IMS Collections
A Festschrift for Jana Jurečková
Vol. 0 (0000) 1–6

arXiv: math.PR/0000000

On the estimation of cross-information quantities in rank-based inference

Delphine Cassart[1], **Marc Hallin**[1,*] and **Davy Paindaveine**[1,†]

Université Libre de Bruxelles

Abstract: Rank-based inference and, in particular, R-estimation, is a red thread running through Jana Jurečková's entire scientific career, starting with her dissertation in 1967, where she laid the foundations of an extension to linear regression of the R-estimation methods that had recently been proposed by Hodges and Lehmann [13]. Cross-information quantities in that context play an essential role. In location/regression problems, these quantities take the form $\int_0^1 \varphi(u)\varphi_g(u)\,\mathrm{d}u$ where φ is a score function and $\varphi_g(u) := g'(G^{-1}(u))/g(G^{-1}(u))$ is the log-derivative of the unknown actual underlying density g computed at the quantile $G^{-1}(u)$; in other models, they involve more general scores. Such quantities appear in the local powers of rank tests and the asymptotic variance of R-estimators. Estimating them consistently is a delicate problem that has been extensively considered in the literature. We provide here a new, flexible, and very general method for that problem, which furthermore applies well beyond the traditional case of regression models.

1. Introduction

1.1. Asymptotic linearity and the foundations of R-estimation

The 1969 volume of the *Annals of Mathematical Statistics* is rightly famous for two pathbreaking papers (Jurečková [15]; Koul [17]) that opened the door to R-estimation procedures in linear regression models. Both papers were their author's first publication. Both were addressing, with different mathematical tools, in slightly different contexts, and under different assumptions, the same essential problem: the uniform asymptotic linearity of residual rank-based statistics in a regression parameter.

The idea of using rank-based test statistics in order to construct point estimators and confidence regions had been proposed, in 1963, by Hodges and Lehmann [13], in the context of one- and two-sample location models. The potential applications of that idea in a much broader context were clear, and immediately triggered a

*Académie Royale de Belgique and CentER, Tilburg University. Research supported by the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (Deutsche Forschungsgemeinschaft) and a Discovery Grant of the Australian Research Council. The financial support and hospitality of ORFE and the Bendheim Center at Princeton University, where part of this work was completed, is gratefully acknowledged. Marc Hallin is also a member of ECORE, the association between CORE and ECARES.

†Research supported by a Mandat d'Impulsion Scientifique of the Fonds National de la Recherche Scientifique, Communauté française de Belgique. Davy Paindaveine is also a member of ECORE, the association between CORE and ECARES.

1 ECARES, Université Libre de Bruxelles, Avenue F.D. Roosevelt, 50, CP114, B-1050 Brussels, Belgium; e-mail: delphine.cassart@dexia.com; mhallin@ulb.ac.be; dpaindav@ulb.ac.be; url: http://homepages.ulb.ac.be/~dpaindav

AMS 2000 subject classifications: Primary 62G99; secondary 62G05, 62G10.

Keywords and phrases: sample, Rank tests, R-estimation, cross-information, local power, asymptotic variance.

surge of activity with the objective of extending the new technique to more general models. The analysis of variance case very soon was developed by Lehmann himself (Lehmann [23]; see also Sen [29]), very much along the same lines as in his original paper with Hodges. But the simple and multiple regression cases were considerably more difficult, the main obstacle to the desired result being a uniform asymptotic linearity property of the rank statistics to be used in the (regression) parameters. That result was more challenging than expected; it is missing, for instance, in Adichie [1]. It was successfully established, simultaneously and independently, in 1967, in two doctoral dissertations, one by Jana Jurečková (in Czech, defended in Prague; advisor Jaroslav Hájek), the other one by Hira Koul (defended in Berkeley; advisor Peter Bickel). Although essentially addressing the same issue, the two contributions (Jurečková [15]; Koul [17]) have little overlap: ranks and Hájek projection methods on one hand, signed-ranks and Billingsley-style weak convergence techniques on the other. Both got published in the same 1969 issue of the *Annals of Mathematical Statistics*.

Those uniform asymptotic linearity results paved the way for a complete theory of rank-based estimation in linear models and their extensions to parametric regression and time series, both linear and nonlinear —see the monographs by Puri and Sen [26], Jurečková and Sen [16], or Koul [18, 19] for systematic expositions.

This modest contribution to the subject is a tribute to Jana Jurečková's pioneering work in the domain.

