stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SW 78/81 DECEMBER
(DEPARTMENT OF MATHEMATICAL STATISTICS)

R.J.M.M. DOES

AN EDGEWORTH EXPANSION FOR SIMPLE LINEAR RANK STATISTICS UNDER THE NULL-HYPOTHESIS

Preprint

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, 413 Kruislaan, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1980 Mathematics subject classification: Primary 62G10, 62G20
Secondary 60F05

An Edgeworth expansion for simple linear rank statistics under the null-hypothesis*)

by

Ronald J.M.M. Does**)

ABSTRACT

An Edgeworth expansion with remainder $o(N^{-1})$ is established for simple linear rank statistics under the null-hypothesis. The theorem is proved for a wide class of scores generating functions which includes the normal quantile function.

KEY WORDS & PHRASES: *simple linear rank statistics, Edgeworth expansions, distributionfree tests*

---

*) This report will be submitted for publication elsewhere.

**) Mathematical Centre, Amsterdam and University of Limburg, Maastricht

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed random variables with a common continuous distribution function F. If $X_{1:N} < X_{2:N} < \ldots < X_{N:N}$ denotes the sequence $X_1, X_2, \ldots, X_N$ arranged in increasing order, then the rank R_{jN} of X_j is defined by $X_j = X_{R_{jN}:N}$ and the antirank D_{jN} is defined by $X_{D_{jN}} = X_{j:N}$, $j = 1,2,\ldots,N$. We consider the simple linear rank statistic

$$(1.1) \qquad T_N = \sum_{j=1}^{N} c_{jN} J\left(\frac{R_{jN}}{N+1}\right) = \sum_{j=1}^{N} c_{D_{jN}N} J\left(\frac{j}{N+1}\right),$$

where $C_{1N}, C_{2N}, \ldots, C_{NN}$, $N = 1,2,\ldots$, is a triangular array of regression constants and J is a scores generating function defined on (0,1). The two-sample linear rank statistic is obviously obtained as a special case by setting $c_{jN} = 0$ for $j = 1,2,\ldots,n$, $c_{jN} = 1$ for $j = n+1,\ldots,N$. If $c_{jN} = j$ for $j = 1,2,\ldots,N$ and $J(t) = t$ for $t \in (0,1)$ then the statistic T_N is distributed as Spearman's rank correlation coefficient ρ under the null-hypothesis of independence.

The statistic T_N may be used for testing the null-hypothesis that all observations are independent and identically distributed against classes of alternatives indicated by the choice of regression constants and scores generating function. Both under the hypothesis and under contiguous and fixed alternatives it was shown that T_N is asymptotically normally distributed under very general conditions (cf. HÁJEK & ŠIDÁK (1967), Chapters V and VI, HÁJEK (1968) and DUPAČ & HÁJEK (1969)). More recently a number of authors have studied the rate of convergence in these limit theorems. Berry-Esseen type bounds of order $\mathcal{O}(N^{-\frac{1}{2}})$ for simple linear rank statistics were established by HUŠKOVÁ (1977,1979), HO & CHEN (1978) and DOES (1981). The purpose of this paper is to establish an Edgeworth expansion for simple linear rank statistics under the hypothesis with remainder $o(N^{-1})$ for a wide class of scores generating functions including the normal quantile function. We note that for the special case of the two-sample linear rank statistic, asymptotic expansions both under the hypothesis and under contiguous alternatives were obtained in BICKEL & VAN ZWET (1978). Asymptotic expansions for the simple linear rank statistics under contiguous alternatives will be discussed in the author's forthcoming Ph.D. thesis.

In Section 2 we formulate our theorem. Section 3 contains a number of preliminaries. The proof of the theorem is contained in Section 4. Finally in the last section we compare our results with those in BICKEL & VAN ZWET (1978) for the two-sample linear rank statistic. In the sequel we suppress the index N whenever it is possible.

2. AN EDGEWORTH EXPANSION

Throughout this paper we make the following assumptions.

ASSUMPTION (A). *The regression constants* $c_{1N}, c_{2N}, \ldots, c_{NN}$ *satisfy*

$$\sum_{j=1}^{N} c_{jN} = 0, \quad \sum_{j=1}^{N} c_{jN}^2 = 1, \quad \max_{1\le j\le N} |c_{jN}| = O(N^{-\frac{1}{2}}).$$

This assumption implies that $ET_N = 0$.

