B}$  is straightforward (simply replace  $F(\mathbf{y})$ ,  $\tau_1$ ,  $\tau_2$  and  $\boldsymbol{\theta}_0$  by the consistent estimators  $\hat{F}(\mathbf{y})$ ,  $\hat{\tau}_1$ ,  $\hat{\tau}_2$  and  $\hat{\boldsymbol{\theta}}$  respectively (see Theorem 5.1 in van der Laan (1996) for the consistency of  $\hat{F}(\mathbf{y})$ )). More attention needs to be paid to the estimation of  $\mathbf{A}$ . An estimator for  $\mathbf{A}$  could be obtained by replacing the function  $g(\mathbf{Y}^h, \mathbf{y})$  by an appropriate estimator. However, the estimation of this function is complicated since it has no explicit formula. For this reason we define in the next section a bootstrap procedure which can be used to obtain a bootstrap estimate for  $\mathbf{A}$ .

### 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  $\mathbf{A}$  in a consistent way.

To explain the proposed bootstrap procedure, consider first the one-dimensional case, which is due to Efron (1981). We use similar notations as for two dimensions. Conditioning on the responses  $T_i$  and censoring times  $C_i$  ( $i = 1, \dots, n$ ) we define the random variables  $T_i^*$  and  $C_i^*$  (independently) as follows :

$$\begin{aligned} T_1^*, \dots, T_n^* &\text{ are independent; } T_i^* \sim \hat{F} \\ C_1^*, \dots, C_n^* &\text{ are independent; } C_i^* \sim \hat{G}, \end{aligned}$$

where

$$\hat{F}(y) = 1 - \prod_{\tilde{T}_{(i)} \leq y} \left( 1 - \frac{1}{n - i + 1} \right)^{\Delta_{(i)}} \quad (3.1)$$

is the Kaplan-Meier estimator and  $\hat{G}(y)$  is the analogous estimator for the censoring times (replace the indicators  $\Delta_{(i)}$  in (3.1) by  $1 - \Delta_{(i)}$ ). Then define, for  $i = 1, \dots, n$ ,

$$\tilde{T}_i^* = \min(T_i^*, C_i^*) \text{ and } \Delta_i^* = I(T_i^* \leq C_i^*).$$

It is readily verified that the above procedure, called the obvious bootstrap, is equivalent to the one where the pairs  $(\tilde{T}_i^*, \Delta_i^*)$  are drawn (with replacement) from  $(\tilde{T}_1, \Delta_1), \dots, (\tilde{T}_n, \Delta_n)$ . The latter procedure is called the simple bootstrap.

The natural extension of the obvious bootstrap to two dimensions exists in replacing the Kaplan-Meier estimators  $\hat{F}(y)$  and  $\hat{G}(y)$  by the van der Laan estimators  $\hat{F}(y_1, y_2)$  and  $\hat{G}(y_1, y_2)$  (where  $\hat{G}(y_1, y_2)$  is obtained by interchanging the role of the survival and censoring times in the definition of  $\hat{F}(y_1, y_2)$ ). For the simple bootstrap, we draw the pairs  $(\tilde{T}_i^*, \Delta_i^*)$  with replacement from  $(\tilde{T}_1, \Delta_1), \dots, (\tilde{T}_n, \Delta_n)$ . Unlike the one-dimensional case, the obvious and the simple bootstrap are not equivalent in two dimensions. This is because for the obvious bootstrap the data points get (slightly) modified for the calculation of the van der Laan estimator, while this is not the case for the simple bootstrap. We prefer here to work with the obvious bootstrap, since this procedure provides us with the bootstrapped censoring times. We need these censoring times to calculate the bootstrapped van der Laan estimator  $\hat{F}^*(y_1, y_2)$ , which is obtained by replacing the original

data in  $\hat{F}(y_1, y_2)$  by the corresponding bootstrap data. Similarly,  $\hat{\tau}_j^* = \hat{H}_j^{*-1}(1 - \varepsilon)$  where  $\hat{H}_j^*(y) = n^{-1} \sum_{i=1}^n I(\tilde{T}_{ij}^* \leq y)$  ( $j = 1, 2$ ).