1.2. Cross-information quantities

Denoting by $\underset{\sim}{Q}(\boldsymbol{\vartheta}_0)$ some rank-based test statistic for a two-sided null hypothesis of the form $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$, an R-estimator $\underset{\sim}{\boldsymbol{\vartheta}}$ of $\boldsymbol{\vartheta}$ is usually defined as a minimizer of $\underset{\sim}{Q}(\boldsymbol{\vartheta})$, that is, $\underset{\sim}{\boldsymbol{\vartheta}} := \text{argmin}_{\boldsymbol{\vartheta}}\, \underset{\sim}{Q}(\boldsymbol{\vartheta})$. Under appropriate regularity conditions, and irrespective of the model under study, the asymptotic performances of the R-estimator $\underset{\sim}{\boldsymbol{\vartheta}}$ and the related rank test typically are the same. More specifically, the local powers of rank tests are monotone functions of quantities of the form

$$\left(\int_0^1 \varphi(u)\varphi_g(u)\,\mathrm{d}u\right)^2, \tag{1.1}$$

whereas the related R-estimators are asymptotically normal, with asymptotic variances proportional to the inverse of the same quantity. Here φ is the score function defining the rank-based statistic $\underset{\sim}{Q}(\boldsymbol{\vartheta})$ from which the R-estimator is constructed, while, in the context of location and regression, $\varphi_g(u) := g'(G^{-1}(u))/g(G^{-1}(u))$ is the log-derivative of the unknown actual underlying density g (with distribution function G) of the error terms underlying the model, computed at $G^{-1}(u)$. All usual score functions φ themselves being of the form φ_f for some reference density f, the integral in (1.1) generally is of the form

$$\mathcal{J}(f;g) := \int_0^1 \varphi_f(u)\,\varphi_g(u)\,\mathrm{d}u = \int_{-\infty}^{\infty} \frac{f'(F^{-1}(G(z)))}{f(F^{-1}(G(z)))}\,\frac{g'(z)}{g(z)}\,g(z)\,\mathrm{d}z.$$

Under that form, and since

$$\mathcal{I}_f := \mathcal{J}(f;f) = \int_{-\infty}^{\infty}\left(\frac{f'(z)}{f(z)}\right)^2 f(z)\,\mathrm{d}z \text{ and } \mathcal{I}_g := \mathcal{J}(g;g) = \int_{-\infty}^{\infty}\left(\frac{g'(z)}{g(z)}\right)^2 g(z)\,\mathrm{d}z$$

are Fisher information quantities (for location), $\mathcal{J}(f;g)$ clearly can be interpreted as a *cross-information quantity*, which explains the terminology and the notation we are using throughout, although φ_f and φ_g in the sequel need not be log-derivatives of probability densities.

That relation between rank tests and R-estimators extends to the multiparameter case, with information and cross-information quantities entering the definition of information and cross-information *matrices*. It also extends to more general models, much beyond the case of linear regression, where information and cross-information quantities still take the form (1.1), but involve scores φ_f and φ_g that are not *location scores* anymore; the notation $\mathcal{J}(g)$ will be used in a generic way for an integral of the form (1.1) where φ is the score of the rank statistic under study, and φ_g the log-derivative of the unknown actual density g with respect to the appropriate parameter of interest.

1.3. One-step R-estimation

An alternative to the classical Hodges–Lehmann argmin definition of an R-estimator was considered recently, for the estimation of the shape matrix of elliptical observations, by Hallin, Oja, and Paindaveine (2006). That method, which is directly connected to Le Cam's one-step approach to estimation problems, actually extends to a very broad range of uniformly locally asymptotically normal (ULAN) models, and is based on the local linearization of a rank-based version of the central sequence of the family.

Such a linearization, in a sense, revives, in the context of Le Cam's asymptotic theory of statistical experiments, an old idea that goes back to van Eeden and Kraft [31] and Antille [2]. The same idea also has been exploited by McKean and Hettmansperger [24], still in the traditional linear model setting, and in the slightly different approach initiated by Jaeckel [14] (which involves the argmin of a function that is not purely rank-based).

One-step estimators avoid some of the computational problems related with argmins of discrete-valued and possibly non-convex objective functions of (in the multiparameter case) several variables. Under their original form (as proposed by van Eeden and Kraft), however, they fail to achieve the same optimality bounds (parametric or nonparametric) as their argmin counterparts. McKean and Hettmansperger [24], in the context of linear models with symmetric noise, and Hallin, Oja, and Paindaveine [12], in the context of shape matrix estimation, solve that problem by introducing an estimated cross-information factor in the linearization step. Although different from (1.1) (since the scores φ_f and φ_g are those related to shape parameters), the cross-information quantity for shape plays exactly the same role in the asymptotic covariance matrix of R-estimators of shape as (1.1) does in the asymptotic variance of R-estimators of location or in the asymptotic covariance matrix of R-estimators of regression coefficients.

Whether entering as an essential ingredient in some one-step form of estimation or not, cross-information quantities explicitly appear in the asymptotic variances of R-estimators, and thus need to be estimated.

