Multivariate Spatial U-Quantiles: a Bahadur-Kiefer Representation, a Theil-Sen Estimator for Multiple Regression, and a Robust Dispersion Estimator

Weihua Zhou[1]

University of North Carolina at Charlotte

and

Robert Serfling[2]

University of Texas at Dallas

Final revision for *Journal of Statistical Planning and Inference*

May 2007

[1]Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA. Email: wzhou2@email.uncc.edu.

[2]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: serfling@utdallas.edu. Website: www.utdallas.edu/~serfling.

Abstract

A leading multivariate extension of the univariate quantiles is the so-called "spatial" or "geometric" notion, for which sample versions are highly robust and conveniently satisfy a Bahadur-Kiefer representation. Another extension of univariate quantiles has been to univariate U-quantiles, on the basis of which, for example, the well-known Hodges-Lehmann location estimator has a natural formulation. Generalizing both extensions, we introduce multivariate spatial U-quantiles and develop a corresponding Bahadur-Kiefer representation. New statistics based on spatial U-quantiles are presented for nonparametric estimation of multiple regression coefficients, extending the classical Theil-Sen nonparametric simple linear regression slope estimator, and for robust estimation of multivariate dispersion. Some other applications are mentioned as well.

AMS 2000 Subject Classification: Primary 62G99 Secondary 60F15.

Key words and phrases: generalized spatial quantiles; multivariate analysis; nonparametric methods; dispersion estimation; Theil-Sen estimator; strong convergence.

1 Introduction and preliminaries

Much of univariate nonparametric inference is very naturally *quantile-based*, and in recent years various notions of *multivariate quantiles* have been proposed (see Serfling (2002a) for a partial review). A leading choice is the *spatial* quantile function (Dudley and Koltchinskii, 1992, and Chaudhuri, 1996) for whose sample version important Bahadur and Bahadur-Kiefer representations have been developed independently in Chaudhuri (1996) and Koltchinskii (1994a), respectively. These extend to spatial quantiles the now classical results (Bahadur, 1966, Kiefer, 1967, 1970) for univariate quantiles (see also Serfling, 1980, Deheuvels and Mason, 1990, Niemiro, 1992, and Arcones and Mason, 1997). Such representations characterize (uniformly over quantiles in the Kiefer version) the remainder in approximating the estimation error between a population quantile and its sample version by a convenient linear term.

Another extension of the original Bahadur result has been to *U-quantiles* (Choudhury and Serfling, 1988, and Arcones, 1996), that is, to quantiles of the $n_{(m)} = n(n-1)\cdots(n-m+1)$ ordered evaluations $h(X_{i_1}, \ldots, X_{i_m})$ of a real-valued "kernel" $h(x_1, \ldots, x_m)$, taken over a sample $X_1, \ldots, X_n$ (from an arbitrary space) and m-tuples $(i_1, \ldots, i_m)$ of distinct indices from $\{1, \ldots, n\}$. A key example is

$$h(x_1, \ldots, x_m) = \frac{x_1 + \cdots + x_m}{m}, \tag{1}$$

which combined with the use of the *median* quantile yields a *generalized Hodges-Lehmann location estimator*: the median of m-wise averages of the sample observations, for given choice of integer $m \geq 1$. This includes for $m = 1$ the usual median and for $m = 2$ the classical Hodges-Lehmann location estimator (Hodges and Lehmann, 1963, and Geertsema, 1970).

In the present paper, combining the above two lines of development, we introduce *vector-valued* kernels $\boldsymbol{h}(x_1, \ldots, x_m)$ and define corresponding *spatial U-quantiles*. In particular, this includes the multivariate form of (1), already treated in Chaudhuri (1992), where a Bahadur representation was established for the spatial median of m-wise averages of sample observations in $\mathbb{R}^d$. As other examples of vector-valued kernels and the scope of application of spatial U-quantiles, in Section 2 we introduce new statistics for two important problems: *nonparametric multiple regression coefficient estimation*, and *robust multivariate dispersion estimation*. This section may be read independently of the rest of the paper.

Our main result is *a Bahadur-Kiefer representation for the sample spatial U-quantile function*. Simultaneously generalizing results in Choudhury and Serfling (1988), Koltchinskii (1994a), Chaudhuri (1992, 1996), and Arcones (1996), this converts the technical handling of a sample spatial U-quantile to that of a (vector-valued) *U-statistic*, for which extensive theory is available. Through the Bahadur-Kiefer representation, spatial U-quantiles as estimators become linked in a natural way with general forms of spatial sign and signed-rank tests for multivariate data, and computations of relative efficiencies become likewise linked (Möttönen and Oja, 1995, Möttönen, Oja, and Tienari, 1997, Möttönen, Oja, and Serfling, 2005, and Zhou and Serfling, 2007).

We next provide background on spatial quantiles and Bahadur-Kiefer representations, formally state our main result (Theorem 1.1), and indicate general lines of application. The proof of Theorem 1.1 is given in Section 3 and utilizes new exponential probability inequalities for the supremum and modulus of continuity functionals of a U-process over a VC class. Of independent interest, these results (Theorems A.1 and A.2) are developed in the Appendix.

1.1 The spatial quantile function and its inverse

Following Chaudhuri (1996), the spatial quantile function corresponding to a given cdf F on $\mathbb{R}^d$ is defined over $\boldsymbol{u}$ in the open unit ball $\mathbb{B}^{d-1}$ as a d-vector $\boldsymbol{Q}_F(\boldsymbol{u})$ given by $\boldsymbol{\theta}$ minimizing

$$E\{\Phi(\boldsymbol{u}, \boldsymbol{X} - \boldsymbol{\theta}) - \Phi(\boldsymbol{u}, \boldsymbol{X})\}, \tag{2}$$

where $\Phi(\boldsymbol{u}, \boldsymbol{t}) = \|\boldsymbol{t}\| + \langle \boldsymbol{u}, \boldsymbol{t}\rangle$ with $\|\cdot\|$ the usual Euclidean norm and $\langle \cdot, \cdot \rangle$ the usual Euclidean inner product. Here $\boldsymbol{Q}_F(\boldsymbol{0})$ is the well-known *spatial median*. Equivalently, in terms of the *spatial sign function* (or *unit vector function*),

$$\boldsymbol{S}(\boldsymbol{x}) = \begin{cases} \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, & \boldsymbol{x} \in \mathbb{R}^d, \boldsymbol{x} \neq \boldsymbol{0}, \\ \boldsymbol{0}, & \boldsymbol{x} = \boldsymbol{0}, \end{cases}$$

the quantile $\boldsymbol{Q}_F(\boldsymbol{u})$ at $\boldsymbol{u}$ may be represented as the solution $\boldsymbol{x}$ of

$$E\{\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X})\} = \boldsymbol{u}. \tag{3}$$

Thus the quantile function $\boldsymbol{Q}_F(\cdot)$ has an *inverse*, given at each point $\boldsymbol{x} \in \mathbb{R}^d$ by the point $\boldsymbol{u}$ in $\mathbb{B}^{d-1}$ for which $\boldsymbol{x}$ has a quantile interpretation as $\boldsymbol{Q}_F(\boldsymbol{u})$, that is, by

$$\boldsymbol{Q}_F^{-1}(\boldsymbol{x}) = E\{\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X})\},$$

the *expected direction* to $\boldsymbol{x}$ from a random point $\boldsymbol{X} \sim F$. The function $\boldsymbol{Q}_F^{-1}(\boldsymbol{x})$ is also known as the *spatial centered rank function* (Möttönen and Oja, 1995). Thus the spatial quantile function and spatial centered rank function are simply inverses of each other.

Sample analogues of $\boldsymbol{Q}_F(\cdot)$ and $\boldsymbol{Q}_F^{-1}(\cdot)$ are readily defined. Let F_n denote the usual empirical df placing equal mass on observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ in $\mathbb{R}^d$, put $\boldsymbol{Q}_{F_n}^{-1}(\boldsymbol{x}) = n^{-1}\sum \boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X}_i)$, and define $\boldsymbol{Q}_{F_n}(\boldsymbol{u})$ as the solution $\boldsymbol{x}$ of $\boldsymbol{Q}_{F_n}^{-1}(\boldsymbol{x}) = \boldsymbol{u}$.

The spatial quantile function and related functions have strong appeal in applications. See Möttönen and Oja (1995) for a spatial sign test, Vardi and Zhang (2000) for a spatial depth function equivalent to an outlyingness function based on spatial quantiles, Chakraborty

(2001) for affine equivariant versions of the spatial quantiles, Serfling (2004) for properties of the spatial quantile function and related nonparametric multivariate descriptive measures for location, spread, skewness, and kurtosis, and Cheng and De Gooijer (2007) for kernel-based conditional spatial quantiles and related Bahadur-type representations.

1.2 Bahadur-Kiefer representation for spatial U-quantiles

We first restate the Bahadur representation for univariate sample quantiles in terminology convenient for the multivariate case. Then we discuss the Bahadur-Kiefer representation for multivariate spatial quantiles given by Koltchinskii (1994a) and formally state our extension to *spatial U-quantiles.*

1.2.1 Classical univariate Bahadur representation

Let F be a cdf on $\mathbb{R}$. Under second order differentiability of F in a neighborhood of $F^{-1}(p)$, with $F'(F^{-1}(p)) > 0$, Bahadur (1966) approximates the sample quantile function by a classical average: for $p \in (0,1)$,

$$F_n^{-1}(p) - F^{-1}(p) = -[F'(F^{-1}(p)]^{-1}\,[F_n(F^{-1}(p)) - p] + \tilde{R}_n(p), \tag{4}$$

where $\tilde{R}_n(p) = O(n^{-3/4}(\log\log n)^{3/4})$ almost surely. See also Kiefer (1967) and Ghosh (1971). In the orientation of Section 1.1 to quantile functions on the unit ball, under transformation via $u = 2p-1$ to $Q_F(u) = F^{-1}(\frac{1+u}{2})$, $-1 < u < 1$, an equivalent center-outward version of (4) is readily obtained: for $u \in (-1,1)$,

$$Q_{F_n}(u) - Q_F(u) = -[(Q_F^{-1})'(Q_F(u))]^{-1}\,\frac{1}{n}\sum[S(Q_F(u) - X_i) - u] + R_n(u), \tag{5}$$

with $R_n(u) = \tilde{R}_n(\frac{1+u}{2})$. Since $Q_{F_n}^{-1}(x) = n^{-1}\sum S(x - X_i)$, we may also express (5) as

$$Q_{F_n}(u) - Q_F(u) = -[(Q_F^{-1})'(Q_F(u))]^{-1}\,[Q_{F_n}^{-1}(Q_F(u)) - u] + R_n(u), \tag{6}$$

the direct analogue of (4) in center-outward form.