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

$$\mathbf{S}_n^*(\hat{\boldsymbol{\theta}}) = n \int_{-\infty}^{\hat{\tau}_1^*} \int_{-\infty}^{\hat{\tau}_2^*} \frac{\mathbf{y} - \hat{\boldsymbol{\theta}}}{\|\mathbf{y} - \hat{\boldsymbol{\theta}}\|} d\hat{F}^*(\mathbf{y}).$$

Calculate this expression for  $B$  bootstrap samples, where  $B$  is a prespecified number. Let  $\mathbf{S}_n^{(j)*}(\hat{\boldsymbol{\theta}})$  be the value of  $\mathbf{S}_n^*(\hat{\boldsymbol{\theta}})$  for resample  $j$  ( $j = 1, \dots, B$ ). Then, define

$$\hat{\mathbf{A}} = \frac{1}{n(B-1)} \sum_{j=1}^B \left[ \mathbf{S}_n^{(j)*}(\hat{\boldsymbol{\theta}}) - B^{-1} \sum_{i=1}^B \mathbf{S}_n^{(i)*}(\hat{\boldsymbol{\theta}}) \right] \left[ \mathbf{S}_n^{(j)*}(\hat{\boldsymbol{\theta}}) - B^{-1} \sum_{i=1}^B \mathbf{S}_n^{(i)*}(\hat{\boldsymbol{\theta}}) \right]^t.$$

## 4 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.

We first compute the van der Laan estimate of the bivariate recurrence time distribution function. Since the censoring times corresponding to uncensored observations are not observed in this example, we followed the suggestion of van der Laan and simulated them using a Kaplan-Meier estimator for the conditional censoring distribution. The last observation was changed to a censored observation so the Kaplan-Meier estimate is a proper distribution. The simulated censoring times are given in Table 1.

Now using van der Laan's algorithm, we find the support points and weights to be used in the estimating equation (2.2). For the algorithm we used a bandwidth of  $h = 24.35$ . This is the bandwidth for which the average number of uncensored observations in each cell is 1. Further,  $\hat{\tau}_1 = \hat{H}_1^{-1}(.995) = 562$ , and  $\hat{\tau}_2 = \hat{H}_2^{-1}(.995) = 511$ . Figure 1 shows the van der Laan support points. The bivariate  $L_1$  median and the component medians are marked in the figure, their values are given in Table 2.

$C_1$	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 1: *Simulated censoring times*

< Figure 1 about here >

	$L_1$ median	marginal medians
$T_1$	90	114
$T_2$	76	63

Table 2:  $L_1$  and marginal medians

Then,

$$\hat{B} = \begin{pmatrix} 0.0042732675 & -0.0004803128 \\ -0.0004803128 & 0.0056800525 \end{pmatrix}$$

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

$$\hat{A} = \begin{pmatrix} 0.7518447 & 0.1293847 \\ 0.1293847 & 0.6597610 \end{pmatrix}$$

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

$$\hat{B}^{-1} \hat{A} \hat{B}^{-1} = \begin{pmatrix} 43451.27 & 11376.54 \\ 11376.54 & 22062.79 \end{pmatrix}$$

Hence, the estimated standard error of  $\hat{\theta}_1$  is  $\sqrt{43451.27/38} = 33.82$  and similarly the estimated standard error of  $\hat{\theta}_2$  is 24.10.

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.

## Appendix : Proof of main results

We start with a result which can be found in van der Laan (1996), and which expresses the estimator  $\hat{F}(\mathbf{y})$  as a sum of i.i.d. terms and a remainder term of lower order. This result is useful for obtaining e.g. the asymptotic normality of  $n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0)$ , since it allows us to apply the central limit theorem. We do not mention the explicit formula of the function  $g$  in the main term of the representation below, because  $g$  has a fairly complicated (non-explicit) formula. Details on this formula can be found in van der Laan (1996).

**Theorem A.1** *Assume (A1), (A2). Then,*

$$\hat{F}(\mathbf{y}) - F(\mathbf{y}) = n^{-1} \sum_{i=1}^n g(\mathbf{Y}_i, \mathbf{y}) + R_n(\mathbf{y}),$$

where

$$\sup_{-\infty < y_1 \leq \tau_1, -\infty < y_2 \leq \tau_2} |R_n(\mathbf{y})| = o_P(n^{-1/2})$$

and  $g(\mathbf{z}, \mathbf{y})$  is defined as in Theorem 5.1 in van der Laan (1996).

