INFERENCES ABOUT THE SLOPE IN LINEAR REGRESSION WITH MISSING RESPONSES: AN EMPIRICAL LIKELIHOOD APPROACH

BY URSULA U. MÜLLER∗ HANXIANG PENG† AND ANTON SCHICK‡

Texas A&M University, Indiana University Purdue University at Indianapolis and Binghamton University

This article considers linear regression models with responses that are allowed to be missing at random, which covers the model with fully observed data as a special case. We assume that the covariates and the errors are independent, without specifying their distributions. The main result of this article is to present two efficient maximum empirical likelihood estimators for the regression parameter, which are easy to obtain numerically. This fills a gap in the literature which does not provide a parameter estimator that is both simple and efficient: the usual efficient approaches require an estimator of the influence function, which can be quite involved. We also present an asymptotic confidence interval for the slope. The proofs are carried out in the model with fully observed data. This suffices since then the transfer principle by Koul et al. (2012) applies, which provides an easy method to adapt the results to the missing data case.

September 5, 2012

1. Introduction. We consider the homoscedastic regression model in which the response variable \( Y \) is linked to a (one-dimensional) covariate \( X \) by the formula

\[
Y = \beta X + \varepsilon,
\]

where \( \beta \) is an unknown real number, the covariate \( X \) has a finite positive variance, and the unobservable error variable \( \varepsilon \) is independent of \( X \), possesses a density \( f \) and has finite variance \( \sigma^2 \). We do not assume that the error variable is centered, i.e., \( E[\varepsilon] = 0 \), as is common in the model with an intercept. We think of the intercept as being absorbed in the error.

∗Supported by NSF Grant DMS-0907014
†Supported by NSF Grant DMS-0940365
‡Supported by NSF Grant DMS-0906551

AMS 2000 subject classifications: Primary 62G05; secondary 62G15,62G20

Keywords and phrases: maximum empirical likelihood estimator; infinitely many constraints; estimated constraint functions; semiparametric inference; efficiency; confidence region; missing at random; complete case analysis, transfer principle
In the ideal case one observes independent copies \((X_1, Y_1), \ldots, (X_n, Y_n)\) of \((X, Y)\). However, in real life data sets some data are typically missing. We are interested in the common situation that responses \(Y\) are missing at random (MAR). This means that we observe copies of the triplet \((\delta, X, \delta Y)\), where \(\delta\) is an indicator variable with \(\delta = 1\) if \(Y\) is observed, and where the probability \(\pi\) that \(Y\) is observed depends only on the covariate, 

\[
P(\delta = 1 | X, Y) = P(\delta = 1 | X) = \pi(X),
\]

with \(E[\pi(X)] = E[\delta] > 0\); we refer to the monographs by Little and Rubin (2002) and by Tsiatis (2006) for further readings. Note that the just described “MAR model” covers the “full model” (in which all data are completely observed) as a special case with \(\pi(X) = 1\). Our goal is to make inference about the slope \(\beta\) in this model treating the error density \(f\) and the distribution of \(X\) as nuisance parameters. We shall do so by using an empirical likelihood approach based on the complete cases, i.e. only the \(N = \sum_{j=1}^{n} \delta_j\) observations \((X_{i_1}, Y_{i_1}), \ldots, (X_{i_N}, Y_{i_N})\) with observed responses are considered. This is the simplest approach to deal with missing data, but it actually yields an semiparametrically efficient estimator of \(\beta\).

To establish efficiency we need to assume that \(f\) has finite Fisher information for location. This means that \(f\) is absolutely continuous and the integral 

\[
J_f = \int \ell_f^2(y) f(y) \, dy
\]

is finite, where \(\ell_f = -f'/f\) denotes the score function for location. It follows from Bickel (1982) that an efficient estimator \(\hat{\beta}\) of \(\beta\) in the full model is characterized by the stochastic expansion

\[
(1.2) \quad \hat{\beta} = \beta + \frac{1}{n} \sum_{j=1}^{n} (X_j - E[X]) \frac{\ell_f(Y_j - \beta X_j)}{J_f \text{Var}(X)} + o_{P}(n^{-1/2}).
\]

Müller (2009, Lemma 5.1) characterizes efficient estimators of \(\beta\) in the MAR model, but with the additional assumption that the errors have mean zero. Inspecting the proofs in that paper it is easy to see that an efficient estimator \(\hat{\beta}_c\) of \(\beta\) in the model considered here is characterized by

\[
(1.3) \quad \hat{\beta}_c = \beta + \frac{1}{N} \sum_{j=1}^{n} \delta_j (X_j - E[X|\delta = 1]) \frac{\ell_f(Y_j - \beta X_j)}{J_f \text{Var}(X|\delta = 1)} + o_{P}(n^{-1/2}).
\]

This is indeed the expansion of a complete case version of an efficient estimator \(\hat{\beta}\) that satisfies (1.2), which follows from the transfer principle for asymptotically linear statistics by Koul et al. (2012). Hence the complete case version \(\hat{\beta}_c\) of an efficient estimator \(\hat{\beta}\) remains efficient in the MAR model.
Efficient estimators were constructed by Bickel (1982), Schick (1987), Jin (1992). Koul and Susarla (1983) do so in the case when \( f \) is also symmetric about zero. See also Schick (1993) and Forrester et al. (2003) 

Our proposed empirical likelihood approach uses an increasing number of estimated constraints. This differs quite drastically from the empirical likelihood introduced by Owen (1988, 2001) which uses a fixed number of known linear constraints. More literature on EL.

Let us explain the idea of our approach, in particular the construction of our empirical likelihood which is crucial since we must exploit the independence between the covariates and the errors in order to obtain efficient maximum empirical likelihood estimators. For reasons of clarity we will begin with the full model. The corresponding approach for the missing data model is then straightforward: just proceed in the same way, now with \( N \) complete cases \((X_{i1}, Y_{i1}), \ldots, (X_{iN}, Y_{iN})\) instead of the \( n \) observations from the full model, and treat the random sample size \( N \) just like \( n \).

Our empirical likelihood \( \mathcal{R}_n = \mathcal{R}_n(b) \) which we want to maximize with respect to \( b \in \mathbb{R} \) is of the form

\[
\mathcal{R}_n(b) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j(X_j - \bar{X})v_n(F_b(Y_j - bX_j)) = 0 \right\}.
\]

Here \( \mathcal{P}_n \) is the probability simplex in dimension \( n \),

\[
\mathcal{P}_n = \{ \pi = (\pi_1, \ldots, \pi_n) \top \in [0, 1]^n : \sum_{j=1}^n \pi_j = 1 \}.
\]

\( \bar{X} \) is the sample mean of the covariates and \( v_n \) is a \( r_n \)-dimensional vector of functions which we will describe below; see (1.4). The constraint \( \sum_{j=1}^n \pi_j(X_j - \bar{X})v_n(F_b(Y_j - bX_j)) = 0 \) in the definition of \( \mathcal{R}_n \) is therefore a vector of \( r_n \) one-dimensional constraints, where the integer \( r_n \) tends to infinity slowly with the sample size \( n \). These constraints emerge from the independence assumption as follows. Independence of \( X \) and \( \varepsilon \) implies that \( \mathbb{E}\{(X - EX)a(\varepsilon)\} = 0 \) for any function \( a \in L_2(F) \), where \( F \) is the distribution function of \( \varepsilon \). If \( a \) is a constant then this is always satisfied. We therefore restrict attention to \( a \in L_{2,0}(F) \) and the family of constraints \( \mathbb{E}\{(X - EX)a(\varepsilon)\} = 0, a \in L_{2,0}(F) \). We assume that \( F \) is continuous. Then \( F(\varepsilon) \) is uniformly distributed on the interval \([0, 1] \), \( F(\varepsilon) \sim \mathcal{U} \), and an orthonormal basis of \( L_{2,0}(F) \) is \( \varphi_1 \circ F, \varphi_2 \circ F, \ldots \) where the \( \varphi_k \) denote a basis of \( L_2(\mathcal{U}) \). This suggests the constraints

\[
\sum_{j=1}^n \pi_j\{X_j - E(X)\}\varphi_k\{F(Y_j - bX_j)\} = 0, \quad k = 1, \ldots, r_n,
\]
which, however, cannot be used since neither $F$ nor the mean of $X$ are known. Replacing them by empirical estimators yields our empirical likelihood $R_n$ from above, with

$$F_b(t) = \frac{1}{n} \sum_{j=1}^{n} 1[Y_j - bX_j \leq t], \quad t \in \mathbb{R}.$$ 