1.2.2 Bahadur-Kiefer representations for the sample spatial quantile function

With F and F_n now on $\mathbb{R}^d$, the d-variate *spatial* analogue of (5) is established in Chaudhuri (1996). With $\boldsymbol{Q}_F(\boldsymbol{u})$ for $Q_F(u)$, etc., and the matrix derivative $\partial \boldsymbol{Q}_F^{-1}(\boldsymbol{x})/\partial \boldsymbol{x}$ for $dQ_F^{-1}(x)/dx$,

$$\boldsymbol{Q}_{F_n}(\boldsymbol{u}) - \boldsymbol{Q}_F(\boldsymbol{u}) = -[\boldsymbol{D}_1(\boldsymbol{Q}_F(\boldsymbol{u}))]^{-1}\,\frac{1}{n}\sum[\boldsymbol{S}(\boldsymbol{Q}_F(\boldsymbol{u}) - \boldsymbol{X}_i) - \boldsymbol{u}] + \boldsymbol{R}_n(\boldsymbol{u}), \tag{7}$$

where

$$\begin{aligned}
\boldsymbol{D}_1(\boldsymbol{x}) &= E\left\{\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{S}(\boldsymbol{x}-\boldsymbol{X})\right\} \\
&= E\left\{\frac{1}{\|\boldsymbol{x}-\boldsymbol{X}\|}\left[\boldsymbol{I}_d - \frac{1}{\|\boldsymbol{x}-\boldsymbol{X}\|^2}(\boldsymbol{x}-\boldsymbol{X})(\boldsymbol{x}-\boldsymbol{X})'\right]\right\}
\end{aligned}$$

and almost surely $\boldsymbol{R}_n(\boldsymbol{u})$ is $o(n^{-\beta})$ for any fixed $\beta \in (0,1)$ in the case $d=2$ and $O(n^{-1}\log n)$ in the case $d \geq 3$. Note that $\boldsymbol{D}_1(\boldsymbol{x})$ is $d \times d$ symmetric and, unless F is supported by a straight line in $\mathbb{R}^d$, positive definite. For the function $\|\boldsymbol{x}\|$, the gradient or first order derivative is given by the sign function $\boldsymbol{S}(\boldsymbol{x})$, and the $d \times d$ Hessian or second order derivative is given by

$$\boldsymbol{D}_2(\boldsymbol{x}) = \left\{\frac{1}{\|\boldsymbol{x}\|}\left[\boldsymbol{I}_d - \frac{1}{\|\boldsymbol{x}\|^2}\boldsymbol{x}\boldsymbol{x}'\right]\right\}.$$

Thus $\boldsymbol{D}_1(\boldsymbol{x}) = E\{\boldsymbol{D}_2(\boldsymbol{x}-\boldsymbol{X})\}$. (For details, see Chaudhuri, 1992, Lemma 5.3.)

Koltchinskii (1994a) develops a *Bahadur-Kiefer version* of the above representation. Let $\boldsymbol{Q}_F^{-1}(\boldsymbol{x})$ be continuously differentiable in an open set $\mathbb{V}$ in $\mathbb{R}^d$ with $\boldsymbol{D}_1(\boldsymbol{x})$ locally Lipschitz in $\mathbb{V}$: for any compact $C \subset \mathbb{V}$, there exists a constant $\alpha_C > 0$ such that

$$\|\boldsymbol{D}_1(\boldsymbol{x}) - \boldsymbol{D}_1(\boldsymbol{y})\|_* \leq \alpha_C \|\boldsymbol{x}-\boldsymbol{y}\|, \text{ for } \boldsymbol{x}, \boldsymbol{y} \in C,$$

where $\|\cdot\|_*$ is the operator norm (for matrix A, $\|A\|_* = \sqrt{\text{largest eigenvalue of } A'A}$). Put $\mathbb{W} = \boldsymbol{Q}_F^{-1}(\mathbb{V}) \subset \mathbb{B}^{d-1}$ and for any compact $K \subset \mathbb{W}$ define $\Delta_n(K) = \sup_{\boldsymbol{u} \in K} \|\boldsymbol{R}_n(\boldsymbol{u})\|$. Then, with Pr^* denoting outer probability, for any compact $K \subset \mathbb{W}$, there exist constants $\beta_1, \beta_2, \beta_3 > 0$ such that

$$Pr^*\{\Delta_n(K) \geq \delta_n\, y\} \leq \beta_1 \exp(-y/\beta_2), \quad y > \beta_3 \log n, \tag{8}$$

where

$$\delta_n := \begin{cases} n^{-1} & \text{for } d \geq 3, \\ n^{-1}(\log n)^{1/2} & \text{for } d = 2, \\ n^{-3/4} & \text{for } d = 1. \end{cases} \tag{9}$$

Under twice continuous differentiability of $\boldsymbol{Q}_F^{-1}$ at $\boldsymbol{x}_0 \in \mathbb{R}^d$, Koltchinskii (1994b) proves δ_n to be the correct normalization for weak convergence of $\Delta(\{\boldsymbol{Q}_F^{-1}(\boldsymbol{x}_0)\})$ to a limit distribution on $\mathbb{R}^d$. By the Borel-Cantelli lemma, (8) yields that almost surely

$$\Delta_n(K) = \begin{cases} O(n^{-1}\log n) & \text{for } d \geq 3, \\ O(n^{-1}(\log n)^{3/2}) & \text{for } d = 2, \\ O(n^{-3/4}\log n) & \text{for } d = 1. \end{cases} \tag{10}$$

This yields *uniformly over compacts* the pointwise rate of Chaudhuri (1996) exactly for $d \geq 3$ and slightly improved for $d = 2$. For $d = 1$, sharper rates are available, of course, as reviewed above.

1.2.3 Bahadur-Kiefer representation for sample spatial U-quantiles

To extend to *U-quantiles*, consider now i.i.d. observations $\{X_1, \ldots, X_n\}$ from a probability distribution P on *any* measurable space $(\mathbb{X}, \mathcal{A})$ and a *vector-valued* kernel $\boldsymbol{h}(x_1, \ldots, x_m)$ mapping $\mathbb{X}^m$ into $\mathbb{R}^d$. (The preceding case corresponds to $\mathbb{X} = \mathbb{R}^d$, $m = 1$, and $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{x}$.) Let $\boldsymbol{h}(X_1, \ldots, X_m)$ have cdf H on $\mathbb{R}^d$ satisfying

(i) H has a density bounded on bounded subsets of $\mathbb{R}^d$, and

(ii) If $d \geq 2$, H is not concentrated on a line.

Associated with H we define a natural *empirical cdf* H_n by placing equal probability mass on the $n_{(m)} = n(n-1)\cdots(n-m+1)$ kernel evaluations $\boldsymbol{h}(X_{i_1}, \ldots, X_{i_m})$ taken over all m-tuples $(i_1, \ldots, i_m)$ of distinct indices chosen from $\{1, \ldots, n\}$. (When $\boldsymbol{h}$ is symmetric under permutation of its arguments, it suffices to define H_n by placing equal probability mass simply on the $\binom{n}{m}$ kernel evaluations $\boldsymbol{h}(X_{i_1}, \ldots, X_{i_m})$ taken over all m-sets $\{i_1, \ldots, i_m\}$ of distinct indices chosen from $\{1, \ldots, n\}$. The kernels in the examples of Section 2 are of this type.)

By substituting a random kernel evaluation for $\boldsymbol{X}$ in (3), i.e., in terms of $\boldsymbol{Y} \sim H$ and the sign function $\boldsymbol{S}(\cdot)$ on $\mathbb{R}^d$, the *spatial U-quantile function* corresponding to H is defined as the solution $\boldsymbol{y} = \boldsymbol{Q}_H(\boldsymbol{u})$ of the equation

$$E\{\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{Y})\} = \boldsymbol{u}. \tag{11}$$

As discussed in Chaudhuri (1996), for any cdf H the solution $\boldsymbol{Q}_H(\boldsymbol{u})$ to (11) always exists for any $\boldsymbol{u}$ if condition (i) holds and is unique under the further condition (ii). The corresponding inverse function is thus $\boldsymbol{Q}_H^{-1}(\boldsymbol{y}) = E\{\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{Y})\}$, with sample analogue the vector-valued U-statistic

$$\boldsymbol{Q}_{H_n}^{-1}(\boldsymbol{y}) = n_{(m)}^{-1} \sum \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{h}(X_{i_1}, \ldots, X_{i_m}))$$

based on the empirical cdf H_n defined above. Accordingly, the sample analogue of $\boldsymbol{Q}_H(\boldsymbol{u})$ is given by the solution $\boldsymbol{y} = \boldsymbol{Q}_{H_n}(\boldsymbol{u})$ of the equation $\boldsymbol{Q}_{H_n}^{-1}(\boldsymbol{y}) = \boldsymbol{u}$. With now

$$\boldsymbol{D}_1(\boldsymbol{y}) = E\{\boldsymbol{D}_2(\boldsymbol{y} - \boldsymbol{h}(X_1, \ldots, X_m))\} = E\{\boldsymbol{D}_2(\boldsymbol{y} - \boldsymbol{Y})\},$$

for $\boldsymbol{Y} \sim H$, by assumption (i) for H and again appealing to Chaudhuri (1992, Lemma 5.3), we have that $\boldsymbol{D}_1(\boldsymbol{y})$ is positive definite. Hence the corresponding extension of (7) is given by

$$\begin{aligned} &\boldsymbol{Q}_{H_n}(\boldsymbol{u}) - \boldsymbol{Q}_H(\boldsymbol{u}) \\ &\quad = \; -[\boldsymbol{D}_1(\boldsymbol{Q}_H(\boldsymbol{u}))]^{-1}\, n_{(m)}^{-1} \sum [\boldsymbol{S}(Q_H(\boldsymbol{u}) - \boldsymbol{h}(X_{i_1}, \ldots, X_{i_m})) - \boldsymbol{u}] + \boldsymbol{R}_n(\boldsymbol{u}), \end{aligned} \tag{12}$$

with $\boldsymbol{R}_n(\boldsymbol{u})$ satisfying the following result, where again $\Delta_n(K) = \sup_{\boldsymbol{u} \in K} \|\boldsymbol{R}_n(\boldsymbol{u})\|$.

Theorem 1.1 *Let the* $\mathbb{R}^d$*-valued kernel* $\boldsymbol{h}(X_1, \ldots, X_m)$ *have cdf* H *satisfying* (i) *and* (ii) *above. Assume* $\boldsymbol{Q}_H^{-1}(\boldsymbol{y})$ *continuously differentiable and* $\boldsymbol{D}_1(\boldsymbol{y})$ *locally Lipschitz for* $\boldsymbol{y}$ *in an open set* $\mathbb{V}$ *in* $\mathbb{R}^d$*. Then, for any compact* $K \subset \mathbb{W} = \boldsymbol{Q}_H^{-1}(\mathbb{V}) \subset \mathbb{B}^{d-1}$*, there exist constants* $\beta_1, \beta_2, \beta_3 > 0$ *such that* $\Delta_n(K)$ *defined via* (12) *satisfies*

$$Pr^*\{\Delta_n(K) \geq \delta_n\, y\} \leq \beta_1 \exp(-y/\beta_2), \quad y > \beta_3 \log n, \tag{13}$$

with δ_n *as in* (9)*. Again the rates in* (10) *follow.*

Note that the leading right-hand term in (12) serves as an approximation to the estimation error $\boldsymbol{Q}_{H_n}(\boldsymbol{u}) - \boldsymbol{Q}_H(\boldsymbol{u})$ and is a *vector-valued U-statistic* in structure. Through this feature, Theorem 1.1 is quite useful in several general lines of application:

- *Joint asymptotic normality of* $\boldsymbol{Q}_{H_n}(\boldsymbol{u_1}), \ldots, \boldsymbol{Q}_{H_n}(\boldsymbol{u_k})$. This follows by straightforward application of the asymptotic multivariate normality of a vector of vector-valued U-statistics.

- *Linkage between estimators and test statistics.* A test statistic for the null hypothesis that H has spatial median $\boldsymbol{\theta}_0$ (i.e., that $Q_H(\boldsymbol{0}) = \boldsymbol{\theta}_0$) is given by the U-statistic in (12) evaluated at $\boldsymbol{u} = \boldsymbol{0}$, $\sum \boldsymbol{S}(\boldsymbol{\theta}_0 - \boldsymbol{h}(X_{i_1}, \ldots, X_{i_m}))$, to which standard U-statistic theory may be applied. Further, this test statistic is linked through (12) with estimation of the spatial median of H by the sample spatial median $\boldsymbol{Q}_{H_n}(\boldsymbol{0})$. Thus performance characteristics for the two statistics are closely related. See Zhou and Serfling (2007).