Next, define

$$\begin{aligned} \mathbf{S}_n^g(\boldsymbol{\theta}) &= \sum_{i=1}^n \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} d(F(\mathbf{y}) + g(\mathbf{Y}_i, \mathbf{y})) \\ &\quad - n \sum_{j=1}^2 \frac{\mathbf{L}'_{j,\boldsymbol{\theta}}(\tau_j)}{h_j(H_j^{-1}(1-\varepsilon))} (\hat{H}_j(H_j^{-1}(1-\varepsilon)) - (1-\varepsilon)). \end{aligned}$$

The function  $\mathbf{S}_n^g(\boldsymbol{\theta})$  is a sum of i.i.d. terms and is the main term in the asymptotic representation for  $\mathbf{S}_n(\boldsymbol{\theta})$  :

**Theorem A.2** *Assume (A1) – (A4). Then,*

$$n^{-1/2} \sup_{\boldsymbol{\theta}} |\mathbf{S}_n(\boldsymbol{\theta}) - \mathbf{S}_n^g(\boldsymbol{\theta})| = o_P(1).$$

**Proof.** Consider

$$\begin{aligned} \mathbf{S}_n(\boldsymbol{\theta}) - \mathbf{S}(\boldsymbol{\theta}) &= n \int_{-\infty}^{\hat{\tau}_1} \int_{-\infty}^{\hat{\tau}_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} d(\hat{F}(\mathbf{y}) - F(\mathbf{y})) \\ &\quad + n \int_{-\infty}^{\hat{\tau}_1} \int_{\tau_2}^{\hat{\tau}_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} dF(\mathbf{y}) + n \int_{\tau_1}^{\hat{\tau}_1} \int_{-\infty}^{\tau_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} dF(\mathbf{y}) \\ &= \sum_{j=1}^3 T_j. \end{aligned}$$

From Theorem A.1 we have :

$$\begin{aligned} T_1 &= \sum_{i=1}^n \int_{-\infty}^{\hat{\tau}_1} \int_{-\infty}^{\hat{\tau}_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} dg(\mathbf{Y}_i^h, \mathbf{y}) + o_P(n^{1/2}) \\ &= \sum_{i=1}^n \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} dg(\mathbf{Y}_i^h, \mathbf{y}) + o_P(n^{1/2}) \end{aligned}$$

using integration by parts and the fact that  $\hat{\tau}_j - \tau_j = O_P(n^{-1/2})$  for  $j = 1, 2$  and  $\sum_{i=1}^n g(\mathbf{Y}_i^h, \mathbf{y}) = O_P(n^{1/2} h_n^{-3/2})$  uniformly in  $\mathbf{y} \in (-\infty, \tau_1] \times (-\infty, \tau_2]$  (see Theorem 5.1 in van der Laan (1996)). Next, write

$$\begin{aligned} T_2 &= n \int_{-\infty}^{\tau_1} \int_{\tau_2}^{\hat{\tau}_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} dF(\mathbf{y}) + o_P(n^{1/2}) \\ &= n(L_{2,\boldsymbol{\theta}}(\hat{\tau}_2) - L_{2,\boldsymbol{\theta}}(\tau_2)) + o_P(n^{1/2}) \\ &= nL'_{2,\boldsymbol{\theta}}(\tau_2)(\hat{H}_2^{-1}(1 - \varepsilon) - H_2^{-1}(1 - \varepsilon)) + o_P(n^{1/2}) \\ &= -nL'_{2,\boldsymbol{\theta}}(\tau_2)(\hat{H}_2(H_2^{-1}(1 - \varepsilon)) - (1 - \varepsilon))/h_2(H_2^{-1}(1 - \varepsilon)) + o_P(n^{1/2}). \end{aligned}$$

The derivation of  $T_3$  is similar, which finishes the proof.