Here we will work with the trigonometric basis

$$\varphi_k(x) = \sqrt{2} \cos(k\pi x), \quad 0 \leq x \leq 1, k = 1, 2, \ldots$$

and take

(1.4) $$v_n = (\varphi_1, \ldots, \varphi_{rn})^\top.$$ 

Let us finally briefly discuss the complete case approach that we propose for the MAR model. In the following a subscript “c” will indicate when a complete case statistic is used, for example $F_{b,c}$ is the complete case version of $F_b$,

$$F_{b,c}(t) = \frac{1}{N} \sum_{j=1}^{n} \delta_j 1[Y_j - bX_j \leq t] = \frac{1}{N} \sum_{j=1}^{N} 1[Y_{ij} - bX_{ij} \leq t], \quad t \in \mathbb{R}.$$ 

The complete case empirical likelihood is

$$R_c(b) = \sup \left\{ \prod_{j=1}^{N} N\pi_j : \pi \in \mathcal{P}_N, \sum_{j=1}^{N} \pi_j (X_{ij} - \bar{X}_c)v_N(F_{b,c}(Y_{ij} - bX_{ij})) = 0 \right\}.$$ 

Note that the choice $\bar{X}_c$ is intentional: when switching from the full data model to the complete cases the distribution of $(X, Y)$ changes to the conditional distribution given $\delta = 1$ (see Koul et al. (2012)). The MAR assumption combined with the independence of $X$ and $\varepsilon$ yield that $\varepsilon$ and $(X, \delta)$ are independent. Hence the parameters $F$ and $\beta$ remain the same. Only the covariate distribution changes to the conditional distribution of $X$ given $\delta = 1$. The estimator $\bar{X}_c = N^{-1} \sum_{j=1}^{n} \delta_j X_j$ indeed estimates $E[X|\delta = 1]$, as desired.

Our empirical likelihood is similar to the one considered for the symmetric location model in Peng and Schick (2012c). We shall derive analogous results as there. The paper is organized as follows. (...
2. Wilks’ theorem. Consider the empirical likelihood $R_n$ and the complete case empirical likelihood $R_c$ from the introduction. Our first result is a version of Wilks’ theorem. Write $\chi_\gamma(d)$ for the $\gamma$-quantile of the chi-square distribution with $d$ degrees of freedom.

**Theorem 1.** Consider the full model and suppose that $X$ has a finite fourth moment and that $r_n$ satisfies $r_n^4 = o(n)$. Then we have

$$P(-2 \log R_n(\beta) \leq \chi_u(r_n)) \to u, \quad 0 < u < 1.$$

Since the statistic $-2 \log R_n$ has a limiting chi-square distribution it is asymptotically distribution free, i.e. its asymptotic distribution does not depend on the distribution of $X$ and $Y$. Hence its complete case version has the same limiting distribution, which follows from the transfer principle by Koul et al. (2012). This only requires that the conditional covariate distribution is in the same model as the unconditional covariate distribution. Although the result for the MAR model is more general than the result for the full model (which is covered as a special case), we will formulate it as a corollary, since we only have to prove Theorem 1.

**Corollary 1.** Consider the MAR model and suppose that the distribution of $X$ given $\delta = 1$ has a finite fourth moment and that $r_N$ satisfies $r_N^4 = o_p(N)$. Then we have

$$P(-2 \log R_c(\beta) \leq \chi_u(r_n)) \to u, \quad 0 < u < 1.$$

Note that the conditions on the number of basis functions $r_n$ and $r_N$ in the full model and the MAR model are equivalent since we require that $E[\delta] > 0$.

The above result shows that

$$\{b \in \mathbb{R} : -2 \log R_c(b) < \chi_{1-\alpha}(r_N)\}$$

is a $1 - \alpha$ confidence region for $\beta$ and that

$$1[-2 \log R_c(\beta_0) \geq \chi_{1-\alpha}(r_N)]$$

is a test of asymptotic size $\alpha$ for testing the null hypothesis $H_0 : \beta = \beta_0$. Note that both the confidence region and the test about the slope also apply to the special case of a full model.
3. Efficient estimation. Our next result gives a uniform local asymptotic normality condition for the local empirical likelihood ratio

\[ \mathcal{L}_n(t) = \log \left( \frac{\mathcal{R}_n(\beta + n^{-1/2}t)}{\mathcal{R}_n(\beta)} \right), \quad t \in \mathbb{R}, \]

in the full model.

**Theorem 2.** Suppose \( X \) has a finite fourth moment, \( f \) has finite Fisher information for location, and \( r_n \) satisfies \( r_n^5 = o(n) \). Then, for every finite \( C \), the uniform expansion

\[
\sup_{|t| \leq C} |\mathcal{L}_n(t) - t \Gamma_n + J_f \text{Var}(X) t^2/2| = o_p(1)
\]

holds with

\[
\Gamma_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j - E[X_j]) \ell_f(X_j - \beta X_j)
\]

which is asymptotically normal with mean zero and variance \( J_f \text{Var}(X) \).

Having obtained this expansion, we can now proceed as in Peng and Schick (2012b) and address maximum empirical likelihood estimation. Peng and Schick show that the random function \( b \mapsto \mathcal{R}_n(b) \) has a local maximum \( \hat{\beta} \) under the uniform local asymptotic normality condition of the previous theorem. We study a version of their approach, namely (guided) one-step maximum empirical likelihood estimation, in order to guarantee that the local maximizer is unique with probability tending to one. If this is met then \( \hat{\beta} \) will obey expansion (1.2). By the discussion in the introduction, the transfer principle for asymptotically linear statistics then implies that the complete case maximum empirical likelihood estimator \( \hat{\beta}_c \) has expansion (1.3), and is therefore efficient. Note that we are primarily interested in obtaining efficient estimators, i.e. in estimators that satisfy expansions (1.2) and (1.3). Hence we need to formulate Theorem 2 only for the full model, since the corresponding result for the MAR model is not relevant for the proofs.

Let \( \hat{\beta} \) denote the least squares estimators in the full model and \( \hat{\beta}_c \) its complete case version in the MAR model,

\[
\hat{\beta} = \frac{\sum_{j=1}^{n} (X_j - \bar{X})Y_j}{\sum (X_j - \bar{X})^2}, \quad \hat{\beta}_c = \frac{\sum_{j=1}^{n} \delta_j (X_j - \bar{X}_c)Y_j}{\sum \delta_j (X_j - \bar{X}_c)^2}.
\]
These estimators are $n^{1/2}$-consistent, i.e., $n^{1/2}(\hat{\beta} - \beta) = O_p(1)$. A one-step maximum empirical likelihood estimator in the full model is defined as

\[
\hat{\beta}_I = \hat{\beta} + \frac{n^{-1/2}\log R_n(\hat{\beta} + an^{-1/2}) - \log R_n(\hat{\beta} - an^{-1/2})}{\log R_n(\hat{\beta} + 2an^{-1/2}) - 2 \log R_n(\hat{\beta}) + \log R_n(\hat{\beta} - 2an^{-1/2})}/2a
\]

for some positive constant $a$. Its complete case version for the full model is

\[
\hat{\beta}_{I,c} = \hat{\beta}_c + 2aN^{-1/2} \times \frac{\log R_c(\hat{\beta}_c + aN^{-1/2}) - \log R_c(\hat{\beta}_c - aN^{-1/2})}{\log R_c(\hat{\beta}_c + 2aN^{-1/2}) - 2 \log R_c(\hat{\beta}_c) + \log R_c(\hat{\beta}_c - 2aN^{-1/2})}
\]

Finally, the guided one-step estimators for the full model and the MAR model are defined as

\[
\hat{\beta}_{II} = \arg \max_{n^{1/2}|b - \hat{\beta}_I| \leq c} R_n(b), \quad \hat{\beta}_{II,c} = \arg \max_{N^{1/2}|b - \hat{\beta}_{I,c}| \leq c} R_c(b),
\]

for some finite $c$.