- *Almost sure convergence and law of iterated logarithm for* $\boldsymbol{Q}_{H_n}(\boldsymbol{u})$. Through (12) and Theorem 1.1, the almost sure behavior for $\boldsymbol{Q}_{H_n}(\boldsymbol{u})$ follows by U-statistic results.

2 Examples of spatial U-quantiles and applications

Here we briefly illustrate spatial U-quantiles and the scope of their applications. In several standard contexts, new types of statistic are formulated. A useful tool for their investigation is provided by the Bahadur-Kiefer representation developed in this paper.

Example A *Spatial U-quantiles for multivariate generalized Hodges-Lehmann scalar spread estimation.* As discussed in Section 1, the *location* statistic

$$\mathrm{HL}_{(m)} = \text{spatial median} \left\{ \frac{\boldsymbol{X}_{i_1} + \cdots + \boldsymbol{X}_{i_m}}{m} \right\}$$

based on the multivariate version of (1) is a U-quantile-based statistic treated already in Chaudhuri (1992). With $\boldsymbol{Q}_H(\cdot)$ the spatial quantile function for this kernel, we also may consider related statistics for measuring *spread*, such as the sample analogues of the *volumes* of "inner" or "central" regions of form $\{\boldsymbol{Q}_H(\boldsymbol{u}) : \|\boldsymbol{u}\| \leq c\}$ (see Chaudhuri, 1996, Liu, Parelius, and Singh, 1999, and Zuo and Serfling, 2000b), analogous to the interquartile range in univariate inference. These will be investigated in a separate study. ■

Example B *Spatial U-quantiles for nonparametric estimation of multiple regression slope coefficients.* As background, consider the simple linear regression model $Y = \alpha + \beta X + \varepsilon$ with i.i.d. $(X_i, Y_i, \varepsilon_i)$, $1 \leq i \leq n$, and errors $\{\varepsilon_i\}$ independent of the random regressors $\{X_i\}$. By the independence assumption, the ratios $(Y_i - Y_j)/(X_i - X_j)$ are symmetrically distributed about β, for which a natural nonparametric estimator is thus given by the *median of these ratios*, which is a version with random regressors of the well-known and classical Theil-Sen estimator (Theil, 1950, and Sen, 1968). This "median of slopes" estimator is robust, having breakdown point (BP) 0.293 resulting from the BP 0.50 for the median (see Rousseeuw and Leroy, 1987, p. 67). One can trade off some robustness in order to gain higher efficiency, however, by using *trimmed slope estimates*, i.e., by taking a weighted average of remaining slopes after trimming away fractions of the lowest and highest ordered slopes (see Frees, 1991). All such estimators can be expressed in terms of the univariate U-quantiles based on

$$h((x_1, y_1), (x_2, y_2)) = (y_2 - y_1)/(x_2 - x_1),$$

and the Bahadur and Bahadur-Kiefer representations of Choudhary and Serfling (1988) and Arcones (1996), respectively, have direct application to the corresponding sample U-quantile function.

Here we generalize the Theil-Sen approach to slope estimation in the *multiple* regression model $Y = \alpha + \boldsymbol{Z}'\boldsymbol{\beta} + \varepsilon$, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_p)'$, as follows. As in the case $p = 1$, we first eliminate the parameter α by reducing the data to the $\binom{n}{2}$ differences

$$Y_i - Y_j = (\boldsymbol{Z}_i - \boldsymbol{Z}_j)'\boldsymbol{\beta} + \varepsilon_i - \varepsilon_j, \ 1 \leq i < j \leq n.$$

Let us denote the $\binom{n}{2}$ pairs (i, j) by $\mathbb{K}$. For each set K of p pairs $\{(i_1, j_1), \ldots, (i_p, j_p)\}$ from $\mathbb{K}$ with all indices $\{i_1, j_1, \ldots, i_p, j_p\}$ *distinct*, let $\boldsymbol{Y}_{(K)}$, $\boldsymbol{Z}_{(K)}$, and $\boldsymbol{\varepsilon}_{(K)}$, respectively, denote the p-vector of differences $Y_{i_m} - Y_{j_m}$, the $p \times p$ matrix of differences $\boldsymbol{Z}_{i_m} - \boldsymbol{Z}_{j_m}$, and the p-vector of differences $\varepsilon_{i_m} - \varepsilon_{j_m}$, for $m = 1, \ldots, p$. Thus

$$\boldsymbol{Y}_{(K)} = \boldsymbol{Z}_{(K)}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{(K)}, \tag{14}$$

for each such K. We now define a relevant kernel as the least squares estimate of $\boldsymbol{\beta}$ based on equation (14). That is, for $K = \{(i_1, j_1), \ldots, (i_p, j_p)\}$,

$$\boldsymbol{h}((\boldsymbol{z}_{i_1}, y_{i_1}), (\boldsymbol{z}_{j_1}, y_{j_1}), \ldots, (\boldsymbol{z}_{i_p}, y_{i_p}), (\boldsymbol{z}_{j_p}, y_{j_p})) = (\boldsymbol{z}_{(K)}'\boldsymbol{z}_{(K)})^{-1}\boldsymbol{z}_{(K)}'\boldsymbol{y}_{(K)}. \tag{15}$$

By distinctness of the indices in K, each $\varepsilon_{i_m} - \varepsilon_{j_m}$ is a difference of independent and identically distributed observations and hence is symmetric about 0, and also the components of $\boldsymbol{\varepsilon}_{(K)}$

are independent. It easily follows that the vector $\boldsymbol{\varepsilon}_{(K)}$ in the model (14) has joint distribution *centrally symmetric* about the p-dimensional origin, i.e.,

$$\boldsymbol{\varepsilon}_{(K)} \stackrel{d}{=} -\boldsymbol{\varepsilon}_{(K)},$$

where " $\stackrel{d}{=}$ " denotes "equal in distribution". This yields, for a random kernel evaluation, that

$$\begin{aligned}
&\boldsymbol{h}((\boldsymbol{Z}_{i_1}, Y_{i_1}), (\boldsymbol{Z}_{j_1}, Y_{j_1}), \ldots, (\boldsymbol{Z}_{i_p}, Y_{i_p}), (\boldsymbol{Z}_{j_p}, Y_{j_p})) - \boldsymbol{\beta} \\
&= (\boldsymbol{Z}_{(K)}{}'\boldsymbol{Z}_{(K)})^{-1}\boldsymbol{Z}_{(K)}{}'\boldsymbol{Y}_{(K)} - \boldsymbol{\beta} \\
&= (\boldsymbol{Z}_{(K)}{}'\boldsymbol{Z}_{(K)})^{-1}\boldsymbol{Z}_{(K)}{}'\boldsymbol{\varepsilon}_{(K)} \\
&\stackrel{d}{=} -(\boldsymbol{Z}_{(K)}{}'\boldsymbol{Z}_{(K)})^{-1}\boldsymbol{Z}_{(K)}{}'\boldsymbol{\varepsilon}_{(K)} \\
&= \boldsymbol{\beta} - (\boldsymbol{Z}_{(K)}{}'\boldsymbol{Z}_{(K)})^{-1}\boldsymbol{Z}_{(K)}{}'\boldsymbol{Y}_{(K)} \\
&= \boldsymbol{\beta} - \boldsymbol{h}((\boldsymbol{Z}_{i_1}, Y_{i_1}), (\boldsymbol{Z}_{j_1}, Y_{j_1}), \ldots, (\boldsymbol{Z}_{i_p}, Y_{i_p}), (\boldsymbol{Z}_{j_p}, Y_{j_p})),
\end{aligned}$$

i.e., the random kernel evaluations have cdf in $\mathbb{R}^p$ *centrally symmetric* about $\boldsymbol{\beta}$, so that a natural nonparametric and robust estimator is thus given by

$$\hat{\boldsymbol{\beta}} = \text{spatial median}\{\boldsymbol{h}((\boldsymbol{Z}_{i_1}, Y_{i_1}), (\boldsymbol{Z}_{j_1}, Y_{j_1}), \ldots, (\boldsymbol{Z}_{i_p}, Y_{i_p}), (\boldsymbol{Z}_{j_p}, Y_{j_p}))\}. \tag{16}$$

(For $p = 1$ the above-discussed Theil-Sen estimator is recovered.) As a consequence of the BP 0.50 of the *spatial* median (Kemperman, 1987), this estimator is readily found to have BP equal to $1-(1/2)^{1/(p+1)}$ (see Rousseeuw and Leroy, 1987, p. 147, for discussion). A somewhat different but related approach was introduced by Oja and Niinimaa (1984), using the *Oja median* (Oja, 1983), which although fully affine equivariant has BP 0 (see Niinimaa, Oja, and Tableman, 1990). Drawing upon the full range of spatial U-quantiles based on the kernel (15), along with the associated Bahadur-Kiefer representation, a variety of related statistics can be constructed and analyzed, for example a (spatial) *U-depth-weighted trimmed mean*, or a *Hodges-Lehmann type estimator* given by the median of pairwise averages of the sample kernel evaluations based on (15). The above estimator $\hat{\boldsymbol{\beta}}$ and these further statistics will be investigated elsewhere. ■

Example C *Spatial U-quantiles for robust dispersion estimation.* For estimation of the *covariance matrix* $\boldsymbol{\Sigma}$ of F, the classical estimator is the *sample covariance matrix* $\boldsymbol{S} = (n-1)^{-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})'$. For purposes of discussion here, we note that it may be expressed equivalently as a U-statistic,

$$\boldsymbol{S} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \boldsymbol{h}(\boldsymbol{X}_i, \boldsymbol{X}_j),$$

based on the matrix-valued kernel

$$\begin{aligned}\boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{x}_1) &= \frac{(\boldsymbol{x}_1-\boldsymbol{x}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)'}{2} \\ &= \frac{1}{2}\begin{pmatrix} (x_{11}-x_{21})^2 & (x_{11}-x_{21})(x_{12}-x_{22}) \\ (x_{11}-x_{21})(x_{12}-x_{22}) & (x_{12}-x_{22})^2 \end{pmatrix},\end{aligned}$$

where $\boldsymbol{x}_1 = (x_{11}, x_{12})'$, etc. While conveniently unbiased for $\boldsymbol{\Sigma}$ ($E\boldsymbol{h}(\boldsymbol{X}_i,\boldsymbol{X}_j) = \boldsymbol{\Sigma}$), the estimator $\boldsymbol{S}$ is *not* robust, of course, and many alternative approaches have been formulated. Here we introduce for consideration a new choice of target parameter, the *spatial median* of the cdf H_F of $\boldsymbol{h}(\boldsymbol{X}_1,\boldsymbol{X}_2)$, say $\boldsymbol{\Sigma}_{(2)}$, along with its sample analogue estimator

$$\widehat{\boldsymbol{\Sigma}_{(2)}} = \text{spatial median}\{\boldsymbol{h}(\boldsymbol{X}_i,\boldsymbol{X}_j)\}, \tag{17}$$

which clearly *is* robust. (The spatial median of the distribution of a random matrix $\boldsymbol{M}$ is defined as the usual spatial median of the distribution of vec $\boldsymbol{M}$.) A corresponding *scalar* dispersion measure is then given by $|\det \boldsymbol{\Sigma}_{(2)}|$, as an analogue of the usual *generalized variance* $|\det \boldsymbol{\Sigma}|$. In view of the BP 0.50 of the spatial median, as noted earlier, the estimator $\widehat{\boldsymbol{\Sigma}_{(2)}}$ has the very favorable BP 0.293, *independently of the dimension of the data.* In the *univariate* case, this estimator has been proposed already in Shamos (1976) and Bickel and Lehmann (1979), and detailed studies within broad settings have been carried out in Rousseeuw and Croux (1993) and Serfling (2002b).