This enables us to prove Theorem 2.1 :

**Proof of Theorem 2.1.** From the above Theorem A.2 together with the central limit theorem, the asymptotic normality of  $n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0)$  follows. Since  $E[g(\mathbf{Y}^h, \mathbf{y})] = 0$  (see p. 614 in van der Laan (1996)), we have that  $E(\mathbf{S}_n^g(\boldsymbol{\theta}_0)) = 0$  after integration by parts and using the fact that  $\mathbf{S}(\boldsymbol{\theta}_0) = 0$ . Hence,  $\mathbf{S}_n(\boldsymbol{\theta}_0)$  is asymptotically unbiased. The asymptotic variance of  $n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0)$  equals the variance of  $n^{-1/2}\mathbf{S}_n^g(\boldsymbol{\theta}_0)$  which is equal to the matrix  $\mathbf{A}$ .

The following is a technical lemma, known as the fourth Pitman condition, which will be needed in many of the proofs of the main results.

**Lemma A.3** *Assume (A5)(ii) and assume that the function  $H_{V,j}(y)$  ( $j = 1, 2$ ) is continuous in  $H_{V,j}^{-1}(1 - \varepsilon)$ . Then, for all  $B > 0$ ,*

$$\sup_{\|\mathbf{b}\| \leq B} |n^{-1/2}\mathbf{S}_n^g(\boldsymbol{\theta}_0 + \mathbf{b}n^{-1/2}) - n^{-1/2}\mathbf{S}_n^g(\boldsymbol{\theta}_0) + \mathbf{B}\mathbf{b}| \xrightarrow{P} 0. \quad (\text{A.1})$$

**Proof.** The function  $\mathbf{S}_n^g(\boldsymbol{\theta})$  naturally splits into three terms. Call them  $\mathbf{S}_n^{gj}(\boldsymbol{\theta})$  ( $j = 1, 2, 3$ ). It is easily seen that  $\sup_{\|\mathbf{b}\| \leq B} |n^{-1/2} \mathbf{S}_n^{gj}(\boldsymbol{\theta}_0 + \mathbf{b}n^{-1/2}) - n^{-1/2} \mathbf{S}_n^{gj}(\boldsymbol{\theta}_0)| = O_P(n^{-1/2})$  for  $j = 2, 3$ . Hence, we need to show that (A.1) holds with  $\mathbf{S}_n^g$  replaced by  $\mathbf{S}_n^{g1}$ . This will be done by applying Theorem 2 in Brown (1985) on  $-\mathbf{S}_n^{g1}(\boldsymbol{\theta})$  (which yields the same estimator  $\hat{\boldsymbol{\theta}}$ ). Therefore we need to verify his conditions (2), (3), (10) and (11). For condition (2) we refer to the proof of Theorem 2.1. Next note that the matrix of derivatives of  $\mathbf{S}_n^{g1}(\boldsymbol{\theta})$  is negative definite and hence the matrix of derivatives of  $-\mathbf{S}_n^{g1}(\boldsymbol{\theta})$  is positive definite, which shows that (11) is satisfied. Condition (3) holds because  $\mathbf{B}$  is positive definite. It remains to show that (10) holds, which we do by verifying the conditions of Theorem 3 in Brown (1985). We first prove that

$$\text{tr}\{n\text{Cov}[n^{-1}\mathbf{S}_n^{g1}(\boldsymbol{\theta}_0 + \mathbf{b}n^{-1/2}) - n^{-1}\mathbf{S}_n^{g1}(\boldsymbol{\theta}_0)]\} \rightarrow 0. \quad (\text{A.2})$$

Let

$$A_j(\mathbf{y}) = \frac{y_j - \theta_{0j} - b_j n^{-1/2}}{\|\mathbf{y} - \boldsymbol{\theta}_0 - \mathbf{b}n^{-1/2}\|} - \frac{y_j - \theta_{0j}}{\|\mathbf{y} - \boldsymbol{\theta}_0\|}$$