Now, the trouble with cross-information quantities is that, being expectations, under the unspecified actual density g, of a function which itself depends on that unknown g, they are not easily estimated. That difficulty may well be one of the main reasons why R-estimation, despite all its attractive theoretical features, never really made its way to everyday practice.

1.4. Estimation of cross-information quantities

A vast literature has been devoted to the problem of estimating (1.1) in the context of linear models with i.i.d. errors (except for Hallin, Oja, and Paindaveine 2006, more general cross-information quantities, to the best of our knowledge, have not been considered so far). Four approaches, mainly, have been investigated.

(a) McKean and Hettmansperger [24] estimate $\mathcal{J}(f;g)$ as the ratio of a $(1-\alpha)$ confidence interval to the corresponding standard normal interquantile range; that idea can be traced back to Lehmann [23] and Sen [29], and requires the arbitrary choice of a confidence level $(1-\alpha)$, which has no consequence in the limit, but for finite n may have quite an impact (Aubuchon and Hettmansperger [3] in the same context propose using the interquartile ranges or median absolute deviations from the median). A similar idea, along with powerful higher-order methods leading to most interesting distributional results, is exploited by Omelka [25], but requires the same choice of a confidence level $(1-\alpha)$.

(b) Some other authors (Antille [2]; Jurečková and Sen [16], p. 321) rely on the asymptotic linearity property of rank statistics, by evaluating the consequence of a $O(n^{-1/2})$ perturbation of $\boldsymbol{\vartheta}_0$ on the test statistic for $\mathcal{H}_0 \colon \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$. This again involves an arbitrary choice—that of the amplitude $cn^{-1/2}$, $c \in \mathbb{R}_0$ (in the multiparameter case, $\mathbf{c}n^{-1/2}$, $\mathbf{c} \in \mathbb{R}^k \setminus \{\mathbf{0}\}$) of the perturbation. Again, different values of c or $\mathbf{c}$ lead, for finite n, to completely different estimators; asymptotically, this has no impact, but finite-n results can be quite dramatically affected.

(c) More sophisticated methods involving window or kernel estimates of g—hence performing poorly under small and moderate sample sizes—have been considered, for Wilcoxon scores, by Schuster [27] and Schweder [28] (see also Cheng and Serfling [7]; Koul, Sievers and McKean [20]; Bickel and Ritov [5]; Fan [8] and, in a more general setting, Section 4.5 of Koul [19]). Instead of a confidence level $(1-\alpha)$ or a deviation $\mathbf{c}$, a kernel and a bandwidth are to be selected. Density estimation methods, moreover, are kind of antinomic to the spirit of rank-based methods: if estimated densities are to be used, indeed, using them all the way by considering semiparametric tests based on estimated scores (in the spirit of Bickel et al. [4]) seems more coherent than considering ranks.)

(d) Finally, jacknifing and the bootstrap also have been utilized in this context: see George and Osborne [9] and George et al. [10] for an investigation of that approach and some empirical findings.

The approach proposed in Hallin, Oja, and Paindaveine [12] is of a different nature. It is based on the asymptotic linearity of a rank-based central sequence, hence requires *uniform local asymptotic normality in the Le Cam sense*, and consists in solving a local linearized likelihood equation. It does not involve any arbitrary choices, and, irrespective of the dimension of the parameter of interest, its implementation involves one-dimensional optimization only. However, it only can handle information quantities entering as a scalar factor in the information matrix of a given model, or, in the case of a block-diagonal information matrix, in some diagonal block thereof. This places a restriction on the quantities to be estimated, and rules out some cases, such as the information quantity for skewness derived in Cassart et al. [6]. In this contribution, we propose a generalization of the Hallin, Oja, and Paindaveine method that does not require uniform local asymptotic normality, and can accomodate much more general situations, including that of Cassart et al. [6].

2. Consistent estimation of cross-information quantities

Let $\mathcal{P}^{(n)} := \{\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g} |\ \boldsymbol{\vartheta} \in \boldsymbol{\Theta}, g \in \mathcal{F}\}$ be a family (actually, a sequence of them, indexed by $n \in \mathbb{N}$) of probability measures over some observation space (usually, $\mathbb{R}^n$, equipped with its Borel σ-field), indexed by a k-dimensional parameter $\boldsymbol{\vartheta} \in \mathbb{R}^k$ and a univariate probability density g; $\boldsymbol{\vartheta}$ ranges over some open subset $\boldsymbol{\Theta}$ of $\mathbb{R}^k$, and g over some broad class of densities $\mathcal{F}$. Associated with that observation, assume that there exists an n-tuple $(Z^{(n)}_1(\boldsymbol{\vartheta}), \ldots, Z^{(n)}_n(\boldsymbol{\vartheta}))$ of *residuals* such that $Z^{(n)}_1(\boldsymbol{\vartheta}_0), \ldots, Z^{(n)}_n(\boldsymbol{\vartheta}_0)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$ are independent and identically distributed with density g iff $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$.