Let us now establish that the associated *covariance functional* $C(F) = \boldsymbol{\Sigma}_{(2)}$ satisfies the condition to be a proper covariance functional in the sense of Rousseeuw and Leroy (1987) and Lopuhaä and Rousseeuw (1991), namely that

$$C(F_{\boldsymbol{AY}+\boldsymbol{b}}) = \boldsymbol{A}\,C(F_{\boldsymbol{Y}})\boldsymbol{A}' \tag{18}$$

holds for all $\boldsymbol{b} \in \mathbb{R}^d$ and for a sufficiently wide class of nonsingular $d \times d$ matrices $\boldsymbol{A}$. Note (Searle, 1982, p. 333) that (18) is equivalent to

$$\text{vec}\; C(F_{\boldsymbol{AY}+\boldsymbol{b}}) = (\boldsymbol{A}\otimes\boldsymbol{A})\,\text{vec}\; C(F_{\boldsymbol{Y}}). \tag{19}$$

Now it is straightforward to verify for the kernel $\boldsymbol{h}$ that

$$\text{vec}\; \boldsymbol{h}(\boldsymbol{A}\boldsymbol{y}_1+\boldsymbol{b}, \boldsymbol{A}\boldsymbol{y}_2+\boldsymbol{b}) = (\boldsymbol{A}\otimes\boldsymbol{A})\,\text{vec}\; \boldsymbol{h}(\boldsymbol{y}_1,\boldsymbol{y}_2). \tag{20}$$

For (19) to hold, we thus need that the spatial median of $H_{F_{\boldsymbol{AY}+\boldsymbol{b}}}$ equals $\boldsymbol{A}\otimes\boldsymbol{A}$ times that of $H_{F_{\boldsymbol{Y}}}$. Using the property that the spatial quantiles are location equivariant with respect to shift, orthogonal, and homogeneous scale transformations (see Serfling, 2004, for detailed discussion), we conclude that (19) holds if $\boldsymbol{A}\otimes\boldsymbol{A}$ is orthogonal (up to a constant of proportionality), which readily follows if the same holds for $\boldsymbol{A}$ itself. In summary, $C(F) = \boldsymbol{\Sigma}_{(2)}$ is *covariance equivariant* with respect to *shift*, *orthogonal*, and *homogeneous scale transformations.*

In particular, for F a *normal model* in $\mathbb{R}^d$, H_F is the Wishart$(\boldsymbol{\Sigma}, 1)$ distribution with mean $\boldsymbol{\Sigma}$. In the univariate case, this is the $\sigma^2\,\chi^2_1$ distribution, having median $\sigma^2_{(2)} = k_1\sigma^2$ with $k_1 = 0.45494$, yielding a robust estimator for σ^2,

$$\hat{\sigma}^2 = k_1^{-1}\,\widehat{\sigma^2_{(2)}} = k_1^{-1}\,\mathrm{median}\{(X_i - X_j)^2/2\},$$

which has Pitman asymptotic relative efficiency 0.864 with respect to the usual sample variance. This estimator is discussed in Rousseeuw and Croux (1993), where variants trading off some ARE for higher BP are developed, and in Serfling (2002b), where variants (the univariate case of Example C* below) trading off some BP for higher ARE are developed.

Further discussion of $\widehat{\boldsymbol{\Sigma}_{(2)}}$ is included in the broader setting of Example C* below. ■

Remark A *Computational burden.* Because of the large number of kernel evaluations involved, the burden for computing U-quantiles is enormous. This can be resolved, however, by taking either a random subpopulation of terms (Rousseeuw and Leroy, p. 147), or a nonrandom selection of terms as with "incomplete" U-statistics and generalized L-statistics (Blom, 1976, Brown and Kildea, 1978, and Hössler, 1996). ■

Example C* *Extension of Example C to a family of estimators.* For integer $m \geq 2$, let us now consider the matrix-valued kernel

$$\boldsymbol{h}^{(m)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \frac{1}{m-1}\sum_{i=1}^{m}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})' = \binom{m}{2}^{-1} \sum_{1 \leq i < j \leq m} \boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{x}_j),$$

with kernel $\boldsymbol{h}$ as in Example C, to which this reduces when $m = 2$. The U-statistic estimator based on $\boldsymbol{h}^{(m)}$ is unbiased for $\boldsymbol{\Sigma}$, $E\boldsymbol{h}^{(m)}(\boldsymbol{X}_1, \boldsymbol{X}_2) = E\boldsymbol{h}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{\Sigma}$, but is nonrobust, so we consider the alternative target parameter $\boldsymbol{\Sigma}_{(m)}$ defined as the *spatial median* of the cdf $H_F^{(m)}$ of $\boldsymbol{h}^{(m)}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)$, with corresponding *scalar* dispersion measure $|\det \boldsymbol{\Sigma}_{(m)}|$. By arguments similar to those for Example C, the functional $C(F) = \boldsymbol{\Sigma}_{(m)}$ is *covariance equivariant* with respect to *shift*, *orthogonal*, and *homogeneous scale transformations.* The BP of the sample analogue estimator

$$\widehat{\boldsymbol{\Sigma}_{(m)}} = \text{spatial median}\{\boldsymbol{h}(\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_m})\} \tag{21}$$

is readily seen to be $1 - (1/2)^{1/m}$, independently of d.

For F a normal model in $\mathbb{R}^d$, $H_F^{(m)}$ is the Wishart$(\boldsymbol{\Sigma}, m-1)$ distribution with mean $\boldsymbol{\Sigma}$. While for $m > d$ this distribution has a density, for $m \leq d$ it is still well-defined (see, e.g., Anderson, 1984, p. 249). In the univariate case, this is the $\sigma^2\,\chi^2_{m-1}$ distribution, having median $\sigma^2_{(m)} = k_{m-1}\sigma^2$ for some constant k_{m-1}. Thus a robust estimator for σ^2 is given by

$$\hat{\sigma}^2 = k_{m-1}^{-1}\,\widehat{\sigma^2_{(m)}} = k_{m-1}^{-1}\ \text{median}\left\{\binom{m}{2}^{-1}\sum_{1\le i<j\le m}(X_i - X_j)^2/2\right\}.$$

These estimators of σ^2 have been studied in Serfling (2002b), with the finding that for $m = 2, 3, 5, 7,$ and 9, the(BP, ARE) values are given by $(0.293, 0.864)$, $(0.206, 0.862)$, $(0.129, 0.910)$, $(0.094, 0.940)$, and $(0.074, 0.956)$, respectively.

Investigation of the dispersion measure $\mathbf{\Sigma}_{(m)}$ for higher dimensional data and various models will be carried out elsewhere. ■

Remark B *Covariance estimation.* As noted above, for a normal model F, in the *univariate case* the parameter $\widehat{\mathbf{\Sigma}}_{(m)} = \sigma^2_{(m)}$ is proportional to the usual variance σ^2. This fortuitous relation enables the estimator of $\sigma^2_{(m)}$ to be converted into a *robust* one for σ^2. It is readily checked that (in the univariate case) such a relation holds quite generally: *if F is a scale change of some fixed model F_0 with mean 0 and variance 1, then*

$$\sigma^2_{(m)} = k_{F_0}\,\sigma^2$$

for some constant k_{F_0} depending only on F_0. In fact, k_{F_0} is simply the univariate median of H_{F_0}, the F_0-distribution of $\binom{m}{2}^{-1}\sum_{1\le i<j\le m}(X_i - X_j)^2/2$.

Now we ask: *does this situation extend to the multivariate case?* That is, if F on $\mathbb{R}^d$ is the distribution of $\boldsymbol{X} = \mathbf{\Sigma}^{1/2}\boldsymbol{Y}$, where $\boldsymbol{Y} \sim F_0$ with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_d$, do we have

$$\mathbf{\Sigma}_{(m)} = k_{F_0}\mathbf{\Sigma}, \tag{22}$$

where the spatial median of $H^{(m)}_{F_0}$ is $k_{F_0}\boldsymbol{I}_d$, for a constant k_{F_0} depending only on F_0? This would yield a *robust* estimator of $\mathbf{\Sigma}$ via

$$\widehat{\mathbf{\Sigma}} = k_{F_0}^{-1}\,\widehat{\mathbf{\Sigma}_{(m)}}, \tag{23}$$

requiring, of course, knowledge of the constant k_{F_0}. For (22) to hold, however, we need (18) to hold with $C(F) = \mathbf{\Sigma}_{(m)}$ and $\boldsymbol{A} = \mathbf{\Sigma}^{1/2}$. But then, as discussed in Example C, we need $\mathbf{\Sigma}^{1/2}$ to be *orthogonal*, i.e., $\mathbf{\Sigma} = \boldsymbol{I}_d$ (within a proportionality constant). Consequently, with $\mathbf{\Sigma}_{(m)}$ defined as the *spatial* median of H_F, (22) does *not* hold for arbitrary $\mathbf{\Sigma}$.

On the other hand, for *fully affine equivariant* notions of multidimensional median, we do indeed have (18) for general $\boldsymbol{A}$ and thus (22) and (23) for arbitrary $\mathbf{\Sigma}$. Well-known choices of affine equivariant multidimensional median are the halfspace (or Tukey) median, the simplicial median, the Oja median, and the projection median (see Zuo and Serfling, 2000a, for discussion). With any such alternative choice of median, the formulations of U-quantiles in Examples A, B, C, and C* remain completely valid, although the relevant breakdown and efficiency properties depend on the particular choice. It is worthy of note that (22) provides a new approach toward *nonparametric* (and robust) estimation of the *correlation matrix* associated with $\boldsymbol{\Sigma}$. Directly using $\widehat{\boldsymbol{\Sigma}_{(m)}}$ we obtain a suitable estimator without involving k_{F_0}, which becomes eliminated. That is, no assumption on F is needed. Among other applications, this supports a nonparametric and robust approach to *principal components analysis on correlation matrices*, an established linear method of *dimension reduction*. A general study of methods using (22), (23), etc., will be carried out elsewhere. ■

3 Proof of Theorem 1.1

As in Section 1.2.3, we consider independent observations $\{X_1, \ldots, X_n\}$ from a probability space $(\mathbb{X}, \mathcal{A}, P)$, an $\mathcal{A}$-measurable *vector-valued* kernel

$$\boldsymbol{h}(x_1, \ldots, x_m)$$

mapping $\mathbb{X}^m$ into $\mathbb{R}^d$, distribution H on $\mathbb{R}^d$ for $\boldsymbol{h}(X_1, \ldots, X_m)$, and corresponding empirical measures P_n based on the sample and H_n based on the sample kernel evaluations. For the proof of Theorem 1.1, it is convenient to use notation compatible with Koltchinskii (1994a). We denote the functions $\boldsymbol{Q}_H^{-1}$ and $\boldsymbol{Q}_H$ by G_H and G_H^{-1}, respectively (likewise for sample analogues), and the matrix derivative $\boldsymbol{D}_1$ of G_H by G'_H. The inverse of a linear operator $A : \mathbb{R}^d \mapsto \mathbb{R}^d$ is denoted by $inv A$ and, in particular, the inverse of the (matrix) operator $\boldsymbol{D}_1(\boldsymbol{Q}_H) = G'_H \circ G_H^{-1}$ by $inv(G'_H \circ G_H^{-1})$. Theorem 1.1 thus asserts that for any compact $K \subset \mathbb{W} = G_H(\mathbb{V}) \subset \mathbb{B}^{d-1}$, the inequality in (13) holds for

$$\Delta_n(K) = \sup_{\boldsymbol{u} \in K} \|G_{H_n}^{-1}(\boldsymbol{u}) - G_H^{-1}(\boldsymbol{u}) + inv(G'_H \circ G_H^{-1}(\boldsymbol{u}))(G_{H_n} - G_H) \circ G_H^{-1}(\boldsymbol{u})\|. \tag{24}$$

For the convenience of compatibility with standard formulations of results involving U-statistics, we suppose the kernel $\boldsymbol{h}$ to be symmetric in its arguments. Proof without this assumption is similar but entails very cumbersome notation.