ASSUMPTION (B). *The scores generating function* J *is three times differentiable on* (0,1) *and*

$$\limsup_{t\to 0,1} t(1-t)\left|\frac{J''(t)}{J'(t)}\right| < 2\,; \tag{2.1}$$

there exist positive numbers $\Gamma > 0$ *and* $\alpha < 3 + 1/14$ *such that its third derivative* J''' *satisfies*

$$|J'''(t)| \le \Gamma\{t(1-t)\}^{-\alpha} \quad \textit{for } t \in (0,1). \tag{2.2}$$

Furthermore

$$\int_0^1 J(t)dt = 0, \quad \int_0^1 J^2(t)dt = 1. \tag{2.3}$$

We note that (2.1) ensures that the function J does not oscillate too wildly near 0 and 1 (see also Appendix 2 of ALBERS, BICKEL & VAN ZWET (1976)). Condition (2.3) can be assumed without loss of generality.

Taking

(2.4) $$\bar{J} = \frac{1}{N} \sum_{j=1}^{N} J(\tfrac{j}{N+1}),$$

we know that the variance σ_N^2 of T_N (cf. (1.1)) is given by

(2.5) $$\sigma_N^2 = \sigma^2(T_N) = \frac{1}{N-1} \sum_{j=1}^{N} \left(J(\tfrac{j}{N+1}) - \bar{J}\right)^2$$

(see e.g. Theorem II 3.1.c of HÁJEK & ŠIDÁK (1967)). Define for each $N \geq 2$

(2.6) $$T_N^* = \sigma_N^{-1} T_N$$

and

(2.7) $$F_N^*(x) = P(T_N^* \leq x) \quad \text{for } -\infty < x < \infty.$$

Furthermore define for each $N \geq 2$ and real x, the function $\tilde{F}_N$ by

(2.8) $$\tilde{F}_N(x) = \Phi(x) - \phi(x)\left\{\frac{\kappa_{3N}}{6}(x^2-1) + \frac{\kappa_{4N}}{24}(x^3-3x) + \frac{\kappa_{3N}^2}{72}(x^5-10x^3+15x)\right\},$$

where Φ denotes the standard normal distribution function, ϕ its density and where the quantities κ_{3N} and κ_{4N} are given by

(2.9) $$\kappa_{3N} = \sum_{j=1}^{N} c_{jN}^3 \left\{\int_0^1 J^3(t)dt\right\}$$

and

(2.10) $$\kappa_{4N} = \sum_{j=1}^{N} c_{jN}^4 \left\{\int_0^1 J^4(t)dt - 3\right\} - \frac{3}{N}\left\{\int_0^1 J^4(t)dt - 1\right\}.$$

Our theorem reads as follows.

<u>THEOREM 2.1</u>. *If the Assumptions* (A) *and* (B) *are satisfied, then as* $N \to \infty$

(2.11) $$\sup_{x \in \mathbb{R}} |F_N^*(x) - \tilde{F}_N(x)| = o(N^{-1}).$$

We note that κ_{3N} and κ_{4N} (cf. (2.9) and (2.10)) are asymptotic expressions for the third and fourth cumulants of T_N^* where terms of order $O(N^{-1})$ have been neglected. Hence $\tilde{F}_N$ may be said to constitute a genuine Edgeworth expansion for F_N^*. We should also point out that Theorem 2.1 allows scores generating functions tending to infinity in the neighbourhood of 0 and 1 at the rate of $\{t(1-t)\}^{-1/14+\varepsilon}$ for some $\varepsilon > 0$. It is clear that this includes the normal quantile function. Whenever we shall suppose in the remainder of this paper that (2.2) in Assumption (B) is satisfied, we shall tacitly and without loss of generality assume that $\alpha \in (3, 3+1/14)$ and define $\delta = 3+1/14-\alpha$. Hence, from now on we replace (2.2) in Assumption (B) by

(2.12) $$|J'''(t)| \le \Gamma\{t(1-t)\}^{-(3+1/14)+\delta} \quad \text{for } t \in (0,1),$$

where

(2.13) $$0 < \delta < 1/14.$$

To conclude this section we define $U_1, U_2, \ldots, U_N$ to be independent and uniformly distributed random variables on (0,1) and $U_{1:N} < U_{2:N} < \ldots < U_{N:N}$ the corresponding uniform order statistics.