**Theorem 3.** Suppose that the error density $f$ has finite Fisher information for location and that $r_n$ satisfies $r_n^5 = o(n)$.

(a) Let the distribution of $X$ have a finite fourth moment. Then the one-step maximum empirical likelihood estimator $\hat{\beta}_I$ and the guided one-step maximum likelihood estimator $\hat{\beta}_{II}$ satisfy expansion (1.2) and hence are efficient in the full model.

(b) Consider the MAR model and assume now that the conditional distribution of $X$ given $\delta = 1$ has a finite fourth moment. Then the complete case versions $\hat{\beta}_{I,c}$ and $\hat{\beta}_{II,c}$ of $\hat{\beta}_I$ and $\hat{\beta}_{II}$ from part (a) satisfy expansion (1.3) and hence are efficient in the MAR model.

4. **Proofs.**

4.1. **Proof of Theorem 1.** Let $\mu$ denote the mean and $\tau$ denote the standard deviation of $X$. We should point out that $R_n(\theta)$ does not change if we replace $(X_j - \bar{X})$ by $(X_j - \bar{X})/\tau = V_j - \bar{V}$, where

\[
V_j = \frac{X_j - \mu}{\tau} \quad \text{and} \quad \bar{V} = \frac{1}{n} \sum_{j=1}^{n} V_j.
\]
Thus, for the purpose of our proofs, we may assume that
\[
\mathcal{R}_n(\partial) = \sup \left\{ \prod_{j=1}^{n} n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^{n} \pi_j(V_j - \bar{V})v_n(\mathbb{F}_\theta(Y_j - \theta x_j)) = 0 \right\}.
\]

In what follows we shall repeatedly use the bounds
\[
|v_n(y)|^2 \leq 2r_n, \quad |v'_n(y)|^2 \leq 2\pi^2r_n^3, \quad \text{and} \quad |v''_n(y)|^2 \leq 2\pi^4r_n^5,
\]
valid for all real \(y\).

Let us set \(Z_j = V_j v_n(F(\varepsilon_j))\) and \(\bar{Z}_j = (V_j - \bar{V})v_n(\bar{F}_\theta(\varepsilon_j))\), \(j = 1, \ldots, n\).

With \(Z = Z_1\), we find the identities \(E[Z] = 0\) and \(E[ZZ^\top] = I_{r_n}\), where \(I_{r_n}\) is the \(r_n \times r_n\) identity matrix, and the bound \(E[|Z|^4] \leq (2r_n)^2E[V^4] = O(r_n^2)\).

As shown in Peng and Schick (2012a) these results yield
\[
\bar{Z}_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j = O_p(r_n^{1/2})
\]
and
\[
\sup_{|u| = 1}\left| \frac{1}{n} \sum_{j=1}^{n} u^\top Z_j \right| - 1 \leq \left| \frac{1}{n} \sum_{j=1}^{n} Z_j Z_j - 1 \right| = O_p(r_n n^{-1/2}).
\]

In view of Corollary 7.6 in Peng and Schick (2012a) and \(r_n^4 = o(n)\), the desired result follows if we verify
\[
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\bar{Z}_j - Z_j) = o_P(1) \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^{n} |\bar{Z}_j - Z_j|^2 = o_P(r_n^3/n).
\]

In view of the identity \(\bar{Z}_j - Z_j = V_j \Delta_j - \bar{V} \Delta_j - \bar{V} v_n(\mathbb{F}_\theta(\varepsilon_j))\), where
\[
\Delta_j = v_n(\mathbb{F}_\theta(\varepsilon_j)) - v_n(F(\varepsilon_j)), \quad j = 1, \ldots, n,
\]
the bound \(|v_n|^2 \leq 2r_n\), and the fact \(n^{1/2}\bar{V} = O_p(1)\), it is easy to see the desired results follow from the following rates,
\[
S_1 = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j \Delta_j = O_p(r_n^{3/2}n^{-1/2}),
\]
\[
S_2 = \frac{1}{n} \sum_{j=1}^{n} \Delta_j = o_P(r_n^{3/2}n^{-1/2}),
\]
\[
S_3 = \frac{1}{n} \sum_{j=1}^{n} v_n(F(\varepsilon_j)) = O_p(r_n^{1/2}n^{-1/2}),
\]
\[
S_4 = \frac{1}{n} \sum_{j=1}^{n} V_j^2 |\Delta_j|^2 = O_p(r_n^3n^{-1}).
\]
Note that $\Delta_1, \ldots, \Delta_n$ are functions of the errors $\varepsilon_1, \ldots, \varepsilon_n$ only and satisfy

$$M_n = \max_{1 \leq j \leq n} |\Delta_j|^2 \leq 2\pi^2 r_n^3 \sup_{t \in \mathbb{R}} |F_\beta(t) - F(t)|^2 = O_p(r_n^3/n).$$

Conditioning on the errors thus yields

$$E[|S_1|^2|\varepsilon_1, \ldots, \varepsilon_n] = E[S_4|\varepsilon_1, \ldots, \varepsilon_n] \leq M_n.$$ 

This establishes the rates for $S_1$ and $S_4$. The other rates follow from $|S_2|^2 \leq M_n$ and $E[|S_3|^2] = E[|v_n(F(\varepsilon))|^2]/n = r/n$.

### 4.2. Proof of Theorem 2

For $t$ in $\mathbb{R}$, we let $\hat{F}_{nt} = F_{\beta + n^{-1/2}t}$ and note that $\hat{F}_{nt}$ is the empirical distribution function of the random variables

$$\varepsilon_{jt} = \varepsilon_j - n^{-1/2}tX_j, \quad j = 1, \ldots, n.$$ 

These random variables are independent with common distribution function $F_{nt}$ given by

$$F_{nt}(y) = E[\hat{F}_{nt}(y)] = E[F(y + n^{-1/2}tX)], \quad y \in \mathbb{R}.$$ 

To simplify notation we introduce

$$\hat{R}_{jt} = \hat{F}_{nt}(\varepsilon_{jt}), \quad R_{jt} = F_{nt}(\varepsilon_{jt}), \quad R_j = F(\varepsilon_j),$$

and

$$\hat{Z}_{jt} = (V_j - \bar{V})v_n(\hat{R}_{jt}), \quad Z_{jt} = V_jv_n(R_{jt}), \quad Z_j = V_jv_n(R_j).$$

Since we work with the form of the empirical likelihood given in the previous section, we have

$$\mathcal{R}_n(\beta + n^{-1/2}t) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j\hat{Z}_{jt} = 0 \right\}, \quad t \in \mathbb{R}.$$ 

The desired result follows if we verify the uniform expansion

$$\sup_{|t| \leq C} |2\log \mathcal{R}_n(\beta + n^{-1/2}t) - |\hat{Z}_n|^2 + 2t\bar{\Gamma}_n - t^2\tau^2 J_f| = o_P(1)$$

for every finite positive constant $C$ with $\hat{Z}_n$ as in (4.1). Now fix such a constant $C$. To verify (4.3) we introduce

$$\mu_n(t) = E[V_n(F(\varepsilon))\rho_{nt}(\varepsilon, X)]$$
with
\[
\rho_{nt}(\varepsilon, X) = \frac{-2n^{1/2}(f^{1/2}(\varepsilon + n^{-1/2}tX) - f^{1/2}(\varepsilon))}{f^{1/2}(\varepsilon)} 1[f(\varepsilon) > 0].
\]

We shall establish (4.3) by verifying the following six conditions.