Denoting $\int f\, d\nu$ by $\nu(f)$ for signed measure ν and measurable function f, we define the empirical "spatial U-quantile inverse process"

$$\xi_n(\boldsymbol{y}) = n^{1/2}(G_{H_n}(\boldsymbol{y}) - G_H(\boldsymbol{y})) = n^{1/2}(H_n - H)(\boldsymbol{S}(\boldsymbol{y} - \cdot)),\ \boldsymbol{y} \in \mathbb{R}^d.$$

Also, for open $U \subset \mathbb{R}^d$ and function x from U into a normed linear space $(L, \|\cdot\|)$, we define

$$\omega_x(\boldsymbol{t}; \delta) = \omega_x^U(\boldsymbol{t}; \delta) := \sup\{\|x(\boldsymbol{s}_1) - x(\boldsymbol{s}_2)\| : \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{B}(\boldsymbol{t}, \delta) \cap U\},\ \boldsymbol{t} \in U,\ \delta > 0,$$

where $\mathbb{B}(\boldsymbol{t}, \delta)$ denotes the open ball with center $\boldsymbol{t} \in \mathbb{R}^d$ and radius $\delta > 0$. Likewise, for compact $K \subset U$, we put $\omega_x(K; \delta) = \omega_x^U(K; \delta) := \sup_{\boldsymbol{t} \in K} \omega_x(\boldsymbol{t}; \delta),\ \delta > 0$.

We treat $\Delta_n(K)$ via results for the supremum and modulus of continuity of ξ_n, using the following lemma, which extends to spatial U-quantiles a result for ordinary spatial quantiles (Koltchinskii, 1994a, equations (4.12) and (4.13)) and is proved similarly.

Lemma 3.1 *For compact $K \subset \mathbb{W} = G_H(\mathbb{V}) \subset \mathbb{B}^{d-1}$, there exist $\delta_0 > 0$ and constants d_1, d_2, d_3, d_4 such that the condition*

$$n^{-1/2}\|\xi_n\|_{\mathbb{V}} + \binom{n}{m}^{-1} \leq \delta_0 \tag{25}$$

implies

$$\begin{aligned}\Delta_n(K) \quad \leq \quad & d_1\, n^{-1/2}\, \omega_{\xi_n}\left(G_H^{-1}(K); d_2\, n^{-1/2}\|\xi_n\|_{\mathbb{V}} + d_3 \binom{n}{m}^{-1}\right) \\ & + d_4 n^{-1}\|\xi_n\|_{\mathbb{V}}^2 + \binom{n}{m}^{-1},\ n \geq m.\end{aligned} \tag{26}$$

Proof. We follow the method of Koltchinskii (1994a) given for the case $m = 1$ and $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{x} \in \mathbb{R}^d$, introducing modifications to treat the collection of sample evaluations $\boldsymbol{h}(X_{i_1}, \ldots, X_{i_m})$ in place of the basic sample observations X_i. Using the notation $\|g\|_C = \sup_{\boldsymbol{t}\in C} |g(\boldsymbol{t}|$, we immediately have

$$\delta(G_{H_n}, G_H) := \|G_H - G_{H_n}\|_{\mathbb{V}} = n^{-1/2}\|\xi_n\|_{\mathbb{V}}. \tag{27}$$

Since H is nonatomic, all sample evaluations $\boldsymbol{h}(X_{i_1}, \ldots, X_{i_m})$ are distinct and it readily follows that $G_{H_n} \circ G_{H_n}^{-1}(\boldsymbol{y})$ equals either $G_{H_n}(\boldsymbol{h}(X_{i_1}, \ldots, X_{i_m}))$, or $\boldsymbol{y}$ according to whether $\boldsymbol{y}$ does or does not belong to the closed ball of radius $\binom{n}{m}^{-1}$ centered at $G_{H_n}(\boldsymbol{h}(X_{i_1}, \ldots, X_{i_m}))$. This leads to $\|G_{H_n} \circ G_{H_n}^{-1}(\boldsymbol{y}) - \boldsymbol{y}\| \leq \binom{n}{m}^{-1}$, i.e., denoting the identity map from $\mathbb{V}$ to $\mathbb{V}$ by $I_{\mathbb{V}}$, we have

$$\|G_{H_n} \circ G_{H_n}^{-1} - I_{\mathbb{V}}\|_{\mathbb{V}} \leq \binom{n}{m}^{-1}. \tag{28}$$

Hence

$$\|G_H - G_{H_n}\|_{\mathbb{V}} + \|G_{H_n} \circ G_{H_n}^{-1} - I_{\mathbb{V}}\|_{\mathbb{V}} \leq n^{1/2}\|\xi_n\|_{\mathbb{V}} + \binom{n}{m}^{-1}. \tag{29}$$

Now, by the assumption that G'_H is locally Lipschitz in $\mathbb{V}$, (27) yields that for compact $K \subset \mathbb{W}$ and constant c,

$$\delta(G_{H_n}, G_H)\, \omega_{G'_H}(G_H^{-1}(K); c\,\delta(G_{H_n}, G_H)) \leq c\,\delta(G_{H_n}, G_H)^2 \leq c\, n^{-1}\|\xi_n\|_{\mathbb{V}}^2, \tag{30}$$

and also that G_H is a diffeomorphism from $\mathbb{V}$ onto $\mathbb{W} = G_H(\mathbb{V})$. Consequently, a general inequality for *functional* inverse functions given in Koltchinskii (1994a, Lemma 3.1 (ii)) may be applied, with $\psi = G_H$, $\eta = G_{H_n}$, and $\eta^{(-1)} = G_{H_n}^{-1}$, and yields via (30) the claim of the lemma. ■

The proof of Theorem 1.1 will also utilize probability inequalities for the supremum and modulus of continuity functionals of certain real-valued empirical processes of U-statistic structure associated *componentwise* with the vector-valued kernel $\boldsymbol{h} = (h^{(1)}, \ldots, h^{(d)})$, as follows. For each $i = 1, \ldots, d$, we consider the (real-valued) *empirical process* defined by

$$W_n^{(i)}(f) := n^{1/2}(H_n^{(i)} - H^{(i)})(f),\ f \in \mathcal{F}^{(i)}, \tag{31}$$

over a suitable class $\mathcal{F}^{(i)}$ of real-valued P-empirically measurable (Giné and Zinn, 1986) functions on $\mathbb{X}$, with $H^{(i)}$ the cdf of $h^{(i)}(X_1, \ldots, X_m)$ and $H_n^{(i)}$ its sample analogue. This is a *U-process* indexed by $f \circ h^{(i)}$, $f \in \mathcal{F}^{(i)}$, as treated extensively in general in Arcones and Giné (1993). In particular, expressing the spatial sign function in $\mathbb{R}^d$ as $\boldsymbol{S} = (S^{(1)}, \ldots, S^{(d)})$, we will use the classes

$$\mathcal{F}^{(i)} = \{S^{(i)}(\boldsymbol{y} - \cdot) : \boldsymbol{y} \in \mathbb{R}^d\},$$

as well as the classes

$$\widetilde{\mathcal{F}}^{(i)} = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F}^{(i)}\},$$

$i = 1, \ldots, d$. These are clearly *uniformly bounded* classes. Also, as pointed out in Koltchinskii (1994a), they are Vapnik-Červonenkis subgraph classes (van der Vaart and Wellner, 1996, and Dudley, 1999), simply, VC-classes, let us say. For later reference, we note that if $\mathcal{F}$ is a VC-class, then so is $\mathcal{F} \circ h$ for a real-valued kernel $h(x_1, \ldots, x_m)$ (see van der Vaart and Wellner, 1996, Lemma 2.6.19). The probability inequalities we need for $W_n^{(i)}(\cdot)$ will draw upon ones of general interest that we develop separately in Appendix A. These results, which will be applied in the proof of Theorem 1.1 below, will concern the metric $\rho_{H,2}$ defined on $\mathcal{F}$ by

$$\begin{aligned}\rho_{H,2}(f,g) &:= (H((f-g)^2) - (H(f-g))^2)^{1/2} \\ &= (\mathrm{Var}(f-g) \circ h(X_1, \ldots, X_m))^{1/2},\ f, g \in \mathcal{F},\end{aligned}$$

and the modulus defined for any function $Z : \mathcal{F} \mapsto \mathbb{R}$ by

$$\omega(Z; \rho_{H,2}; \delta) := \sup\{|Z(f) - Z(g)| : f, g \in \mathcal{F},\ \rho_{H,2}(f,g) \leq \delta\},\ \delta > 0.$$

Our final preliminary is to state a result of Koltchinskii (1994a) for metrics on $\mathbb{R}^d$ of the form

$$d_{P,2}(\boldsymbol{s}, \boldsymbol{t}) := \left(\int_{\mathbb{R}^d} \|\boldsymbol{S}(\boldsymbol{s} - \boldsymbol{r}) - \boldsymbol{S}(\boldsymbol{t} - \boldsymbol{r})\|^2 \, dP(\boldsymbol{r})\right)^{1/2}, \boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^d,$$

for P a distribution on $\mathbb{R}^d$.

Lemma 3.2 *If P is a distribution on $\mathbb{R}^d$ with a bounded density, then there exists a constant $c_0 > 0$ such that for any $\boldsymbol{s} \neq \boldsymbol{t} \in \mathbb{R}^d$,*

$$d_{P,2}(\boldsymbol{s}, \boldsymbol{t}) \leq c_0 \begin{cases} \|\boldsymbol{s} - \boldsymbol{t}\|^{1/2}, & \text{for } d = 1 \\ \|\boldsymbol{s} - \boldsymbol{t}\| \log^{1/2} \frac{1}{\|\boldsymbol{s}-\boldsymbol{t}\|}, & \text{for } d = 2 \\ \|\boldsymbol{s} - \boldsymbol{t}\|, & \text{for } d > 2. \end{cases}$$

Proof of Theorem 1.1. Since we follow closely the method of Koltchinskii (1994a, Theorem 2.1), with appropriate substitutions of our results for U-processes in place of results for classical empirical processes, it suffices to present this proof as a sketch. Componentwise application of Theorem A.1, inequality (A.4), to the vector-valued process ξ_n yields

$$\Pr{}^*(\|\xi_n\|_{\mathbb{V}} \geq t) \leq d_1 \exp(-t^2/d_2), \; t > 0. \tag{32}$$

From the above definitions we have, for $1 \leq i \leq d$,

$$\begin{aligned} &\rho_{H,2}(S^{(i)}(\boldsymbol{y}_1 - \cdot), S^{(i)}(\boldsymbol{y}_2 - \cdot)) \\ &= \left(\mathrm{Var}(S^{(i)}(\boldsymbol{y}_1 - h) - S^{(i)}(\boldsymbol{y}_2 - h))\right)^{1/2} \\ &\leq \left(\int_{\mathbb{R}_d} \left|S^{(i)}(\boldsymbol{y_1} - h) - S^{(i)}(\boldsymbol{y_2} - h)\right|^2 dH\right)^{1/2} \\ &= d_{H,2}(\boldsymbol{y}_1, \boldsymbol{y}_2), \; \boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{R}^d. \end{aligned}$$