Denoting by $R^{(n)}_i(\boldsymbol{\vartheta})$ the rank of $Z^{(n)}_i(\boldsymbol{\vartheta})$ among $Z^{(n)}_1(\boldsymbol{\vartheta}), \ldots, Z^{(n)}_n(\boldsymbol{\vartheta})$, the vector $\mathbf{R}^{(n)}(\boldsymbol{\vartheta}) := (R^{(n)}_1(\boldsymbol{\vartheta}), \ldots, R^{(n)}_n(\boldsymbol{\vartheta}))$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$ is uniformly distributed over the $n!$ permutations of $\{1, \ldots, n\}$, irrespective of g—a distribution-freeness property which serves as the starting point of rank tests and R-estimation of $\boldsymbol{\vartheta}$ in the family $\mathcal{P}^{(n)}$.

Our goal is to estimate consistently a cross-information quantity $\mathcal{J}(g) > 0$ that enters the picture through the following assumption.

Assumption (A) There exists a sequence $\underset{\sim}{\mathbf{S}}^{(n)}(\boldsymbol{\vartheta})$ of k-dimensional $\mathbf{R}^{(n)}(\boldsymbol{\vartheta})$-measurable statistics such that, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$,

(i) $\underset{\sim}{\mathbf{S}}^{(n)}(\boldsymbol{\vartheta})$, $n \in \mathbb{N}$ is uniformly tight and asymptotically bounded away from the origin; more precisely, for all $\varepsilon > 0$, there exist $\delta_\varepsilon > 0$, M_ε and N_ε such that, for all $n \geq N_\varepsilon$,

$$\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}\Big[\delta_\varepsilon \leq \|\underset{\sim}{\mathbf{S}}^{(n)}(\boldsymbol{\vartheta})\| \leq M_\varepsilon\Big] \geq 1 - \varepsilon$$

(uniformity here is with respect to n, not $\boldsymbol{\vartheta}$);

(ii) there exists a continuous mapping $\boldsymbol{\vartheta} \mapsto \boldsymbol{\Upsilon}^{-1}(\boldsymbol{\vartheta})$, where $\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\vartheta})$ is a full-rank $k \times k$ matrix such that

$$\underset{\sim}{\mathbf{S}}^{(n)}(\boldsymbol{\vartheta} + n^{-1/2}\mathbf{t}^{(n)}) = \underset{\sim}{\mathbf{S}}^{(n)}(\boldsymbol{\vartheta}) - \mathcal{J}(g)\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\vartheta})\mathbf{t}^{(n)} + o_{\mathrm{P}}(1) \quad \text{as } n \to \infty \tag{2.1}$$

for any bounded sequence $\mathbf{t}^{(n)} \in \mathbb{R}^k$.

We will also need

Assumption (B) A root-n consistent estimator $\hat{\boldsymbol{\vartheta}}^{(n)}$ of $\boldsymbol{\vartheta}$ is available, such that, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$, $\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)})$ is asymptotically bounded away from zero: for all $\varepsilon > 0$, there exist δ_ε and N_ε such that

$$\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}\Big[\|\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)})\| \geq \delta_\varepsilon\Big] \geq 1 - \varepsilon$$

for all $n \geq N_\varepsilon$.

Note that part (i) of Assumption (A) is rather mild, as it is satisfied as soon as $\underset{\sim}{\mathbf{S}}^{(n)}(\boldsymbol{\vartheta})$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$ is converging in distribution to a random vector that has no atom at the origin. As for part (ii), it does not require the asymptotic linearity (2.1) to be uniform. Similarly, Assumption (B) requires that $\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)})$ asymptotically has no atom at $\mathbf{0}$. The statistic $\underset{\sim}{\mathbf{S}}^{(n)}$ indeed is to provide, via its local behavior (2.1), an estimator for $\mathcal{J}(g)$—not a test statistic, nor (through some estimating equation) an estimator for $\boldsymbol{\vartheta}$: Assumption (B) thus explicitly rules out an estimator that would be obtained as $\hat{\boldsymbol{\vartheta}}^{(n)} = \text{argmin}_{\boldsymbol{\vartheta}} \| \underset{\sim}{\mathbf{S}}^{(n)}(\boldsymbol{\vartheta})\|$.