3. PRELIMINARY LEMMAS

The aim in this section is threefold. In the first place we approximate $(N-1)\sigma_N^2$ (cf. (2.5)) by a integral. For this we shall draw heavily on the results in Appendix 2 of ALBERS, BICKEL & VAN ZWET (1976). Secondly we study the behaviour of the characteristic function of T_N^* (cf. (2.6)) for large values of the argument. To this end we shall provide a lemma which is a special case of Theorem 2.1 of VAN ZWET (1980). Finally we prove two technical lemmas, the purpose of which will become clear in Section 4.

LEMMA 3.1. *If* J *satisfies Assumption* (B), *then*

(3.1) $$\sum_{j=1}^{N} \left(J(\tfrac{j}{N+1}) - \bar{J} \right)^2 = N + O(N^{1/7-2\delta}).$$

PROOF. Take δ as in (2.12) and (2.13), let h be a function on (0,1) with $h'(t) \equiv \Gamma\{t(1-t)\}^{-15/14+\delta}$ and write $\lambda_j = j/(N+1)$. Since

$$\limsup_{t\to 0,1} t(1-t)\left|\frac{h''(t)}{h'(t)}\right| < \frac{3}{2},$$

Lemma A.2.3 of ALBERS, BICKEL & VAN ZWET (1976) yields

$$E\{h(U_{j:N}) - h(\lambda_j)\}^2 = \mathcal{O}\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-8/7+2\delta}}{N}\right)$$

uniformly in j. Because $|J'(t)| \le h'(t)$ we have $|J(s)-J(t)| \le |h(s)-h(t)|$ for every $s,t \in (0,1)$ and hence

$$\text{(3.2)} \qquad E\{J(U_{j:N})-J(\lambda_j)\}^2 = \mathcal{O}\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-8/7+2\delta}}{N}\right)$$

uniformly in j. As J satisfies (2.1), we also have, in view of (A.2.11) in ALBERS, BICKEL & VAN ZWET (1976),

$$\text{(3.3)} \qquad |EJ(U_{j:N}) - J(\lambda_j)| = \mathcal{O}\left(\frac{\lambda_j(1-\lambda_j) + |J'(\lambda_j)|}{N}\right) = \mathcal{O}\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-15/14+\delta}}{N}\right)$$

uniformly in j. Since $\int J = 0$ and $\delta \in (0,1/14)$ (cf. (2.3) and (2.13)), it follows that

$$\text{(3.4)} \qquad \left|\frac{1}{N}\sum_{j=1}^{N} J(\lambda_j)\right| = \left|\frac{1}{N}\sum_{j=1}^{N} \{J(\lambda_j) - EJ(U_{j:N})\}\right| = \mathcal{O}(N^{-13/14-\delta}).$$

Furthermore, in view of (3.2) and (3.3) and since $\int J^2 = 1$ and $\delta \in (0,1/14)$

$$\left|\sum_{j=1}^{N} J^2(\lambda_j) - N\right| = \left|\sum_{j=1}^{N} \{J^2(\lambda_j) - EJ^2(U_{j:N})\}\right| \le$$

$$\le \sum_{j=1}^{N} E\{J(U_{j:N}) - J(\lambda_j)\}^2 + 2\sum_{j=1}^{N} |J(\lambda_j)|\,|EJ(U_{j:N}) - J(\lambda_j)| =$$

$$= \mathcal{O}(N^{1/7-2\delta}),$$

which proves the lemma. □

We now consider the behaviour of the characteristic function of T_N^* for large values of the argument. Let

$$(3.5) \qquad \psi_N(t) = Ee^{itT_N^*}.$$

LEMMA 3.2. *Suppose that the assumptions of Theorem* 2.1 *are satisfied. Then there exist positive numbers* B, β *and* γ *such that*

$$(3. \qquad |\psi_N(t)| \le BN^{-\beta \log N}$$

for $\log N \le |t| \le \gamma N^{3/2}$ *and* $N = 2,3,\ldots$.

PROOF. The present lemma is a special case of Theorem 2.1 of VAN ZWET (1980). Since we are concerned with independent and identically distributed random variables $X_1, X_2, \ldots, X_N$ - which we may assume to be uniformly distributed without loss of generality - Condition (2.7) of this theorem is clearly satisfied. Moreover, the assumptions of our theorem guarantee that there exists a positive fraction of the scores which are at a distance of at least $N^{-3/2} \log N$ apart from each other, so Assumption (2.6) of Theorem 2.1 of VAN ZWET (1980) is also fulfilled. Finally, it follows from Section 3 in VAN ZWET (1980) that the existence of positive numbers c and C such that

$$(3.7) \qquad \sum_{j=1}^{N} c_j^2 \ge c, \qquad \sum_{j=1}^{N} c_j^4 \le CN^{-1},$$

$$(3.8) \qquad \sum_{j=1}^{N} \left(J\left(\frac{j}{N+1}\right) - \bar{J}\right)^2 \ge cN, \qquad \sum_{j=1}^{N} \left(J\left(\frac{j}{N+1}\right) - \bar{J}\right)^4 \le CN$$

suffices to prove the present lemma. Assumption (A) guarantees the validity of (3.7) and (3.8) is an immediate consequence of Assumption (B) and the continuity of J (cf. also (3.1)). □