Put $\tilde{\delta} = \sup\{d_{H,2}(\boldsymbol{y}_1, \boldsymbol{y}_2) : |\boldsymbol{y}_1 - \boldsymbol{y}_2| \leq \delta\}$. Then, for $1 \leq i \leq d$,

$$\begin{aligned} &\omega_{\xi_n^{(i)}}(G_H^{-1}(K); \delta) \\ &= \sup_{\boldsymbol{t} \in G_H^{-1}(K)} \{\left|W_n^{(i)}(S^{(i)}(\boldsymbol{y}_1, \cdot)) - W_n^{(i)}(S^{(i)}(\boldsymbol{y}_2, \cdot))\right| : \boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{B}(\boldsymbol{t}, \delta)\} \\ &\leq \sup\{|W_n^{(i)}(f_1) - W_n^{(i)}(f_2)| : f_1, f_2 \in \mathcal{F}^{(i)}, \rho_{H,2}(f_1, f_2) \leq \sup_{|\boldsymbol{y}_1 - \boldsymbol{y}_2| \leq \delta} d_{H,2}(\boldsymbol{y}_1, \boldsymbol{y}_2)\} \\ &= \omega(W_n^{(i)}; \rho_{H,2}; \tilde{\delta}). \end{aligned} \tag{33}$$

Based on (33), we apply Lemma 3.2 and Theorem A.2 to obtain the following key results for $\omega_{\xi_n^{(i)}}(G_H^{-1}(K); \delta)$. For $d = 1$, with

$$\delta_1 = \sup_{|\boldsymbol{y}_1 - \boldsymbol{y}_2| \leq \frac{t}{n^{1/2}}} d_{H,2}(\boldsymbol{y}_1, \boldsymbol{y}_2),$$

and $t \geq c_5 \log n$, we have

$$\begin{aligned} \Pr{}^*\left(\{\omega_{\xi_n}(G_H^{-1}(K); \frac{t}{n^{1/2}}) \geq \frac{c_1 t}{n^{1/4}}\}\right) &\leq \Pr{}^*\left(\{\omega(W_n; \rho_{H,2}; \delta_1) \geq \frac{c_1 t}{n^{1/4}}\}\right) \\ &\leq \Pr{}^*\left(\{\omega(W_n; \rho_{H,2}; \frac{c_2 t^{1/2}}{n^{1/4}}) \geq \frac{c_1 t}{n^{1/4}}\}\right) \\ &\leq c_3 e^{-t/c_4}. \end{aligned} \tag{34}$$

For $d = 2$ and $t \geq c_5 \log n$, similar steps lead to

$$\Pr{}^* \left(\{\omega_{\xi_n}(G_H^{-1}(K); \frac{t}{n^{1/2}}) \geq \frac{c_1 t (\log n)^{1/2}}{n^{1/2}}\} \right) \leq c_3 e^{-t/c_4}. \tag{35}$$

For $d > 2$ and $t \geq c_5 \log n$, we obtain

$$\Pr{}^* \left(\{\omega_{\xi_n}(G_H^{-1}(K); \frac{t}{n^{1/2}}) \geq \frac{c_1 t}{n^{1/2}}\} \right) \leq c_3 e^{-t/c_4}. \tag{36}$$

We continue the proof only in the case $d > 2$ (the other cases just using (34) or (35) instead of (36)). Let

$$A = \left\{ \Delta_n(K) \geq n^{-1} \left[(d_1^2 + d_4)t + n \binom{n}{m}^{-1} \right] \right\},$$

$$B = \left\{ \omega_{\xi_n} \left(G_H^{-1}(K); \frac{d_2 \|\xi_n\|_{\mathbb{V}}}{n^{1/2}} + d_3 \binom{n}{m}^{-1} \right) \geq d_1 t n^{-1/2} \right\},$$

$$C = \{\|\xi_n\|_{\mathbb{V}}^2 \geq t\},$$

and

$$D = \left\{ \|\xi_n\|_{\mathbb{V}} \geq n^{1/2}\delta_0 - \binom{n}{m}^{-1} n^{1/2} \right\}.$$

Because under D^c we have (25), it follows by Lemma 3.1 that $D^c \cap B^c \cap C^c$ implies A^c, yielding $P(A) \leq P(B) + P(C) + P(D)$, i.e.,

$$\begin{aligned}
&\Pr{}^* \left(\Delta_n(K) \geq n^{-1} \left[(d_1^2 + d_4)t + n \binom{n}{m}^{-1} \right] \right) \\
&\leq \Pr{}^* \left(\omega_{\xi_n} \left(G_H^{-1}(K); \frac{d_2 \|\xi_n\|_{\mathbb{V}}}{n^{1/2}} + d_3 \binom{n}{m}^{-1} \right) \geq d_1 t n^{-1/2} \} \right) \\
&\quad + \Pr{}^*(\|\xi_n\|_{\mathbb{V}}^2 \geq t) + \Pr{}^* \left(\|\xi_n\|_{\mathbb{V}} \geq n^{1/2}\delta_0 - \binom{n}{m}^{-1} n^{1/2} \right).
\end{aligned} \tag{37}$$

From (32), for

$$0 < t < n\delta_0^2 - 2n \binom{n}{m}^{-1} \delta_0 + n \binom{n}{m}^{-2},$$

we obtain

$$
\begin{aligned}
&\Pr{}^*(\|\xi_n\|_{\mathbb{V}}^2 \geq t) + \Pr{}^*\left(\|\xi_n\|_{\mathbb{V}} \geq n^{1/2}\delta_0 - \binom{n}{m}^{-1} n^{1/2}\right) \\
&\leq \; d_1 \exp(-t/d_2) + d_1 \exp\left(-\frac{n\delta_0^2 - 2n\binom{n}{m}^{-1}\delta_0 + \binom{n}{m}^{-2} n}{d_2}\right) \\
&\leq \; d_3 \exp(-t/d_2). \qquad (38)
\end{aligned}
$$

On the other hand, using

$$
\begin{aligned}
&\left\{\omega_{\xi_n}\left(G_H^{-1}(K); \frac{t^{1/2}}{n^{1/2}}\right) \geq \frac{d_1 t}{n^{1/2}}\right\}^c \cap \left\{\|\xi_n\|_{\mathbb{V}} \geq \frac{t^{1/2}}{d_2}\right\}^c \\
&\subset \; \left\{\omega_{\xi_n}\left(G_H^{-1}(K); \frac{d_2\|\xi_n\|_{\mathbb{V}}}{n^{1/2}} + d_3\binom{n}{m}^{-1}\right) \geq d_1 t n^{-1/2}\right\}^c
\end{aligned}
$$

with (36) and (32), for $t \geq c_5 \log n$ we have

$$
\begin{aligned}
&\Pr{}^*\left(\omega_{\xi_n}\left(G_H^{-1}(K); \frac{d_2\|\xi_n\|_{\mathbb{V}}}{n^{1/2}} + d_3\binom{n}{m}^{-1}\right) \geq d_1 t n^{-1/2}\right) \\
&\leq \; \Pr{}^*\left(\omega_{\xi_n}\left(G_H^{-1}(K); \frac{t^{1/2}}{n^{1/2}}\right) \geq \frac{d_1 t}{n^{1/2}}\right) + \Pr{}^*\left(\|\xi_n\|_{\mathbb{V}} \geq \frac{t^{1/2}}{d_2}\right) \\
&\leq \; c_3 \exp(-t/c_4). \qquad (39)
\end{aligned}
$$

Based on (37), (38) and (39), we obtain

$$
\Pr{}^*(\{\Delta_n(K) \geq n^{-1}t\}) \leq c_3 \exp(-t/c_4), \qquad (40)
$$

where $c_5 \log n \leq t \leq n\delta_0^2 - 2n\binom{n}{m}^{-1}\delta_0 + \binom{n}{m}^{-2} n$. In order to prove the inequality for large t, we note that

$$
\begin{aligned}
\Delta_n(K) \;&=\; \|G_{H_n}^{-1}(\boldsymbol{u}) - G_H^{-1}(\boldsymbol{u}) + \mathit{inv}(G'_H \circ G_H^{-1}(\boldsymbol{u}))(G_{H_n} - G_H) \circ G_H^{-1}(\boldsymbol{u})\|_K \\
&\leq\; \|G_{H_n}^{-1}(\boldsymbol{u}) - G_H^{-1}(\boldsymbol{u})\|_K \\
&\quad + \|\mathit{inv}(G'_H \circ G_H^{-1}(\boldsymbol{u}))(G_{H_n} - G_H) \circ G_H^{-1}(\boldsymbol{u})\|_K \\
&\leq\; \|G_{H_n}^{-1}(\boldsymbol{u}) - G_H^{-1}(\boldsymbol{u})\|_K + c\|G_{H_n} - G_H\|_{G_H^{-1}(K)}. \qquad (41)
\end{aligned}
$$

From Lemma 3.1, noting that G_H^{-1} is also locally Lipschitz in $\mathbb{V}$, we have

$$
\begin{aligned}
\|G_{H_n}^{-1}(\boldsymbol{u}) - G_H^{-1}(\boldsymbol{u})\|_K \;&\leq\; \omega_{G_H^{-1}}(K; \gamma(G_H, G_{H_n}, G_{H_n}^{-1})) \\
&=\; \sup_{t \in K}\{|G_H^{-1}(\boldsymbol{y_1}) - G_H^{-1}(\boldsymbol{y_2})| : \boldsymbol{y_1}, \boldsymbol{y_2} \in \mathbb{B}(t, \gamma)\} \\
&\leq\; \sup_{t \in K}\{c|\boldsymbol{y_1} - \boldsymbol{y_2}| : |\boldsymbol{y_1} - \boldsymbol{y_2}| \leq \gamma(G_H, G_{H_n}, G_{H_n}^{-1})\} \\
&\leq\; c\gamma(G_H, G_{H_n}, G_{H_n}^{-1}). \qquad (42)
\end{aligned}
$$

From (27), (28), (41), and (42), we obtain

$$\begin{aligned}
\Delta_n(K) &\leq c_1\gamma(G_H, G_{H_n}, G_{H_n}^{-1}) + c_2\|G_{H_n} - G_H\|_{\mathbb{V}} \\
&= c_1\gamma(G_H, G_{H_n}, G_{H_n}^{-1}) + c_2\delta(G_{H_n}, G_H) \\
&= c_1[\delta(G_{H_n}, G_H) + \theta(G_{H_n}, G_{H_n}^{-1})] + c_2\delta(G_{H_n}, G_H) \\
&\leq c_3\delta(G_{H_n}, G_H) + c_4\binom{n}{m}^{-1} \\
&\leq c_5\left(n^{-1/2}\|\xi_n\|_{\mathbb{V}} + \binom{n}{m}^{-1}\right). \qquad (43)
\end{aligned}$$

Now for any $\mathcal{F}_i$, $1 \leq i \leq d$ and $t \geq cn$, we can apply Theorem A.2 to obtain

$$\Pr{}^*\left(\|\xi_n\|_{\mathbb{V}} \geq tn^{-1/2}\right) \leq c_3\exp(-\frac{t^2}{nc_4}) \leq c_3\exp(-\frac{cnt}{nc_4}) \leq c_3\exp(-t/c_5). \qquad (44)$$

Combining (43) and (44), we obtain, for any $t \geq c_1 n$,

$$\begin{aligned}
&\Pr{}^*\left(\|\Delta_n(K)\| \geq tn^{-1}\right) \\
&\leq \Pr{}^*\left(c_5(n^{-1/2}\|\xi_n\|_{\mathbb{V}} + \binom{n}{m}^{-1}) \geq tn^{-1}\right) \\
&\leq \Pr{}^*\left(c_5(\|\xi_n\|_{\mathbb{V}} + n^{1/2}\binom{n}{m}^{-1}) \geq tn^{-1/2}\right) \\
&\leq \Pr{}^*(\|\xi_n\|_{\mathbb{V}} \geq c_2 tn^{-1/2}) \\
&\leq c_3\exp(-t/c_4). \qquad (45)
\end{aligned}$$

It follows from (40) and (45), when $d > 2$,

$$\Pr{}^*(\Delta_n(K) \geq n^{-1}t) \leq \beta_1\exp(-t/\beta_2),\ \ t > \beta_3\log n.$$

Similar arguments lead to the desired inequalities for $d = 1$ and $d = 2$. ■

Appendix A: Probability inequalities for some U-process functionals

We give in Theorem A.1 exponential probability inequalities for the supremum functional of a U-process over a VC class, for example the real-valued U-processes as defined in (31). This yields a similar result (Theorem A.2) for the modulus of continuity functional. We obtain Theorem A.1 from an exponential inequality (Lemma A.4) for the moment generating function of the supremum functional for a U-process. In turn, Lemma A.4 follows from a similar result (Lemma A.3) developed for the classical empirical process over a VC class. All of these results are of independent interest.