( $j = 1, 2$ ). We need to calculate

$$\begin{aligned} & n\text{Var} \left[ n^{-1} \sum_{i=1}^n \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} A_j(\mathbf{y}) dg(\mathbf{Y}_i, \mathbf{y}) \right] \\ &= E \left[ \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} A_j(\mathbf{y}) dg(\mathbf{Y}, \mathbf{y}) \right]^2 \\ &\leq 4 \left\{ A_j^2(\boldsymbol{\tau}) E[g^2(\mathbf{Y}, \boldsymbol{\tau})] + E \left[ \int_{-\infty}^{\tau_1} g(\mathbf{Y}, y_1, \tau_2) dA_j(y_1, \tau_2) \right]^2 \right. \\ &\quad \left. + E \left[ \int_{-\infty}^{\tau_2} g(\mathbf{Y}, \tau_1, y_2) dA_j(\tau_1, y_2) \right]^2 + E \left[ \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} g(\mathbf{Y}, \mathbf{y}) dA_j(\mathbf{y}) \right]^2 \right\}, \quad (\text{A.3}) \end{aligned}$$

where the last inequality follows from integration by parts and the fact that  $(\sum_{j=1}^k T_j)^p \leq k^{p-1} \sum_{j=1}^k T_j^p$  for any random variables  $T_1, \dots, T_k$ , from which it follows that  $E[(\sum_{j=1}^k T_j)^2] \leq k \sum_{j=1}^k E(T_j^2)$ . The first term of (A.3) is  $o(1)$  since  $A_j(\boldsymbol{\tau}) \rightarrow 0$ . The second term equals

$$4 \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_1} E(g(\mathbf{Y}, y_1, \tau_2) g(\mathbf{Y}, z_1, \tau_2)) dA_j(y_1, \tau_2) dA_j(z_1, \tau_2)$$

which tends to zero since  $\int_{-\infty}^{\tau_1} |dA_j(y_1, \tau_2)| \rightarrow 0$ . Analogously, the third term is  $o(1)$ . Also the last term of (A.3) has zero limit since  $\int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} |dA_j(\mathbf{y})| \rightarrow 0$ . This shows that condition (A.2) is satisfied. It remains to show that

$$\mathbf{B} = -\nabla E_{\theta_0}[n^{-1}\mathbf{S}_n^{g^1}(\boldsymbol{\theta})]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

where  $\nabla$  denotes differentiation with respect to the components of  $\boldsymbol{\theta}$ . The right hand side equals

$$\begin{aligned} & -\nabla \left[ \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} \frac{\mathbf{y} - \boldsymbol{\theta}}{\|\mathbf{y} - \boldsymbol{\theta}\|} dF(\mathbf{y}) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \int_{-\infty}^{\tau_1} \int_{-\infty}^{\tau_2} \left[ \frac{\mathbf{I}}{\|\mathbf{y} - \boldsymbol{\theta}_0\|} - \frac{1}{\|\mathbf{y} - \boldsymbol{\theta}_0\|^3} (\mathbf{y} - \boldsymbol{\theta}_0)(\mathbf{y} - \boldsymbol{\theta}_0)^t \right] dF(\mathbf{y}), \end{aligned}$$

which equals  $\mathbf{B}$ .

A first consequence of this result is the weak uniform consistency of  $\hat{\boldsymbol{\theta}}$  :

**Lemma A.4** *Assume (A1) – (A5). Then,*

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1).$$

**Proof.** This follows from Theorem 2 in Brown (1985). For the verification of the conditions stated in that theorem, we refer to the proof of Lemma A.3 where the same result was used.

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

**Proof of Theorem 2.2.** Since  $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1)$ , we can apply the fourth Pitman condition (Lemma A.3) on  $\mathbf{b} = n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$  yielding

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

Because  $\mathbf{S}_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ , it follows that  $\mathbf{S}_n^g(\hat{\boldsymbol{\theta}}) = o_P(n^{1/2})$  and hence

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



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

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

$$n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0) = n^{1/2} \int_{-\infty}^{\hat{H}_1^{-1}(1-\varepsilon)} \int_{-\infty}^{\hat{H}_2^{-1}(1-\varepsilon)} \frac{\mathbf{y} - \boldsymbol{\theta}_0}{\|\mathbf{y} - \boldsymbol{\theta}_0\|} d\hat{F}(\mathbf{y})$$