(4.4) \( \sup_{|t| \leq C \ |u| = 1} \left| \frac{1}{n} \sum_{j=1}^{n} (u^\top \hat{Z}_{jt})^2 - 1 \right| = o_P(1/r_n), \)

(4.5) \( \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\hat{Z}_{jt} - Z_{jt}) \right| = o_P(r_n^{3/2}n^{-1/2} \log n), \)

(4.6) \( \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (Z_{jt} - Z_j - E[Z_{jt} - Z_j]) \right| = o_P(r_n^{3/2}n^{-1/2}), \)

(4.7) \( \sup_{|t| \leq C} \left| n^{1/2}E[Z_{1t} - Z_1] + \mu_n(t) \right| = O(r_n^{3/2}n^{-1/2}), \)

(4.8) \( \sup_{|t| \leq C} \left| \left\| \mu_n(t) \right\|^2 - t^2 \tau^2 J_f \right| \to 0, \)

(4.9) \( \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [\mu_n(t) \top Z_j - t\tau V_j \ell_f(\varepsilon_j)] \right| = o_P(1). \)

Conditions (4.6)–(4.9) are verified in Section 4.3; Section 4.4 contains the proofs of conditions (4.4) and (4.5). Here we establish their sufficiency.

**Lemma 1.** The conditions (4.4)–(4.9) imply (4.3).

To prove this lemma, we use the following result which is a special case of Lemma 5.2 in Peng and Schick (2012a). This version was already used in Schick (2012).

**Lemma 2.** Let \( x_1, \ldots, x_n \) be \( m \)-dimensional vectors. Set
\[
\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j, \quad x^* = \max_{1 \leq j \leq n} |x_j|, \quad \nu_4 = \frac{1}{n} \sum_{j=1}^{n} |x_j|^4, \quad S = \frac{1}{n} \sum_{j=1}^{n} x_j x_j^\top,
\]
and let \( \lambda \) denote the smallest and \( \Lambda \) the largest eigenvalue of the matrix \( S \). Then the inequality \( \lambda > 5|\bar{x}|x^* \) implies

\[
-2 \log(\mathcal{R}) - n\bar{x}^T S^{-1}\bar{x} \leq \frac{n|\bar{x}|^3(\Lambda \nu_4)^{1/2}}{(\lambda - |\bar{x}|x^*)^3} + \frac{4n\Lambda^2|\bar{x}|^4\nu_4}{\lambda^2(\lambda - |\bar{x}|x^*)^4}
\]

where

\[
\mathcal{R} = \sup \left\{ \prod_{j=1}^{n} n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^{n} \pi_j x_j = 0 \right\}
\]

**Proof of Lemma 1.** We introduce

\[
T(t) = \frac{1}{n} \sum_{j=1}^{n} \hat{Z}_{jt} \quad \text{and} \quad S(t) = \frac{1}{n} \sum_{j=1}^{n} \hat{Z}_{jt} \hat{Z}_{jt}^\top,
\]

and let \( \lambda_n(t) \) and \( \Lambda_n(t) \) denote the smallest and largest eigenvalues of \( S(t) \),

\[
\lambda_n(t) = \inf_{|u|=1} u^T S(t) u = \inf_{|u|=1} \frac{1}{n} \sum_{j=1}^{n} (u^T \hat{Z}_{jt})^2
\]

and

\[
\Lambda_n(t) = \sup_{|u|=1} u^T S(t) u = \sup_{|u|=1} \frac{1}{n} \sum_{j=1}^{n} (u^T \hat{Z}_{jt})^2.
\]

By (4.4), we have

\[
\sup_{|t| \leq C} |\lambda_n(t) - 1| = o_p(1) \quad \text{and} \quad \sup_{|t| \leq C} |\Lambda_n(t) - 1| = o_p(1).
\]

The conditions (4.5)–(4.7) imply

\[
\sup_{|t| \leq C} |n^{1/2} T(t) - \hat{Z}_n + \mu_n(t)| = o_p(r_n^{-1/2}).
\]

This, (4.1) and (4.8) yield

\[
\sup_{|t| \leq C} n|T(t)|^2 = O_p(r_n).
\]

Next, we find

\[
\sup_{|t| \leq C} \max_{1 \leq j \leq n} |\hat{Z}_{jt}| \leq (2r_n)^{1/2} \max_{1 \leq j \leq n} |V_j|^4 = o_p(r_n^{1/2}n^{1/4})
\]

and

\[
\sup_{|t| \leq C} \frac{1}{n} \sum_{j=1}^{n} |\hat{Z}_{jt}|^4 \leq (2r_n)^2 \frac{1}{n} \sum_{j=1}^{n} |V_j|^4 = O_p(r_n^2).
\]
Thus we derive
\begin{equation}
\sup_{|t| \leq C} \left| -2 \log \mathcal{R}_n(\beta + n^{-1/2}t) - nT(t)^\top (S(t)y)^{-1}T(t) \right| = o_P(1)
\end{equation}

since by Lemma 2 the left-hand side is of order $O_P(r^{5/2}n^{-1/2} + r_1^4/n)$. For a positive definite matrix $A$ and a compatible vector $z$, we have

\begin{equation}
|x^\top A^{-1}x - x^\top x| \leq x^\top A^{-1}x \sup_{|u| = 1} |1 - u^\top Au| \leq \frac{|x|^2}{\lambda} \sup_{|u| = 1} |1 - u^\top Au|
\end{equation}

with $\lambda$ the smallest eigenvalue of $A$. Using this, (4.4) and (4.12) we derive
\begin{equation}
\sup_{|t| \leq C} \left| n|T(t)^\top (S(t))^{-1}T(t) - T(t)^\top T(t) | = o_P(1).
\end{equation}

With the help of (4.1), (4.11) and (4.8) we verify
\begin{equation}
\sup_{|t| \leq C} \left| n|T(t)^\top (S(t))^{-1}T(t) - T(t)^\top T(t) | = o_P(1).
\end{equation}

The desired result (4.3) follows from (4.13), (4.14), (4.15) (4.8) and (4.9).

4.3. Proofs of (4.6)-(4.9). We begin by mentioning properties of $f$ and $F$ which are crucial to the proofs. Since $f$ has finite Fisher information for location and $X$ has a finite second moment, we have
\begin{equation}
E \int (f^{1/2}(y - tX) - f^{1/2}(y) - (tX/2)\ell_f(y)f^{1/2}(y))^2 dy = o(|t|^2).
\end{equation}

One also has the inequalities
\begin{equation}
\int (f^{1/2}(y + t) - f^{1/2}(y + s))^2 dy \leq J_f(t - s)^2/4,
\end{equation}
\begin{equation}
\int |f(y + t) - f(y + s)| dy \leq B_1|t - s|,
\end{equation}
\begin{equation}
|F(t) - F(s)| \leq B_1|t - s|,
\end{equation}
\begin{equation}
|F(t + s) - F(t) - sf(t)| \leq B_2|s|^{3/2}
\end{equation}

for all real $s$ and $t$, and some constants $B_1$ and $B_2$, see e.g. Peng and Schick (2012c).
Next, we look at the process

\[ H_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [h_{nt}(X_j, Y_j) - E[h_{nt}(X, Y)]], \quad t \in \mathbb{R}, \]

where \( h_{nt} \) are measurable functions from \( \mathbb{R}^2 \) to \( \mathbb{R}^{m_n} \) such that \( h_{n0} = 0 \).