In order to control for the uniformity of local behaviors, a discretized version $\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}$ of $\hat{\boldsymbol{\vartheta}}^{(n)}$ will be considered in theoretical asymptotic statements. Such a version can be obtained, for instance, by letting

$$\left(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}\right)_i := (cn^{1/2})^{-1}\text{sign}\Big((\hat{\boldsymbol{\vartheta}}^{(n)})_i\Big)\lceil cn^{1/2}|(\hat{\boldsymbol{\vartheta}}^{(n)})_i|\rceil, i = 1, \ldots, k$$

for some arbitrary discretization constant $c > 0$. This discretization trick, which is due to Le Cam, is quite standard in the context of one-step estimation. While retaining root-n consistency, discretized estimators indeed enjoy the important property of *asymptotic local discreteness*, that is, as $n \to \infty$, they only take a bounded number of distinct values in $\boldsymbol{\vartheta}$-centered balls with $O(n^{-1/2})$ radius. In fixed-n practice, however, such discretizations are irrelevant (one cannot work with an infinite number of decimal values, and c can be chosen arbitrarily large). The reason why discretization is required in asymptotic statements is that (see, for instance, Lemma 4.4 of Kreiss [21]), (2.1) then also holds with $n^{1/2}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#} - \boldsymbol{\vartheta})$ substituted for $\mathbf{t}^{(n)}$, yielding

$$\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) = \underset{\sim}{\mathbf{S}}^{(n)}(\boldsymbol{\vartheta}) - n^{1/2}\mathcal{J}(g)\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\vartheta})(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#} - \boldsymbol{\vartheta}) + o_{\mathrm{P}}(1) \tag{2.2}$$

as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$. This stochastic form of (2.1) in a sense takes care of uniformity problems.

We now describe the construction of our estimator of $\mathcal{J}(g)$. For any $\lambda \in \mathbb{R}^+$, define

$$\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda} := \hat{\boldsymbol{\vartheta}}^{(n)}_{\#} + n^{-1/2}\lambda\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\, \underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}). \tag{2.3}$$

When λ ranges over the positive real line, $\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda}$ for fixed n thus moves, monotonically with respect to λ, along a half-line with origin $\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}$. Note that any $\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda}$, once discretized into $\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda\#}$, provides a new root-$n$ consistent and asymptotically locally discrete estimator of $\boldsymbol{\vartheta}$ to which (2.2) applies. It follows that

$$\underset{\sim}{\mathbf{S}}^{(n)}(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda\#}) - \underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) = -\lambda\mathcal{J}(g)\, \underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}_{\#}) + o_{\mathrm{P}}(1), \tag{2.4}$$

still as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$. Moreover, $\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda\#}$ also can serve as the starting point for an iteration of the type (2.3), yielding, for any $\mu \in \mathbb{R}^+$, a further root-n consistent estimator of the form

$$\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda\#} + n^{-1/2}\mu\boldsymbol{\Upsilon}(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda\#})\, \underset{\sim}{\mathbf{S}}^{(n)}(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda\#}). \tag{2.5}$$

From (2.4) we thus obtain, for all $\lambda > 0$,

$$\underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#}) \tag{2.6}$$

$$= (1-\lambda\mathcal{J}(g))\,\underset{\sim}{\mathbf{S}}^{(n)\prime}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#}) + o_{\mathrm{P}}(1)$$

$$= (1-\lambda\mathcal{J}(g))\,\underset{\sim}{\mathbf{S}}^{(n)\prime}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\boldsymbol{\Upsilon}'(\boldsymbol{\vartheta})\boldsymbol{\Upsilon}(\boldsymbol{\vartheta})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#}) + o_{\mathrm{P}}(1). \tag{2.7}$$

The intuition behind our method lies in the fact that (2.6), which is the scalar product of the increments in (2.3) and (2.5), is, up to $o_{\mathrm{P}}(1)$'s, a decreasing linear function (2.7) of λ: since $\boldsymbol{\Upsilon}$ has full-rank, the quadratic form in (2.7) indeed is positive definite. That function takes positive values for λ close to zero, and changes sign at $\lambda = \mathcal{J}^{-1}(g)$.

Let therefore (c is an arbitrary discretization constant that plays no role in practical implementations)

$$\lambda^{(n)}_{-} := \min_{\ell}\Big\{\lambda_\ell := \frac{\ell}{c}$$

$$\text{such that}\;\; \underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda_{\ell+1}\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda_{\ell+1}\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#}) < 0\Big\} \tag{2.8}$$

and $\lambda^{(n)}_{+} := \lambda^{(n)}_{-} + \frac{1}{c}$. Defining $\mathcal{J}^{(n)}(g) := (\lambda^{(n)})^{-1}$, where $\lambda^{(n)}$ is based on a linear interpolation between $\lambda^{(n)}_{-}$ and $\lambda^{(n)}_{+}$, namely

$$\lambda^{(n)} := \lambda^{(n)}_{-} + \frac{(\lambda^{(n)}_{+}-\lambda^{(n)}_{-})\,\underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{-}\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{-}\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})}{[\underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{-}\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{-}\#}) - \underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{+}\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{+}\#})]\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})}$$

$$= \lambda^{(n)}_{-} + \frac{1}{c}\,\frac{\underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{-}\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{-}\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})}{[\underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{-}\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{-}\#}) - \underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{+}\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda^{(n)}_{+}\#})]\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#})},$$

we have the following result (see the Appendix for the proof).