For i.i.d. observations $X_1, \ldots, X_n$ having measure P and empirical measure P_n, we define

$$Z_n(f) := n^{1/2}(P_n - P)(f),\ f \in \mathcal{F},$$

and put

$$\sigma_0^2 = \sigma_0^2(\mathcal{F}) := \sup_{f \in \mathcal{F}} (P(f^2) - P(f)^2) := \sup_{f \in \mathcal{F}} \mathrm{Var}(f(X_1)).$$

Given a real-valued kernel $q(x_1, \ldots, x_m)$ symmetric in its arguments, let us denote by $U_n(q)$ the U-statistic based on $X_1, \ldots, X_n$ and define associated functions

$$\pi_{k,m} q(x_1, \ldots, x_k) = (\delta_{x_1} - P) \cdots (\delta_{x_k} - P) P^{m-k}(q),\ k = 0, \ldots, m,$$

where $\alpha_1 \ldots \alpha_m(q) = \int q(y_1, \ldots, y_m)\, d\alpha_1(y_1) \ldots d\alpha_m(y_m)$ and δ_x denotes the distribution with mass 1 at x. The corresponding Hoeffding decomposition (Serfling, 1980, and Arcones and Giné, 1993) of $U_n(q)$ is thus $\sum_{k=0}^m \binom{m}{k} U_n(\pi_{k,m} q)$. We will make use of the following lemma, extracted from Arcones (1995, Proposition 4), giving a uniform exponential probability inequality for the nonlinear components of the Hoeffding decompositions for a uniformly bounded VC class of kernels.

Lemma A.1 *Let $\mathcal{Q}$ be a uniformly bounded VC-class of (symmetric) kernels q. Then there exists a constant $c > 0$ such that*

$$Pr\left(n^{1/2} \left\| \sum_{k=2}^m \binom{m}{k} U_n(\pi_{k,m} q) \right\|_{\mathcal{Q}} \geq t \right) \leq 2 \exp(-t n^{1/2}/c),\ t > 0. \tag{A.1}$$

As a further preliminary, we introduce, for a family $\mathcal{F}$ of real functions and a real-valued (symmetric) kernel $h(x_1, \ldots, x_m)$ with associated cdf H and empirical cdf H_n, the corresponding U-process indexed by $f \circ h$, $f \in \mathcal{F}$,

$$W_n(f) := n^{1/2}(H_n - H)(f),\ f \in \mathcal{F},$$

and put

$$\sigma^2 = \sigma^2(\mathcal{F}) := \sup_{f \in \mathcal{F}} \mathrm{Var}(\pi_{1,m} f \circ h(X_1)).$$

We now state and prove our main results, Theorems A.1–A.2 and Lemmas A.3–A.4. Our starting point is the following lemma for $Z_n(\cdot)$ extracted from Alexander (1984) via Koltchinskii (1994a).

Lemma A.2 *Let $\mathcal{F}$ be a VC-class. Then there exist constants $c_1, c_2 > 0$ such that*

$$Pr^*(\sup_{\mathcal{F}} |Z_n(f)| \geq t) \leq c_1 \exp(-t^2/c_2), \ t > 0, \tag{A.2}$$

and constants $c_3, \ldots, c_7 > 0$ such that, for $t > (c_3 n^{-1/2} \log n) \vee (c_4 \sigma_0 \log^{1/2}(1/\sigma_0)$,

$$Pr^*(\sup_{\mathcal{F}} |Z_n(f)| \geq t) \leq c_5 \exp(-t^2/c_6\sigma_0^2) \bigvee \exp(-tn^{1/2}/c_7). \tag{A.3}$$

Extension of this result to $W_n(\cdot)$, with σ_0 replaced by σ, is given by

Theorem A.1 *Let $\mathcal{F}$ be a uniformly bounded VC-class. Then there exist constants $d_1, d_2 > 0$ such that*

$$Pr^*(\sup_{\mathcal{F}} |W_n(f)| \geq t) \leq d_1 \exp(-t^2/d_2), \ t > 0, \tag{A.4}$$

and constants $d_3, \ldots, d_8 > 0$ such that, for $t > (d_3 n^{-1/2} \log n) \vee (d_4 \sigma \log^{1/2}(1/\sigma)$,

$$Pr^*(\sup_{\mathcal{F}} |W_n(f)| \geq t) \leq d_5 \exp(-t^2/d_6\sigma^2) \bigvee d_7 \exp(-tn^{1/2}/d_8). \tag{A.5}$$

As will be seen next, Theorem A.1 yields a parallel result for the modulus of continuity of $W_n(\cdot)$ corresponding to the metric $\rho_{H,2}$ on $\mathcal{F}$ and the modulus $\omega(Z; \rho_{H,2}; \delta)$ defined in Section 3. We will prove

Theorem A.2 *Let $\{f - g : f, g \in \mathcal{F}\}$ be a VC-class based on a uniformly bounded $\mathcal{F}$. Then, for any $\alpha \in (0, 1]$, there exist constants $c_1, \ldots, c_6$ such that*

$$Pr^*(\omega(W_n; \rho_{H,2}; t^{1/2}/n^{\alpha/2}) \geq t/n^{\alpha/2}) \leq c_1 \exp(-t/c_2), \ t \geq c_3 \log n, \tag{A.6}$$

and

$$\begin{aligned} &Pr^*(\omega(W_n; \rho_{H,2}; t^{1/2}/n^{1/2} \log^{1/2} n) \geq t/n^{1/2} \log^{1/2} n) \\ &\leq c_4 \exp(-t/c_5), \ t \geq c_6 \log n. \end{aligned} \tag{A.7}$$

For $m = 1$ and $h(x) = x$, this reduces to Lemma 4.2 of Koltchinskii (1994a).

As mentioned earlier, Theorem A.1 will be derived from moment generating function inequalities, the first of which is the following companion to Lemma A.2.

Lemma A.3 *Let $\mathcal{F}$ be a uniformly bounded VC-class. Then there exist constants $C_0, C_1 > 0$ such that*

$$E\{\exp s\|P_n - P\|_{\mathcal{F}}\} \leq (1 + C_0 s n^{-1/2}) \exp(C_1 s^2/4n), \ s > 0, \ n \geq 1. \tag{A.8}$$

This result extends to U-quantiles:

Lemma A.4 *Let $\mathcal{F}$ be a uniformly bounded VC-class. Then there exist constants $C_3, C_4 > 0$ such that*

$$E\{\exp s\|H_n - H\|_{\mathcal{F}}\} \le (1 + C_3 s[n/m]^{-1/2}) \exp(C_4 s^2/4[n/m]),\ \ s > 0,\ \ n \ge m. \qquad \text{(A.9)}$$

With $\mathcal{F}$ the class of indicator functions $\{f(x) = \mathbf{1}(x \le \cdot),\ x \in \mathbb{R}\}$, Lemmas A.3 and A.4 reduce to Lemmas 2.2 and 2.3 of Helmers, Janssen and Serfling (1988), whose technique we use for the present versions.

PROOF OF LEMMAS A.3 AND A.4. Without loss of generality, let the uniform bound on $\mathcal{F}$ equal 1/2. By a well-known inequality and an application of Lemma A.2, equation (A.2), we have, for $s > 0$,

$$\begin{aligned} E\{\exp s\|P_n - P\|_{\mathcal{F}}\} &= \int_0^\infty P(\exp s\|P_n - P\|_{\mathcal{F}} > t)\,dt \\ &= 1 + \int_1^{e^s} P(\|P_n - P\|_{\mathcal{F}} > (\log t)/s)\,dt \\ &\le 1 + \int_1^{e^s} \exp(-n(\log t)^2/cs^2)\,dt, \qquad \text{(A.10)} \end{aligned}$$

where we have used $0 \le \|P_n - P\|_{\mathcal{F}} \le 1$. It is straightforward to express the integral in (A.10) as, denoting by $N(\mu, \sigma^2)$ a normal random variable with mean μ and variance σ^2,

$$(c\pi)^{1/2} s n^{-1/2}\, P(0 \le N(cs^2/2n, cs^2/2n) < s)\ \exp(cs^2/4n),$$

which is dominated by $csn^{-1/2}\exp(cs^2/4n)$. Then (A.8) follows immediately, with $C_1 = (c\pi)^{1/2}$ and $C_2 = c$.

To obtain (A.9), we use the well-known representation (Hoeffding, 1963, and Serfling, 1980, p. 180)

$$H_n - H = \frac{1}{n!}\sum_{i=1}^{n!} (\alpha_n^{(i)} - H),$$

where each $\alpha_n^{(i)}$ is a classical empirical distribution based on $[n/m]$ terms $h(X_{i_1}, \ldots, X_{i_m})$, $h(X_{i_{m+1}}, \ldots, X_{i_{2m}})$, ... corresponding to a particular permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$. Via convexity of the exponential function and Jensen's inequality, (A.8) leads to (A.9). ■

PROOF OF THEOREM A.1. We apply Lemma A.4 in connection with the elementary relation

$$P(\|H_n - H\|_{\mathcal{F}} \ge t) \le e^{-st} E\{e^{s\|H_n - H\|_{\mathcal{F}}}\}$$

and take $s = 2[n/m]t/C_4$. Then

$$\begin{aligned} P(\|W_n\|_{\mathcal{F}} \ge t) &= P(\|H_n - H\|_{\mathcal{F}} \ge tn^{-1/2}) \\ &\le e^{-st} E\{e^{s\|H_n - H\|_{\mathcal{F}}}\} \\ &\le (1 + C_3([n/m]/n)^{1/2} t) \exp(-[n/m]t^2/C_4 n) \\ &\le c_1 \exp(-t^2/c_2), \end{aligned}$$

for suitable c_1 and c_2, proving (A.4). Now apply the Hoeffding decomposition $U_n(f \circ h) = \sum_{k=0}^{m} \binom{m}{k} U_n(\pi_{k,m} f \circ h)$ to write

$$\begin{aligned} W_n(f) &= n^{1/2}(H_n - H)(f) = n^{1/2}(U_n(f \circ h) - P^m(f \circ h)) \\ &= n^{1/2}\left(m\, U_n(\pi_{1,m} f \circ h) + \sum_{k=2}^{m} \binom{m}{k} U_n(\pi_{k,m} f \circ h) \right) \\ &= n^{1/2} m\, n^{-1} \sum_{i=1}^{n} \pi_{1,m} f \circ h(X_i) + n^{1/2} \sum_{k=2}^{m} \binom{m}{k} U_n(\pi_{k,m} f \circ h)). \end{aligned}$$