We are interested in the cases \( m_n = 1 \) and \( m_n = r_n \). A version of the following lemma has already been used in Peng and Schick (2012c).

**Lemma 3.** Suppose that the map \( t \mapsto h_{nt}(x, y) \) is continuous for all \( x \) and \( y \) in \( \mathbb{R} \) and

\[ E[|h_{nt}(X, Y) - h_{ns}(X, Y)|^2] \leq K_n |t - s|^2, \quad s, t \in \mathbb{R}, \]

for positive constants \( K_n \). Then we have the rate

\[ \sup_{|t| \leq C} |H_n(t)| = O_p(K_n^{1/2}). \]

If also \( E[|h_{nt}(X, Y)|^2] = o(K_n) \) holds for all \( t \), then we have the rate

\[ \sup_{|t| \leq C} |H_n(t)| = O_p(K_n^{1/2}). \]

**Proof of (4.6).** The desired result follows from Lemma 3 applied with

\[ h_{nt}(X, Y) = V[v_n(F_{nt}  \varepsilon - n^{-1/2} t X)) - v_n(F  \varepsilon)), \quad t \in \mathbb{R}, \]

and \( K_n = 2\pi^2 r_n^3 B^2 n E[V(X_1 - X)^2]/n \). Indeed, we have \( h_{n0} = 0 \) and (4.21) in view of (4.19). \( \square \)

**Proof of (4.7).** Since \( V \) has mean zero, we obtain the identity

\[ E[Z_{1t} - Z_1] = E[V_1(v_n(R_{1t}) - v_n(R_1))] = \Delta_1(t) + \Delta_2(t) \]

with

\[ \Delta_1(t) = E[V \int v_n(F(y)(f(y + n^{-1/2} t X) - f(y)) dy] \]

\[ \Delta_2(t) = E[V \int v_n(F_{nt}(y)) - v_n(F(y))(f(y + n^{-1/2} t X) - f(y)) dy] \]

Using the identity \( a - b = (\sqrt{a} - \sqrt{b})^2 + 2(\sqrt{a} - \sqrt{b})\sqrt{b} \) valid for nonnegative \( a \) and \( b \), and the inequality (4.17) we see that

\[ \sup_{|t| \leq C} \left| E[n^{1/2} \Delta_1(t) + \mu_n(t)] \right| \leq (2r_n)^{1/2} C^2 n^{-1/2} J_f E[|V||X|^2]/4. \]
It follows from (4.18) and (4.19) that
\[
\sup_{|t| \leq C} n^{1/2} |\Delta_2(t)| \leq (2\pi^2 m_n^2)^{1/2} B_1 C E[|X|] B_1 C E[|V X|] n^{-1/2}.
\]
From these bounds we conclude (4.7). \qed

**Proof of (4.8).** From (4.16) and (4.17) we derive
\[
E[(\rho_{nt}(Z) - \rho_{ns}(Z))^2] \leq J_f E[X^2](t - s)^2, \quad s, t \in \mathbb{R},
\]
and
\[
E[(\rho_{nt}(Z) - tX\ell_f(\varepsilon))^2] \to 0, \quad t \in \mathbb{R}.
\]
Note the functions \(V\varphi_1(F(\varepsilon)), V\varphi_2(F(\varepsilon)), \ldots\) form an orthonormal basis of the space \(\mathcal{V} = \{V a(\varepsilon) : a \in L_{2,0}(F)\}\). Thus \(\mu_n(t)\) is the vector consisting of the first \(r_n\) Fourier coefficients of \(\rho_n(\varepsilon, X)\) with respect to this basis. Thus (4.22) and Bessel’s inequality yield \(|\mu_n(t) - \mu_n(s)|^2 \leq J_f E[X^2](t - s)^2\) for all real \(s\) and \(t\), while (4.23) and Parceval’s theorem yield \(|\mu_n(t)|^2 \to E[(t(X - \mu)\ell_f(\varepsilon))^2] = t^2 \tau^2 J_f\) for all real \(t\). In the last step we also used the fact that \(t(X - \mu)\ell_f(\varepsilon)\) is the projection of \(tX\ell_f(\varepsilon)\) onto \(\mathcal{V}\). It is now easy to see that (4.8) holds. \qed

**Proof of (4.9).** We apply Lemma 3 with \(h_{nt}(X, Y) = V \mu_n(t)^	op v_n(F(\varepsilon)) - t(X - \mu)\ell_f(\varepsilon)\). In view of the inequality \(E[(h_{nt}(X, Y) - h_{ns}(X, Y))^2] \leq 2|\mu_n(t) - \mu_n(s)|^2 + 2|t - s|^2 \tau^2 J_f\), we obtain (4.21) with \(K_n = 4\tau^2 J_f\). We also have \(E[h_{nt}^2(X, Y)] \to 0\) in view of (4.23). Thus the desired (4.9) follows from Lemma 3. \qed

4.4. Proofs of (4.4) and (4.5). We begin by deriving properties of \(\hat{R}_{ij}\) and \(\hat{R}_{jt}\) which we need in the proofs of (4.4) and (4.5). For this we introduce the leave-one-out version \(\tilde{R}_{ij}\) of \(\hat{R}_{ij}\) defined by
\[
\tilde{R}_{ij} = \frac{1}{n - 1} \sum_{i \neq j} 1[\varepsilon_{it} \leq \varepsilon_{jt}] = \frac{n}{n - 1} \hat{R}_{ij} - \frac{1}{n - 1} 1[\varepsilon_{jt} \leq \varepsilon_{jt}]
\]
which satisfies
\[
|\tilde{R}_{ij} - \hat{R}_{ij}| \leq \frac{2}{n - 1}.
\]
We abbreviate \(\tilde{R}_{i0}\) by \(\tilde{R}_i\). In the ensuing arguments we rely on the following properties of these quantities, where \(B_1\) and \(B_2\) are the constants appearing in (4.19) and (4.20).
\[
\max_{1 \leq j \leq n} \sup_{|t| \leq C} |\tilde{R}_{ij} - R_{ij} - \tilde{R}_i + R_i| = O_p(n^{-3/4}(\log n)^{1/2}),
\]
\[
\sum_{|t| \leq C} |\tilde{R}_{ij} - R_{ij} - \tilde{R}_i + R_i| = O_p(n^{-1/4}(\log n)^{1/2}).
\]
(4.26) \[ \max_{1 \leq j \leq n} |\hat{R}_j - R_j| = O_p(n^{-1/2}), \]

(4.27) \[ \sup_{|t| \leq C} |R_{jt} - R_j| \leq B_1 C n^{-1/2}(|X_j| + E[|X|]), \]

(4.28) \[ \sup_{|t| \leq C} |R_{jt} - R_j + n^{-1/2} t(X_j - \mu) f(\varepsilon_j)| \leq B_2 C^3 n^{-3/4} \sqrt{2(|X_j|^{3/2} + E[|X|^{3/2}]}. \]

The first statement is proved in Section 5, (4.26) follows from properties of the empirical distribution function, and (4.27) and (4.28) follow from (4.19) and (4.20), respectively. It follows from (4.24) – (4.27) that we have the bounds

(4.29) \[ \sup_{|t| \leq C} |\hat{R}_{jt} - R_j| \leq B_1 C n^{-1/2} |X_j| + n^{-1/2} \xi_n, \quad j = 1, \ldots, n \]

where the positive random variable \( \xi_n \) satisfies \( \xi_n = O_p(1) \).