Proposition 2.1. *Let Assumptions (A) and (B) hold. Then* $\mathcal{J}^{(n)}(g) = \mathcal{J}(g) + o_{\mathrm{P}}(1)$ *as* $n\to\infty$*, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$.

As already mentioned, discretizing the estimators is a mathematical device which is needed in the proof of asymptotic results but makes little sense in a fixed-n practical situation, as a very large discretization constant can be chosen. In practice, still assuming that Assumptions (A) and (B) hold, we recommend directly computing $(\mathcal{J}^{(n)}(g))^{-1}$ as

$$(\mathcal{J}^{(n)}(g))^{-1} := \lambda^{(n)} := \inf\Big\{\lambda \text{ such that } \underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}{}^{(n)})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}{}^{(n)}) < 0\Big\}.$$

Indeed, for large values of the discretization constant c, $\hat{\boldsymbol{\vartheta}}{}^{(n)}_{\#}$ and $\hat{\boldsymbol{\vartheta}}{}^{(n)}$ are arbitrarily close, as well as $\lambda^{(n)}_{-}$ and $\lambda^{(n)}_{+}$ defined in (2.8).

3. Conclusion

Proposition 2.1 establishes the consistency of the proposed estimator of cross-information quantities. Consistency indeed is the only property required from estimators of cross-information quantities—be it in the construction of a one-step R-estimator $\underset{\sim}{\boldsymbol{\vartheta}}$ of $\boldsymbol{\vartheta}$ or in the estimation of its asymptotic variance (with the purpose, for instance, of computing asymptotically valid confidence regions for $\boldsymbol{\vartheta}$). We do not provide (and, to the best of our knowledge, nobody, in that context, ever has) any indication about the consistency rates and asymptotic distribution of $\mathcal{J}^{(n)}(g)$ as an estimator of $\mathcal{J}(g)$—even less about its optimality. While they have no impact on the asymptotic behavior of $\underset{\sim}{\boldsymbol{\vartheta}}$, the choices of (i) the sequence of rank-based statistics $\underset{\sim}{\mathbf{S}}^{(n)}(\boldsymbol{\vartheta})$, (ii) the initial estimator $\hat{\boldsymbol{\vartheta}}^{(n)}$, and (iii) the discretization constant c are likely to affect its finite-sample performances. However, the magnitude of such effects can be expected to be negligible when compared to the estimation error $(\underset{\sim}{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})$ itself.

Appendix A: Proof of Proposition 2.1

To start with, let us show that $\lambda_-^{(n)}$, defined in (2.8), hence also $\lambda_+^{(n)}$, is $O_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$. Assume therefore it is not: then, there exist $\epsilon > 0$ and a sequence $n_i \uparrow \infty$ such that, for all $L \in \mathbb{R}$ and i, $\mathrm{P}^{(n_i)}_{\boldsymbol{\vartheta};g}[\lambda_-^{(n_i)} > L] > \epsilon$. This implies, for arbitrarily large L, that

$$\mathrm{P}^{(n_i)}_{\boldsymbol{\vartheta};g}\Big[\underset{\sim}{\mathbf{S}}^{(n_i)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n_i)}_{L\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n_i)}_{L\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n_i)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n_i)}(\hat{\boldsymbol{\vartheta}}^{(n_i)}_{\#}) > 0\Big] > \epsilon,$$

hence, in view of (2.7),

$$\mathrm{P}^{(n_i)}_{\boldsymbol{\vartheta};g}\Big[(1 - L\mathcal{J}(g))\,\underset{\sim}{\mathbf{S}}^{(n_i)\prime}(\hat{\boldsymbol{\vartheta}}^{(n_i)}_{\#})\boldsymbol{\Upsilon}'(\boldsymbol{\vartheta})\boldsymbol{\Upsilon}(\boldsymbol{\vartheta})\,\underset{\sim}{\mathbf{S}}^{(n_i)}(\hat{\boldsymbol{\vartheta}}^{(n_i)}_{\#}) + \zeta^{(n_i)} > 0\Big] > \epsilon$$

for all i, where $\zeta^{(n)}$, $n \in \mathbb{N}$ is some $o_{\mathrm{P}}(1)$ sequence. For $L > (\mathcal{J}(g))^{-1}$, this entails, for all i,

$$\mathrm{P}^{(n_i)}_{\boldsymbol{\vartheta};g}\Big[0 < \underset{\sim}{\mathbf{S}}^{(n_i)\prime}(\hat{\boldsymbol{\vartheta}}^{(n_i)}_{\#})\boldsymbol{\Upsilon}'(\boldsymbol{\vartheta})\boldsymbol{\Upsilon}(\boldsymbol{\vartheta})\,\underset{\sim}{\mathbf{S}}^{(n_i)}(\hat{\boldsymbol{\vartheta}}^{(n_i)}_{\#}) < |\zeta^{(n_i)}|/(L\mathcal{J}(g) - 1)\Big] > \epsilon,$$

which contradicts Assumption (B) that $\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)})$ is bounded away from zero. It follows that $\lambda_-^{(n)}$ is $O_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$; actually, we have shown the stronger result that, for any $L > (\mathcal{J}(g))^{-1}$, $\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}[\lambda_-^{(n)} > L] = 0$.