Let [x] denote the largest integer not exceeding x. Define $m = [N^{8/15}]$ and $I = \{1,2,\ldots,m,N-m+1,\ldots,N-1,N\}$.

LEMMA 3.3. *If Assumtions* (A) *and* (B) *are satisfied, then*

$$(3.9)\qquad E\Big|\sum_{j\in I} c_{D_j}\, J(\tfrac{j}{N+1})\Big|^5 = O(N^{-1-7\delta/3}),$$

$$(3.10)\qquad \Big\{\frac{1}{N-2m}\sum_{j=m+1}^{N-m} J(\tfrac{j}{N+1})\Big\}^2 E\Big(\sum_{j=m+1}^{N-m} c_{D_j}\Big)^2 = O(N^{-4/3-14\delta/15}).$$

PROOF. According to Assumption (A) $\sum c_j = 0$, $\sum c_j^2 = 1$ and

$$\sum_{j=1}^{N} |c_j|^k \le \max_{1\le j\le N} |c_j|^{k-2} \sum_{j=1}^{N} c_j^2 = O(N^{1-k/2}),$$

for $k > 2$. It follows that for distinct $i,j,h,g,k,\ell \in I$

$$Ec_{D_i}^6 = O(N^{-3}),\quad Ec_{D_i}^5 c_{D_j} = O(N^{-4}),\quad Ec_{D_i}^4 c_{D_j}^2 = O(N^{-3}),$$

$$Ec_{D_i}^3 c_{D_j}^3 = O(N^{-3}),\quad Ec_{D_i}^4 c_{D_j} c_{D_h} = O(N^{-4}),\quad Ec_{D_i}^3 c_{D_j}^2 c_{D_h} = O(N^{-4}),$$

$$Ec_{D_i}^2 c_{D_j}^2 c_{D_h}^2 = O(N^{-3}),\quad Ec_{D_i}^3 c_{D_j} c_{D_h} c_{D_g} = O(N^{-5}),$$

$$Ec_{D_i}^2 c_{D_j}^2 c_{D_h} c_{D_g} = O(N^{-4}),\quad Ec_{D_i}^2 c_{D_j} c_{D_h} c_{D_g} c_{D_k} = O(N^{-5}),$$

$$Ec_{D_i} c_{D_j} c_{D_h} c_{D_g} c_{D_k} c_{D_\ell} = O(N^{-6}).$$

Furthermore, Hölder's inequality yields

$$(3.11)\qquad E\Big|\sum_{j\in I} c_{D_j}\, J(\tfrac{j}{N+1})\Big|^5 \le \Big\{E\Big(\sum_{j\in I} c_{D_j}\, J(\tfrac{j}{N+1})\Big)^6\Big\}^{5/6}.$$

In view of (2.12) and (2.13) we have for $k = 1,2,\dots,6$

$$(3.12)\qquad \frac{1}{N}\sum_{j\in I} |J(\tfrac{j}{N+1})|^k = O\Big(\int_0^{\frac{m}{N+1}} \{t(1-t)\}^{-k/14+k\delta}\, dt\Big) = O\Big(\Big\{\frac{m}{N+1}\Big\}^{1-k/14+k\delta}\Big).$$

Direct computation of the right-hand side of (3.11) produces (3.9). Since $\sum c_j = 0$, $Ec_{D_j}^2 = N^{-1}$ and $Ec_{D_i} c_{D_j} = -\{N(N-1)\}^{-1}$ for $i \ne j$, we have

$$E\Big(\sum_{j=m+1}^{N-m} c_{D_j}\Big)^2 = E\Big(\sum_{j\in I} c_{D_j}\Big)^2 = O(\tfrac{m}{N}).$$

Similarly we find that, in view of (3.4) and (3.12),

$$\left|\frac{1}{N}\sum_{j=m+1}^{N-m} J(\tfrac{j}{N+1})\right| = \left|\frac{1}{N}\sum_{j\in I} J(\tfrac{j}{N+1})\right| + O(N^{-13/14-\delta}) =$$