**Proof of (4.4).** In view of (4.2), it suffices to verify

(4.30) \[ \sup_{|u| = 1} \sup_{|t| \leq C} \left| \frac{1}{n} \sum_{j=1}^{n} (u^\top \hat{Z}_{jt})^2 - \frac{1}{n} \sum_{j=1}^{n} (u^\top Z_j)^2 \right| = O_p(1/r_n). \]

Using the Cauchy-Schwarz inequality we bound the left-hand side of (4.30) by \( 2(D_n \Lambda_n)^{1/2} + D_n \) with

\[ \Lambda_n = \sup_{|u| = 1} \frac{1}{n} \sum_{j=1}^{n} (u^\top Z_j)^2 \quad \text{and} \quad D_n = \sup_{|t| \leq C} \frac{1}{n} \sum_{j=1}^{n} |\hat{Z}_{jt} - Z_j|^2. \]

Thus in view of (4.2) it suffices to prove \( D_n = O_p(1/r_n^2) \). This follows from the inequality

\[ D_n \leq \sup_{|t| \leq C} \frac{1}{n} \sum_{j=1}^{n} \left( 2 \hat{V}_n^2 |v_n(\hat{R}_{jt})|^2 + 2 \hat{V}_j^2 |v_n(\hat{R}_{jt}) - v_n(R_j)|^2 \right) \]
\[ \leq 4 \hat{V}_n^2 + 4 \pi^2 r_n^3 \frac{1}{n} \sum_{j=1}^{n} V_j^2 \sup_{|t| \leq C} |\hat{R}_{jt} - R_j| = O_p(r_n^3/n) \]

and the rate \( r_n^5 = o(n) \). \( \square \)
Proof of (4.5). In view of the rate $\bar{V} = O_p(n^{-1/2})$ and the identity
\[ \hat{Z}_{jt} - Z_{jt} = V_j(v_n(\hat{R}_{jt}) - v_n(R_j)) - \bar{V}(v_n(\hat{R}_{jt}) - v_n(R_j)) - \bar{V}v_n(R_j), \]
the desired (4.5) is implied by the following three statements.

(4.31) \[ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} v_n(R_j) = o_p(r_n^{1/2}), \]

(4.32) \[ \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (v_n(\hat{R}_{jt}) - v_n(R_j)) \right| = O_p(r_n^{3/2}), \]

(4.33) \[ \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j[v_n(\hat{R}_{jt}) - v_n(\hat{R}_{jt})] \right| = O_p(r_n^{3/2}n^{-1/2}(\log n)^{1/2}), \]

We obtain the first statement from $E[v_n(F(\epsilon))] = 0$ and $E[|v_n(F(\epsilon)|^2] = r_n$. The second statement follows from the fact that its left-hand side is bounded by
\[ (2\pi^2 r_n^3)^{1/2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} |\hat{R}_{jt} - R_j| \leq (2\pi^2 r_n^3)^{1/2} \left( \xi_n + \frac{1}{n} \sum_{j=1}^{n} B_1 C |X_j| \right) = O_p(r_n^{3/2}). \]

Using (4.24) we find
\[ \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j[v_n(\hat{R}_{jt}) - v_n(R_j)] \right| = O_p(r_n^{3/2}n^{-1/2}). \]

Taylor expansions, the bound $|v_n'''| \leq 2\pi^6 r_n^7$ and (4.29) show that
\[ \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j[v_n(\hat{R}_{jt}) - v_n(R_j) - v_n'(R_j)(\hat{R}_{jt} - R_j) - \frac{1}{2} v_n''(R_j)(\hat{R}_{jt} - R_j)^2] \right| \]

and
\[ \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j[v_n(R_j) - v_n(R_j) - v_n'(R_j)(R_j - R_j) - \frac{1}{2} v_n''(R_j)(R_j - R_j)^2] \right| \]

are of order $r_n^{7/2}n^{-1}$. Using the identity
\[ (a + b + c)^2 - a^2 - b^2 + 2ab = c^2 + 2(a + d)b + 2(a + c)c \]
with \( a = R_{jt} - R_j, b = \tilde{R}_j - R_j, c = \tilde{R}_{jt} - R_{jt} - \tilde{R}_j - R_j \) and \( d^{-1/2}t(X_j - \mu)f(\varepsilon_j) = n^{-1/2}t\tau V_jf(\varepsilon_j) \) and then Fubini’s theorem, we calculate and obtain

\[
\sup_{|t| \leq C} |(\tilde{R}_{jt} - R_j)^2 - (R_{jt} - R_j)^2 - (\tilde{R}_j - R_j)^2 + \frac{2t\tau V_jf(\varepsilon_j)}{n^{1/2}}(\tilde{R}_j - R_j)| \\
\leq \zeta_n(1 + |X_j|)^{3/2}, \quad j = 1, \ldots, n,
\]

with \( \zeta_n = O_p(n^{-5/4}\log^{1/2} n) \). If follows from the above that the left-hand side of (4.33) is bounded by \(|T_1|/2 + C\tau |T_2| + T_3 + T_4\), where

\[
T_1 = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j v''_n(R_j)(\tilde{R}_j - R_j)^2,
\]

\[
T_2 = \frac{1}{n} \sum_{j=1}^{n} V_j^2 v''_n(R_j)f(\varepsilon_j)(\tilde{R}_j - R_j),
\]

\[
T_3 = \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j v'_n(R_j)(\tilde{R}_{jt} - \tilde{R}_{jt}) \right|,
\]

and

\[
T_4 = O_p(r_n^{3/2}n^{-1/2} + r_n^{7/2}n^{-1} + r_n^{5/2}n^{-3/4}(\log n)^{1/2}) = O_p(r_n^{3/2}n^{-1/2}).
\]

We calculate

\[
E[|T_1|^2|\varepsilon_1, \ldots, \varepsilon_n] = \frac{1}{n} \sum_{j=1}^{n} |v''_n(R_j)|^2(\tilde{R}_j - R_j)^4 = O_p(r_n^{-2}).
\]

Thus \( |T_1| = o_p(r_n^{3/2}n^{-1/2}) \). Next, we write \( T_2 \) as the vector U-statistic

\[
T_2 = \frac{1}{n(n-1)} \sum_{i \neq j} V_j^2 v''_n(F(\varepsilon_j))f(\varepsilon_j)(1[\varepsilon_i \leq \varepsilon_j] - F(\varepsilon_j))
\]

and obtain

\[
E[|T_2|^2] \leq \frac{E[|k(\varepsilon)|^2]}{n} + \frac{2E[|V_2|^4|v''_n(F(\varepsilon))|^2f^2(\varepsilon_2)(1[\varepsilon_1 \leq \varepsilon_2] - F(\varepsilon_2))^2]}{n(n-1)}
\]

with \( k(x) = E[v''_n(F(\varepsilon))|f(\varepsilon)|1[x \leq \varepsilon] - F(\varepsilon)] \). Using the representation (??) and then Fubini’s theorem, we calculate

\[
k(x) = \int_{-\infty}^{\infty} v''_n(F(y))f(y)(1[x \leq y] - F(y))f(y) \, dy
\]

\[
= \int_{x}^{\infty} \left( v'_n(F(z)) - v'_n(F(x))\ell_f(z)f(z) \right) \, dz
\]

\[
- \int_{-\infty}^{x} \left[ v'(F(z))F(z) - v(F(z)) \right] \ell_f(z)f(z) \, dz.
\]
Thus $|k|$ is bounded by a constant times $r_{n}^{3/2}$ and we see that $E|T_2|^2 = O(r_{n}^{3/2} + r_{n}^{5}/n^2)$. This proves $|T_2| = O_P(r_{n}^{3/2} n^{-1/2})$.