Thus, by Lemma A.1 and Lemma A.2, equation (A.3), and the fact that $E\{\pi_{k,m} f \circ h(X_1, \ldots, X_k)\} = P(\pi_{k,m} f \circ h) = 0$ for $k \geq 1$, we obtain

$$\begin{aligned} P(\sup_{\mathcal{F}} \|W_n(f)\| \geq t) &\leq P\left(\left\| n^{-1/2} \sum_{i=1}^{n} \pi_{1,m} f \circ h(X_i) \right\|_{\mathcal{F}} \geq t/2m \right) \\ &\quad + 2\exp(-tn^{1/2}/c) \\ &\leq P(\|Z_n(\pi_{1,m} f \circ h)\|_{\mathcal{F}} \geq t/2m) + 2\exp(-tn^{1/2}/c) \\ &\leq \max\{c_5 \exp(-t^2/c_6\sigma^2),\ \ \exp(-tn^{1/2}/c_7)\} \\ &\quad + 2\exp(-tn^{1/2}/c) \\ &\leq \max\{d_5 \exp(-t^2/d_6\sigma^2),\ d_7 \exp(-tn^{1/2}/d_8)\} \end{aligned}$$

for suitable choices of constants, proving (A.5). ■

Proof of Theorem A.2. Without loss of generality we assume $Ef = 0$. Then

$$\begin{aligned} \mathrm{Var}(\pi_{1,m} f(X)) &= E([\pi_{1,m} f(X)]^2) \\ &= E\{[Ef(X, X_1, \ldots, X_{m-1})|X]^2\} \\ &\leq E\{Ef^2(X, X_1, \ldots, X_{m-1})|X\} \\ &= Ef^2(X_1, \ldots, X_m) = \mathrm{Var}(f(X_1, \ldots, X_m)), \end{aligned}$$

whence

$$\mathrm{Var}(\pi_{1,m}(f-g) \circ h) \leq \mathrm{Var}[(f-g) \circ h].$$

Thus

$$\begin{aligned} \sigma &= \left(\sup_{f,g \in \mathcal{F}} \mathrm{Var}[\pi_{1,m}(f-g) \circ h] \right)^{1/2} \\ &\leq \left(\sup_{f,g \in \mathcal{F}} \mathrm{Var}[(f-g) \circ h] \right)^{1/2} = \sup_{f,g \in \mathcal{F}} \rho_{H,2}(f,g). \qquad \text{(A.11)} \end{aligned}$$

Let us now apply Theorem A.1 to the class

$$\mathcal{C} = \{f - g : f, g \in \mathcal{F} \text{ and } \rho_{H,2}(f,g) \leq t^{1/2} n^{-\alpha/2}\}.$$

From (A.11) we have

$$\sigma \leq \sup_{f,g\in\mathcal{F}} \rho_{H,2}(f,g) \leq t^{1/2} n^{-\alpha/2}.$$

Since there exists a constant $c > 0$ such that for any $t > c \log n$

$$tn^{-\alpha/2} > (cn^{-1/2} \log n) \bigvee (c\sigma(\log 1/\sigma)^{1/2}),$$

Theorem A.1 thus yields, for suitable constants,

$$\begin{aligned}
&P\left(\omega\left(W_n; \rho_{H,2}; \frac{t^{1/2}}{n^{\alpha/2}}\right) \geq tn^{-\alpha/2}\right) \\
&= \; P\left(\sup_{f\in\mathcal{C}} |W_n(f)| \geq tn^{-\alpha/2}\right) \\
&\leq \; d_5 \exp(-t^2/(d_6\sigma^2 n^\alpha)) \bigvee d_7 \exp(-tn^{(1-\alpha)/2}/d_8) \\
&\leq \; c_1 \exp(-t/c_2).
\end{aligned}$$

Hence (A.6) follows and similar arguments lead to (A.7). ■

Acknowledgment

The authors thank Evarist Giné for providing relevant background sources on Lemma A.1 and its role in proving Theorem A.1, and Xin Dang for pointing out a needed clarification. Helpful remarks by two reviewers and an Associate Editor are greatly appreciated. Also, support by NSF Grants DMS-0103698 and CCF-0430366 is gratefully acknowledged.

References

Alexander, K., 1984. Probability inequalities for general empirical processes and a law of the iterated logarithm. Ann. Probab. 12, 1041–1076.

Anderson, T. W., 1984. An Introduction to Multivariate Statistical Analysis, Second Edition. Wiley, New York.

Arcones, M., 1995. A Bernstein-type inequality for U-statistics and U-processes. Statist. Probab. Lett. 22, 239–247.

Arcones, M., 1996. The Bahadur-Kiefer representation for U-quantiles. Ann. Statist. 24, 1400–1422.

Arcones, M. A., Giné, E., 1993. Limit theorems for U-processes. Ann. Probab. 21, 1494–1542.

Arcones, M. A., Mason, D. M., 1997. A general approach to Bahadur-Kiefer representations for M-estimators. Math. Meth. Statist. 6, 267–292.

Bahadur, R. R., 1966. A note on quantiles in large samples. Ann. Math. Statist. 37, 577–580.

Bickel, P. J., Lehmann, E. L., 1979. Descriptive statistics for nonparametric models. IV. Spread. In: Jurečková,, J. (Ed.), Contributions to Statistics: Jaroslav Hájek Memorial Volume, Academia, Prague, 33–40.

Blom, G., 1976. Some properties of incomplete U-statistics. Biometrika 63, 573–580.

Brown, B. M., Kildea, D. G., 1978. Reduced U-statistics and the Hodges-Lehmann estimator. Ann. Statist. 6, 828–835.

Chakraborty, B., 2001. On affine equivariant multivariate quantiles. Ann. Inst. Statist. Math. 53, 380–403.

Chaudhuri, P., 1992. Multivariate location estimation using extension of R-estimates through U-statistics type approach. Ann. Statist. 20, 897–916.

Chaudhuri, P., 1996. On a geometric notion of quantiles for multivariate data. J. Amer. Statist. Assoc. 91, 862–872.

Cheng, Y., De Gooijer, J. G., 2007. On the uth geometric conditional quantile. J. Statist. Plann. Inference 137, 1914–1930.

Choudhury, J., Serfling, R., 1988. Generalized order statistics, Bahadur representations, and sequential nonparametric fixed-width confidence intervals. J. Statist. Plann. Inference 19, 269–282.

Deheuvels, P., Mason, D. M., 1990. Bahadur-Kiefer type processes. Ann. Probab. 18, 669–697.

Dudley, R. M., 1999. Uniform Central Limit Theorems. Cambridge University Press, Cambridge.

Dudley, R. M., Koltchinskii, V. I., 1992. The spatial quantiles. Unpublished manuscript.

Frees, E. W., 1991. Trimmed slope estimates for simple linear regression. J. Statist. Plann. Inference 27, 203–221.

Geertsema, J. C., 1970. Sequential confidence intervals based on rank tests. Ann. Math. Statist. 41, 1016–1026.

Ghosh, J. K., 1971. A new proof of the Bahadur representation of sample quantiles and an application. Ann. Math. Statist. 42, 1957–1961.

Giné, E., Zinn, J., 1986. Lectures on the central limit theorem for empirical processes. In: Lecture Notes in Mathematics 1221, Springer, New York, 50–113.

Helmers, R., Janssen, P., Serfling, R., 1988. Glivenko-Cantelli properties of some generalized empirical df's and strong convergence of generalized L-statistics. Probab. Th. and Related Fields 79, 75–93.

Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 36, 369–408.

Hodges, J. L., Lehmann, E. L., 1963. Estimates of location based on rank tests. Ann. Math. Statist. 34, 598–611.

Hössler, O. (1996). Incomplete generalized L-statistics. Ann. Statist. 24, 2631–2654.

Kemperman, J. H. B., 1987. The median of a finite measure on a Banach space. In: Dodge, Y. (Ed.), Statistical Data Analysis Based On the L_1-Norm and Related Methods, North-Holland, 217–230,

Kiefer, J., 1967. On Bahadur's representation of sample quantiles. Ann. Math. Statist. 38, 1323–1342.

Kiefer, J., 1970. Deviations between the sample quantile process and the sample df. In: Puri, M. L. (Ed.), Proceedings of Conference on Nonparametric Techniques in Statistical Inference, Cambridge University Press, 299-319.

Koltchinskii, V., 1994a. Bahadur-Kiefer approximation for spatial quantiles. In: Hoffman-Jørgensen, J., Kuelbs, J., Marcus, M. (Eds.), Probability in Banach Spaces 9, Birkhaüser, 401–415.

Koltchinskii, V., 1994b. Nonlinear transformations of empirical processes: functional inverses and Bahadur-Kiefer representations. In: Proceedings of 6th International Vilnius Conference on Probability and Mathematical Statistics, VSP-TEV, Netherlands, 423–445.

Liu, R. Y., Parelius, J. M., Singh, K., 1999. Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). Ann. Statist. 27, 783–858.

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

Möttönen, J., Oja, H., 1995. Multivariate spatial sign and rank methods. J. Nonpar. Statist. 5, 201–213.

Möttönen, J., Oja, H., Serfling, R., 2004. Multivariate generalized spatial signed-rank methods. J. Statist. Res. 39, 25–48.

Möttönen, J., Oja, H., Tienari, J., 1997. On the efficiency of multivariate spatial sign and rank tests. Ann. Statist. 25, 542–552.

Niemiro, W., 1992. Asymptotics for M-estimators defined by convex minimization. Ann. Statist. 20, 1514–1533.

Niinimaa, A., Oja, H., Tableman, M., 1990. On the finite sample breakdown point of the Oja bivariate median and of the corresponding half-samples version. Statist. Probab. Lett. 10, 325–328.

Oja, H., 1983. Descriptive statistics for multivariate distributions. Statist. Probab. Lett. 1, 327–333.

Oja, H., Niinimaa, A., 1984. On robust estimation of regression coefficients. Research report, Department of Applied Mathematics and Statistics, University of Oulu, Finland.

Rousseeuw, P. J., Croux, C., 1993. Alternatives to the median absolute deviation. J. Amer. Statist. Assoc. 88, 1273–1283.

Rousseeuw, P. J., Leroy, A. M., 1987. Robust Regression and Outlier Detection. John Wiley & Sons, New York.

Searle, S. R., 1982. Matrix Algebra Useful for Statistics. Wiley, New York.

Sen, P. K., 1968. Estimates of the regression coefficient based on Kendall's tau. J. Amer. Statist. Assoc. 63, 1379–1389.

Serfling, R., 1980. Approximation Theorems of Mathematical Statistics. John Wiley & Sons, New York.

Serfling, R., 2002a. Quantile functions for multivariate analysis: approaches and applications. Statist. Neerl. 56, 214–232.

Serfling, R., 2002b. Efficient and robust fitting of lognormal distributions. N. Amer. Actuarial J. 4, 95–109.

Serfling, R., 2004. Nonparametric multivariate descriptive measures based on spatial quantiles. J. Statist. Plann. Inference 123, 259–278

Shamos, M. I., 1976. Geometry and statistics: problems at the interface. In: Traub, J. F. (Ed.), New Directions and Recent Results in Algorithms and Complexity, Academic Press, 251–280.

Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis, III. Proc. Kon. Ned. Akad. v. Wetensch. A 53, 1397–1412.

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

Vardi, Y., Zhang, C.-H., 2000. The multivariate L_1-median and associated data depth. Proc. Nat. Acad. Sci. USA 97, 1423–1426.

Zhou, W., Serfling, R., 2007. Generalized multivariate rank type test statistics via spatial U-quantiles. Statist. Probab. Lett., to appear.

Zuo, Y., Serfling, R., 2000a. General notions of statistical depth function. Ann. Statist. 28, 461–482.

Zuo, Y., Serfling, R., 2000b. Nonparametric notions of multivariate "scatter measure" and "more scattered" based on statistical depth functions. J. Multiv. Analysis 75, 62–78.