In view of Assumption (B), for all $\eta > 0$, there exist $\delta_\eta > 0$ and an integer N_η such that

$$\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}\left[\underset{\sim}{\mathbf{S}}^{(n)\prime}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) \geq \delta_\eta\right] \geq 1 - \eta/2$$

for all $n \geq N_\eta$. In view of (2.4), the fact that $\lambda_-^{(n)}$ and $\lambda_+^{(n)}$ are $O_{\mathrm{P}}(1)$, and Assumption (A), for all $\eta > 0$ and $\varepsilon > 0$, there exists an integer $N_{\varepsilon,\delta} \geq N_\eta$ such that, for all $n \geq N_{\varepsilon,\delta}$ (with $\lambda_\pm^{(n)}$ standing for either $\lambda_-^{(n)}$ or $\lambda_+^{(n)}$),

$$\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}\Bigg[(1 - \mathcal{J}(g)\lambda_\pm^{(n)})\,\underset{\sim}{\mathbf{S}}^{(n)\prime}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) \in \Big[\underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda_\pm\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}{}^{(n)}_{\lambda_\pm\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) \pm \varepsilon\Big]\Bigg] \geq 1 - \eta/2.$$

It follows that for all $\eta > 0$, $\varepsilon > 0$ and $n \geq N_{\varepsilon,\delta}$, letting $\delta = \delta_\eta$,

$$\begin{aligned} \mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}\Big[A^{(n)}_{\varepsilon,\delta}\Big] := \mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}\Bigg[&(1-\mathcal{J}(g)\lambda^{(n)}_{\pm})\,\underset{\sim}{\mathbf{S}}^{(n)\prime}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) \\ &\in \Big[\,\underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda_\pm\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda_\pm\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) \pm \varepsilon\Big] \\ &\text{and } \underset{\sim}{\mathbf{S}}^{(n)\prime}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) \geq \delta\Bigg] \\ &\geq 1-\eta. \end{aligned}$$

Next, denote by $\hat{D}^{(n)}$, $D^{(n)}$ and $D^{(n)}_{\pm}$ the graphs of the mappings

$$\begin{aligned} \lambda \quad \mapsto \quad & \underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda_-\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda_-\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) \\ & -c(\lambda-\lambda_-)[\,\underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda_-\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda_-\#}) - \underset{\sim}{\mathbf{S}}^{(n)\prime}(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda_+\#})\boldsymbol{\Upsilon}'(\underset{\sim}{\boldsymbol{\vartheta}}^{(n)}_{\lambda_+\#})] \\ & \qquad \times\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}), \\ \lambda \quad \mapsto \quad & (1-\mathcal{J}(g)\lambda)\,\underset{\sim}{\mathbf{S}}^{(n)\prime}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}'(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}), \end{aligned}$$

and

$$\lambda \quad \mapsto \quad (1-\mathcal{J}(g)\lambda)\,\underset{\sim}{\mathbf{S}}^{(n)\prime}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}'(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\boldsymbol{\Upsilon}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#})\,\underset{\sim}{\mathbf{S}}^{(n)}(\hat{\boldsymbol{\vartheta}}^{(n)}_{\#}) \pm \varepsilon,$$

respectively. These graphs take the form of four random straight lines, intersecting the horizontal axis at $\lambda^{(n)}$ (our estimator of $(\mathcal{J}^{(n)}(g))^{-1}$), $\lambda_0 := (\mathcal{J}(g))^{-1}$, λ_0^+ and λ_0^-, respectively. Since $D^{(n)}_{\pm}$ and $D^{(n)}$ are parallel, with a negative slope, we have that

$$\lambda_0^- \leq \lambda_0 \leq \lambda_0^+.$$

Under $A^{(n)}_{\varepsilon,\delta}$, that common slope has absolute value at least $\mathcal{J}(g)\delta$, which implies that

$$\lambda_0^+ - \lambda_0^- \leq 2\varepsilon/\mathcal{J}(g)\delta.$$

Still under $A^{(n)}_{\varepsilon,\delta}$, for λ values between $\lambda^{(n)}_-$ and $\lambda^{(n)}_+$, $\hat{D}^{(n)}$ is lying between $D^{(n)}_-$ and $D^{(n)}_+$, which entails

$$\lambda_0^- \leq \lambda^{(n)} \leq \lambda_0^+.$$

Summing up, for all $\eta > 0$ and $\varepsilon > 0$, there exist $\delta = \delta_\eta > 0$, and $N = N_{\varepsilon\mathcal{J}(g)\delta/2,\delta}$ such that, for any $n \geq N$, with $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g}$ probability larger than $1-\eta$,

$$|\lambda^{(n)} - \lambda_0| \leq \lambda_0^+ - \lambda_0^- \leq \varepsilon. \qquad \square$$

Acknowledgement. We gratefully acknowledge the insightful comments by two anonymous referees.