We bound $T_3$ by the sum $T_{31} + T_{32} + T_{33}$ where

$$T_{31} = \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} W_j v_n'(R_j)(\hat{R}_{jt} - R_{jt}) \right|,$$

$$T_{32} = \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j 1[|V_j| > n^{1/4}] v_n'(R_j)(\hat{R}_{jt} - R_{jt}) \right|,$$

$$T_{33} = \sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E[V 1[|V| > n^{1/4}]] v_n'(R_j)(\hat{R}_{jt} - R_{jt}) \right|,$$

and $W_j = V_j 1[|V_j| \leq n^{-1/4}] - E[V 1[|V| \leq n^{1/4}]]$. Since $V$ has a finite fourth moment, we obtain the rates $\max_{1\leq j \leq n} |V_j| = o_P(n^{1/4})$ and $|E[V 1[|V| > n^{1/4}]]| \leq n^{3/4} E[V^4 1[|V| > n^{1/4}]] = o(n^{-3/4})$. Thus we find $P(T_{32} > 0) = P(\max_{1\leq j \leq n} |V_j| > n^{1/4}) \rightarrow 0$ and $T_{33} = o_P(n^{-3/4} r_{n}^{3/2})$ in view of (4.25) and (4.26). To deal with $T_{31}$ we express it as

$$T_{31} = \sup_{|t| \leq C} n^{1/2} \left| \frac{1}{n(n-1)} \sum_{j \neq j} W_j v_n'(F(\varepsilon_j)) \left( 1[\varepsilon_{jt} \leq \varepsilon_{jt}] - F_{nt}(\varepsilon_{jt}) \right) \right|.$$

Let us set

$$k_{nt}(z) = E[W v_n(F(z + n^{-1/2}tX))], \quad z \in \mathbb{R}.$$

Using (4.19) we obtain the bound

$$E[|k_{nt}(\varepsilon_{jt}) - k_{ns}(\varepsilon_{js})|^2] \leq 2\pi^2 r_{n}^{3} D_{t}^{2} E[W^2(X_j - X)^2]|t - s|^2/n$$

and derive with the help of Lemma 3

$$\sup_{|t| \leq C} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (k_{nt}(\varepsilon_{jt}) - E[k_{nt}(\varepsilon_{jt})]) \right| = O_P(r_{n}^{3/2} n^{-1/2}).$$

In view of this we obtain the rate $T_3 = O_P(r_{n}^{3/2} n^{-1/2} \log n)$, if we verify

$$\sup_{|t| \leq C} |U(t)| = O_P(r_{n}^{3/2} n^{-1/2} \log n),$$

where $U(t)$ is the vector U-statistic equaling

$$\frac{1}{n(n-1)} \sum_{j \neq j} W_j v_n'(F(\varepsilon_j)) \left( 1[\varepsilon_{jt} \leq \varepsilon_{jt}] - F_{nt}(\varepsilon_{jt}) \right) + k_{nt}(\varepsilon_{jt}) - E[k_{nt}(\varepsilon_{jt})].$$
It is easy to verify that $U(t)$ is degenerate. Let $t_k = -C + 2kC/n$, $k = 0, \ldots, n$. Then we have

$$\sup_{|t| \leq C} |U(t)| \leq \max_{1 \leq k \leq n} \left( |U(t_k)| + \sup_{t_{k-1} \leq t \leq t_k} |U(t) - U(t_k)| \right).$$

For $t \in [t_{k-1}, t_k]$, we find

$$|U(t) - U(t_k)| \leq (2n^2 t_n^3)^{1/2} \left(n^{1/4} (N_k^+ + N_k^-) + 2B_1 C n^{-3/2} S \right)$$

with

$$S = \frac{1}{n} \sum_{j=1}^{n} \left( |W_j|(|X_j| + E[|X|]) + E[|W|]|X_j| + 2E[|W|]X_j| + E[|W|]E[|X|] \right),$$

$$N_k^+ = \frac{1}{n(n-1)} \sum_{i \neq j} 1[t_{k-1} - 1 < t_k - t_{ij}] 1[D_{ij} < e_i - e_j \leq t_k D_{ij}] 1[D_{ij} \geq 0],$$

$$N_k^- = \frac{1}{n(n-1)} \sum_{i \neq j} 1[t_k D_{ij} < e_i - e_j \leq t_{k-1} - t_{ij}] 1[D_{ij} < 0],$$

and $D_{ij}^{-1/2}(X_i - X_j)$. We write $U_i(t)$ for the $l$-th component of the vector $U(t)$. Then we have

$$P\left( \max_{1 \leq k \leq n} |U(t_k)| > \eta \right) \leq \sum_{k=1}^{n} \sum_{l=1}^{r_n} P(\max_{|t| \leq C} |U_i(t)| > \eta)^{1/2}, \quad \eta > 0.$$  

Since $U_i(t)$ is a degenerate U-statistic whose kernel is bounded by $b_l = \sqrt{2n^{1/4}(\pi l + 2)} \leq 8n^{1/4}l$ and has second moment bounded by $2(\pi l)^2$, we derive from part (c) of Proposition 2.3 of Arcones and Giné (1993) that

$$\sup_{|t| \leq C} P(\max_{|t| \leq C} |U_i(t)| > \eta) \leq c_1 \exp \left( - \frac{c_2 \eta}{\sqrt{2\pi l + b_l^{2/3}\eta^{1/3}n^{-1/3}}} \right)$$

for universal constants $c_1$ and $c_2$. Using the above we obtain

$$P\left( \max_{1 \leq k \leq n} |U(t_k)| > K^{3}\log(n)r_n^{3/2}/n - 1 \right) \leq \sum_{k=1}^{n} \sum_{l=1}^{r_n} P((n - 1)|U_i(t_k)| > K^{3}\log(n)r_n)$$

$$\leq n r_n c_1 \exp \left( \frac{-c_2K^{3}\log(n)}{\sqrt{2\pi l + 4K(\log n)^{1/3}n^{-1/6}}} \right), \quad K > 0.$$
This shows that
\[ \max_{1 \leq k \leq n} |U(t_k)| = O_p(r_n^{3/2}n^{-1} \log n). \]

To deal with \( N_k \) we introduce the degenerate U-statistic
\[
\tilde{N}_k = \frac{1}{n(n-1)} \sum_{i \neq j} 1[D_{ij} \geq 0] \xi_k(i, j)
\]
with
\[
\xi_k(i, j) = 1[t_{k-1}D_{ij} < \varepsilon_i - \varepsilon_j \leq t_kD_{ij}] - F(\varepsilon_j + t_kD_{ij}) + F(\varepsilon_j + t_{k-1}D_{ij})
\]
\[
- F(\varepsilon_i - t_{k-1}D_{ij}) + F(\varepsilon_i - t_kD_{ij}) - F_2(t_kD_{ij}) - F(t_{k-1}D_{ij})
\]
and \( F_2 \) the distribution function of \( \varepsilon_1 - \varepsilon_2 \). It is easy to see that
\[
|N_k^+ - \tilde{N}_k^+| \leq 6CB_1n^{-3/2}\frac{1}{n(n-1)} \sum_{i \neq j} |X_i - X_j|.
\]

The kernel of the U-statistic \( \tilde{N}_k^+ \) is bounded by 8 and has second moment bounded by \( Dn^{-3/4} \) with \( D = 2B_1CE[|X_1 - X_2|] \). Thus, by part (c) of Proposition 2.3 in Arcones and Giné (1993), we obtain that the corresponding degenerate U-statistic \( \tilde{N}_k \) satisfies
\[
\sum_{k=1}^n P(|\tilde{N}_k| > K^3\log^{3/2} nn^{-1/2}) \leq nc_1 \exp\left(-\frac{c_2K^3\log^{3/2} n}{D^{1/2}n^{-1/4} + 2K\log^{1/3} n}\right).
\]

The above show that
\[ \max_{1 \leq k \leq n} N_k^+ = O_p(n^{-3/2} \log^{3/2} n). \]