under  $H_1$  equals the asymptotic distribution under  $H_0$  of

$$\begin{aligned} & n^{1/2} \int_{-\infty}^{\hat{H}_1^{-1}(1-\varepsilon) + \frac{\gamma_1}{n^{1/2}}} \int_{-\infty}^{\hat{H}_2^{-1}(1-\varepsilon) + \frac{\gamma_2}{n^{1/2}}} \frac{\mathbf{y} - \boldsymbol{\theta}_0}{\|\mathbf{y} - \boldsymbol{\theta}_0\|} d\hat{F}(\mathbf{y} - n^{-1/2}\boldsymbol{\gamma}) \\ &= n^{1/2} \int_{-\infty}^{\hat{H}_1^{-1}(1-\varepsilon)} \int_{-\infty}^{\hat{H}_2^{-1}(1-\varepsilon)} \frac{\mathbf{y} - \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\gamma}}{\|\mathbf{y} - \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\gamma}\|} d\hat{F}(\mathbf{y}) \\ &= n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0 - n^{-1/2}\boldsymbol{\gamma}) \\ &= n^{-1/2}\mathbf{S}_n^g(\boldsymbol{\theta}_0 - n^{-1/2}\boldsymbol{\gamma}) + o_P(1) \\ &= n^{-1/2}\mathbf{S}_n^g(\boldsymbol{\theta}_0) + \mathbf{B}\boldsymbol{\gamma} + o_P(1) \end{aligned}$$

by the fourth Pitman condition. Since the latter converges to a bivariate normal distribution with mean  $\mathbf{B}\boldsymbol{\gamma}$  and variance matrix  $\mathbf{A}$ , the result follows. For the scale model the proof is similar : the asymptotic distribution of  $n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0)$  under  $H_1$  equals the asymptotic distribution under  $H_0$  of

$$\begin{aligned} & n^{1/2} \int_0^{\hat{H}_1^{-1}(1-\varepsilon)e^{\gamma n^{-1/2}}} \int_0^{\hat{H}_2^{-1}(1-\varepsilon)e^{\gamma n^{-1/2}}} \frac{\mathbf{y} - \boldsymbol{\theta}_0}{\|\mathbf{y} - \boldsymbol{\theta}_0\|} d\hat{F}(\mathbf{y}e^{-\gamma n^{-1/2}}) \\ &= n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0 e^{-\gamma n^{-1/2}}) \\ &= n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0 - \gamma\boldsymbol{\theta}_0 n^{-1/2}) + o_P(1) \\ &= n^{-1/2}\mathbf{S}_n^g(\boldsymbol{\theta}_0) + \gamma\mathbf{B}\boldsymbol{\theta}_0 + o_P(1), \end{aligned}$$

from which the result follows.

**Proof of Theorem 2.4.** Let

$$\begin{aligned}\Sigma &= \mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1} \\ \hat{\Sigma} &= \hat{\mathbf{B}}^{-1}\hat{\mathbf{A}}\hat{\mathbf{B}}^{-1}.\end{aligned}$$

Since  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{B}}$  are weakly consistent estimators for respectively  $\mathbf{A}$  and  $\mathbf{B}$ , we have that  $\hat{\Sigma}^{-1} - \Sigma^{-1} \xrightarrow{P} 0$  and hence, since  $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1)$ ,

$$n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t(\hat{\Sigma}^{-1} - \Sigma^{-1})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{P} 0,$$

and also

$$\frac{n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t(\hat{\Sigma}^{-1} - \Sigma^{-1})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}{n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t\Sigma^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)} \xrightarrow{P} 0$$

or equivalently

$$\frac{n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t\hat{\Sigma}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}{n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t\Sigma^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)} \xrightarrow{P} 1. \tag{A.4}$$

From Theorem 2.2 it follows that

$$n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t\mathbf{B}\mathbf{A}^{-1}\mathbf{B}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(2)$$

and hence using (A.4) and Slutsky's theorem,

$$n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(2).$$

This proves the result.

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

## References

- 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.
- Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.*, **76**, 312-319.
- Gini, C. and Galvani L. (1929). Di talune estensioni dei concetti di media ai caratteri qualitativi. *Metron*, **8**, 3-209.
- Gower, J. C. (1974). The mediancenter. *Applied Statist.*, **32**, 466-470.
- Haldane, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika*, **35**, 414-415.
- Hayford, J. F. (1902). What is the center of an area, or the center of a population? *J. Amer. Statist. Assoc.*, **8**, 47-58.
- Hettmansperger, T. P. and McKean, J. W. (1998). *Robust nonparametric statistical methods*. Arnold, London.
- 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 properties of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.*, **19**, 229-248.
- 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.
- 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.

van der Laan Support Points