References

[1] Adichie, J. N. (1967). Estimates of regression parameters based on rank tests. *Annals of Mathematical Statistics* **38** 894–904.

[2] Antille, A. (1974). A linearized version of the Hodges–Lehmann estimator. *Annals of Statistics* **2** 1308–1313.

[3] Aubuchon, J. C. and Hettmansperger, T.P. (1984). A note on the estimation of the integral of $f^2(x)$. *Journal of Statistical Planning and Inference* **9** 321–331.

[4] Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Statistical Inference for Semiparametric Models*, Johns Hopkins University Press, Baltimore.

[5] Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives. *Sankhya A* **50** 381–393.

[6] Cassart, D., Hallin, M., and Paindaveine, D. (2010). A class of optimal signed-rank tests for symmetry. Submitted.

[7] Cheng, K. F. and Serfling, R. J. (1981). On estimation of a class of efficiency-related parameters. *Scandinavian Actuarial Journal* **8** 83–92.

[8] Fan, J. (1991). On the estimation of quadratic functionals. *Annals of Statistics* **19** 1273–1294.

[9] George, K. J. and Osborne, M. (1990). The efficient computation of linear rank statistics. *Journal of Statistical Computation and Simulation* **35** 227–237.

[10] George, K. J., McKean, J. W., Schucany, W.R., and Sheather, S. J. (1995). A comparison of confidence intervals from R-estimators in regression. *Journal of Statistical Computation and Simulation* **53** 13–22.

[11] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

[12] Hallin, M., Oja, H., and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape: II Optimal R-estimation of shape. *Annals of Statistics* **34** 2757–2789.

[13] Hodges, J. L., Jr. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Annals of Mathematical Statistics* **34** 598–611.

[14] Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics* **43** 1449–1458.

[15] Jurečková, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *Annals of Mathematical Statistics* **40** 1889–1900.

[16] Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations*. New York: Wiley.

[17] Koul, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Annals of Mathematical Statistics* **40** 1950–1979.

[18] Koul, H. L. (1992). *Weighted Empiricals and Linear Models*, IMS lecture Notes-Monograph **21**, Institute of Mathematical Statitics.

[19] Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd edition. New York: Springer Verlag.

[20] Koul, H. L., Sievers, G. L., and McKean, J. W. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scandinavian Journal of Statistics* **14** 131–141.

[21] Kreiss, J.-P. (1987). On adaptative estimation in stationary ARMA processes. *Annals of Statistics* **15** 112–133.

[22] Le Cam, L. M. (1986). *Asymptotic Methods in Statistical Decision Theory.* New-York: Springer-Verlag.

[23] Lehmann, E. L. (1963). Nonparametric confidence intervals for a shift parameter. *The Annals of Mathematical Statistics* **34** 1507–1512.

[24] McKean, J. W. and Hettmansperger, T. P. (1978). A robust analysis of the general linear model based on one-step R-estimates. *Biometrika* **65** 571–579.

[25] Omelka, M. (2008). Comparison of two types of confidence intervals based on Wilcoxon-type R-estimators. *Statistics and Probability Letters* **78** 3366–3372.

[26] Puri, M. L. and Sen, P. K. (1985). *Nonparametric Methods in General Linear Models.* New York: J. Wiley.

[27] Schuster, E. (1974). On the rate of convergence of an estimate of a functional of a probability density. *Scandinavian Actuarial Journal* **1** 103–107.

[28] Schweder, T. (1975). Window estimation of the asymptotic variance of rank estimators of location. *Scandinavian Journal of Statistics* **2** 113–126.

[29] Sen, P. K. (1966). On a distribution-free method of estimating asymptotic efficiency of a class of nonparametric tests. *The Annals of Mathematical Statistics* **37** 1759–1770.

[30] van Eeden, C. (1972). An analogue, for signed-rank statistics, of Jurečková's asymptotic linearity theorem for rank statistics. *The Annals of Mathematical Statistics* **43** 791–802.

[31] van Eeden, C. and Kraft, C. H. (1972). Linearized rank estimates and signed-rank estimates for the general linear hypothesis. *The Annals of Mathematical Statistics* **43** 42–57.