Similarly one obtains
\[ \max_{1 \leq k \leq n} N_k^- = O_p(n^{-3/2} \log^{3/2} n). \]

Thus we obtain
\[ \sup_{|t| \leq C} |U(t)| = O_p(r_n^{3/2}n^{-1} \log n) \]
and conclude \( T_3 = O_p(r_n^{3/2}n^{-1/2} \log n) \). This concludes the proof of (4.4).
5. Auxiliary Results. Let $X$ and $Y$ be independent random variables. Let $(X_1, Y_1), \ldots, (X_m, Y_m)$ be independent copies of $(X, Y)$. For reals $t$, $x$ and $y$, set

$$N(t, x, y) = \sum_{i=1}^{m} (1[Y_j - tX_j \leq y - tx] - 1[Y_j \leq y]),$$

and

$$\tilde{N}(t, x, y) = N(t, x, y) - E[N(t, x, y)].$$

**Lemma 4.** Suppose $X$ has finite expectation and the distribution function $F$ of $Y$ is Lipschitz: $|F(y) - F(x)| \leq \Lambda |y - x|$ for all $x, y$ and some finite constant $\Lambda$. Then

$$P\left( \sup_{|t| \leq \delta} |\tilde{N}(t, x, y)| > 4\eta \right) \leq (4M + 2) \exp\left( \frac{-\eta^2}{2m\delta E[|X - x|] + 2\eta/3} \right)$$

for $\eta > 0$, $\delta > 0$, real $x$ and $y$ and every integer $M \geq m\Lambda E[|X - x|]/\eta$. In particular, for all positive $C$ and $K$,\n
$$P\left( \sup_{|t| \leq C/m^{1/2}} |\tilde{N}(t, x, y)| > 4Km^{1/4}(\log m)^{1/2} \right)$$

$$\leq \left( 6 + \frac{4m^{1/4}ACE[|X - x|]}{K(\log m)^{1/2}} \right) \exp\left( -\frac{K^2 \log(m)}{2(ACE[|X - x|] + K)} \right).$$

and

$$\sup_{|x| \leq n^{1/4}} P\left( \sup_{|t| \leq C/m^{1/2}} |\tilde{N}(t, x, y)| > 4Km^{3/8}(\log m)^{1/2} \right)$$

$$\leq \left( 6 + \frac{4m^{1/4}ACE[|X|] + m^{1/4}}{K(\log m)^{1/2}} \right) \exp\left( -\frac{K^2 \log(m)}{2(ACE(1 + E[|X|]m^{-1/4}) + K)} \right).$$

**Proof.** Fix $x$ and $y$. Abbreviate $N(t, x, y)$ by $N(t)$ and $\tilde{N}(t, x, y)$ by $\tilde{N}(t)$, set

$$N_+(t) = \sum_{i=1}^{m} (1[Y_j - t(X_j - x) \leq y] - 1[Y_j \leq y])1[X_j - x \geq 0]$$

$$N_-(t) = \sum_{i=1}^{m} (1[Y_j - t(X_j - x) \leq y] - 1[Y_j \leq y])1[X_j - x < 0]$$

and let $\tilde{N}_+(t) = N_+(t) - E[N_+(t)]$ and $\tilde{N}_-(t) = N_-(t) - E[N_-(t)]$. Since $F$ is Lipschitz, we obtain

$$|E[N_+(t_1)] - E[N_+(t_2)]| \leq m\Lambda|t_1 - t_2|E[|X - x|].$$
For \( s \leq t \leq u \), we have \( N_+(u) - E[N_+(s)] \leq N_+(t) - E[N_+(t)] \leq N_+(s) - E[N_+(u)] \) and thus
\[
\tilde{N}_+(u) - m\Lambda|t-u|E[|X-x|] \leq \tilde{N}_+(t) \leq \tilde{N}_+(s) + m\Lambda|t-s|E[|X-x|].
\]
It is now easy to see that
\[
\sup_{|t| \leq \delta} |\tilde{N}_+(t)| \leq \max_{k=-M,-...,M} |N_+(k\delta/M)| + m\Lambda\delta E[|X-x|]/M
\]
for every integer \( M \). From this we obtain the bound
\[
P(\sup_{|t| \leq \delta} |\tilde{N}_+(t)| \geq 2\eta) \leq \sum_{k=-M}^{M} P(|\tilde{N}_+(k\delta/M) > \eta) + P(m\Lambda\delta/M > \eta)
\]
The Bernstein inequality and the fact that the variance of \( (1[Y-t(X-x) \leq y] - 1[Y \leq y])1[X \leq x] \) is bounded by \( \Lambda|t|E[|X-x|] \) yield
\[
P(|N_+(k\delta/M)| > \eta) \leq 2\exp(-\frac{\eta^2}{2m\Lambda\delta E[|X-x|] + 2\eta/3}).
\]
Thus we have
\[
P(\sup_{|t| \leq \delta} |\tilde{N}_+(t)| > 2\eta) \leq 2(2M + 1)\exp(-\frac{\eta^2}{2m\Lambda\delta E[|X-x|] + 2\eta/3}).
\]
for \( M > m\lambda\delta E[|X-x|]/\eta \). Similarly, one verifies for such \( M \),
\[
P(\sup_{|t| \leq \delta} |\tilde{N}_-(t)| > 2\eta) \leq 2(2M + 1)\exp(-\frac{\eta^2}{2m\Lambda\delta E[|X-x|] + 2\eta/3}).
\]
Since \( \tilde{N}(t) = \tilde{N}_+(t) + N_-(t) \), we obtain the first result. The second result follows from the first one by taking \( \delta = Cm^{-1/2}, \eta = Km^{1/4}(\log m)^{1/2} \) and observing that \( (\log m)^{1/2}m^{-1/4} \leq 1 \).

**Proof of (4.25).** Let \( \xi_j(t) = \tilde{R}_{jt} - R_{jt} - \tilde{R}_j + R_j \) and \( m - 1 \). Then \( (n-1)\xi_n(t) \) equals \( \tilde{N}(n^{-1/2}t, X_n, \varepsilon_n) \) with \( Y_i = \varepsilon_i \). Thus,
\[
P(\max_{1 \leq j \leq n} \sup_{|t| \leq C} |\xi_j(t)| > 4K(n-1)^{-3/4}(\log(n-1))^{1/2})
\]
\[
\leq nP(\sup_{|t| \leq C} |\xi_n(t)| > 4Km^{-3/4}(\log m)^{1/2})
\]
\[
\leq nE[P(\sup_{|t| \leq Cm^{-1/2}} |\tilde{N}(t, X_n, \varepsilon_n)| > 4Km^{1/4}(\log m)^{1/2}|X_n, \varepsilon_n])
\]
\[
\leq nE\left(6\frac{4m^{1/4}B_1C\mu(X_n)}{K(\log m)^{1/2}}\right)\exp\left(-\frac{K^2\log(m)}{2(B_1C\mu(X_n) + K)}\right)
\]
for all finite \( C \) and \( K \) with \( \mu(X_n) = E[|X_1 - X_n||X_n] \). The desired (4.25) is now immediate. \( \square \)
References.


**Ursula U. Müller**  
Department of Statistics  
Texas A&M University  
College Station, TX 77843-3143, USA  
E-mail: uschi@stat.tamu.edu  
URL: http://www.stat.tamu.edu/~uschi  

**Hanxiang Peng**  
Indiana University Purdue University  
At Indianapolis  
Department of Mathematical Sciences  
Indianapolis, IN 46202-3267  
USA  
E-mail: hpeng@math.iupui.edu  
URL: http://www.math.iupui.edu/~hpeng  

**Anton Schick**  
Binghamton University  
Department of Mathematical Sciences  
Binghamton, NY 13902-6000  
USA  
E-mail: anton@math.binghamton.edu  
URL: http://www.math.binghamton.edu/anton