(3.13)

$$= O\Big(\Big\{\frac{m}{N}\Big\}^{13/14+\delta}\Big)$$

and the lemma follows. □

To conclude this section we prove

LEMMA 3.4. *If Assumption* (A) *is satisfied, then for any* $\gamma < 1$ *and* $N \to \infty$

(3.14) $$P\Big(\sum_{j\in I} c^2_{D_j} \geq 1-\gamma\Big) = O(N^{-22/15}).$$

PROOF. Since $E(\sum_{j\in I} c^2_{D_j}) = 2mN^{-1}$ and

$$E\Big(\sum_{j\in I} c^2_{D_j} - \frac{2m}{N}\Big)^2 = \frac{2m(N-2m)}{N(N-1)}\Big(\sum_{j=1}^{N} c_j^4 - \frac{1}{N}\Big),$$

the Bienaymé-Chebyshev inequality ensures that for every $\gamma < 1$

$$P\Big(\Big|\sum_{j\in I} c^2_{D_j} - \frac{2m}{N}\Big| \geq \frac{1-\gamma}{2}\Big) \leq \frac{4}{(1-\gamma)^2} E\Big(\sum_{j\in I} c^2_{D_j} - \frac{2m}{N}\Big)^2 = O(N^{-22/15}).$$

The lemma follows because $mN^{-1} \to 0$ as $N \to \infty$. □

4. PROOF OF THE THEOREM

To prove Theorem 2.1 we start with an application of Esseen's smoothing lemma (see e.g. FELLER (1971), p.538), which implies that for all $\gamma > 0$

(4.1) $$\sup_{x\in\mathbb{R}} |F_N^*(x)-\tilde{F}_N(x)| \leq \frac{1}{\pi}\int_{-\gamma N^{3/2}}^{\gamma N^{3/2}} \frac{|\psi_N(t)-\lambda_N(t)|}{|t|}\,dt + O(N^{-3/2}),$$

where ψ_N denotes the characteristic function of T_N^* (cf. (3.5)) and λ_N denotes the Fourier-Stieltjes transform of $\tilde{F}_N$, i.e.

$$\lambda_N(t) = \int_{-\infty}^{\infty} e^{itx}\, d\tilde{F}_N(x) = e^{-\frac{1}{2}t^2}\left\{1 - \frac{\kappa_{3N}}{6} it^3 + \frac{\kappa_{4N}}{24} t^4 - \frac{\kappa_{3N}^2}{72} t^6\right\}. \tag{4.2}$$

The derivative of λ_N is uniformly bounded and also

$$\left|\frac{d\psi_N(t)}{dt}\right| \le E|T_N^*| \le 1.$$

Because $\psi_N(0) = \lambda_N(0) = 1$, we have

$$\int_{|t|\le N^{-3/2}} \frac{|\psi_N(t)-\lambda_N(t)|}{|t|}\, dt = O(N^{-3/2}). \tag{4.3}$$

Similarly, Lemma 3.2 and (4.2) ensure that

$$\int_{\log N \le |t| \le \gamma N^{3/2}} \frac{|\psi_N(t)-\lambda_N(t)|}{|t|}\, dt = O(N^{-3/2}). \tag{4.4}$$

From (4.1), (4.3) and (4.4) it follows that, in order to prove Theorem 2.1, it suffices to show that

$$\int_{t\in A} \frac{|\psi_N(t)-\lambda_N(t)|}{|t|}\, dt = o(N^{-1}), \tag{4.5}$$

where $A = \{t : N^{-3/2} \le |t| \le \log N\}$.

To solve this problem we use a conditioning argument. We take δ as in (2.12) and (2.13) and define $m = [N^{8/15}]$ and $I = \{1,2,\ldots,m,N-m+1,\ldots,N-1,N\}$ as in Section 3. Let $\Omega = \{D_j : j\in I\}$ be the set of antiranks D_j with indices in I and let $\omega = \{d_j : j\in I\}$ be a possible realization of Ω. Finally define

$$Z_N = \sum_{j\in I} c_{D_j} J\left(\frac{j}{N+1}\right). \tag{4.6}$$

Because $(T_N - Z_N)$ and Z_N are conditionally independent given Ω, we have

$$\begin{aligned}\psi_N(t) &= Ee^{itT_N^*} = E\left[E\left(e^{it\sigma_N^{-1}(T_N-Z_N)}\middle|\Omega\right)E\left(e^{it\sigma_N^{-1}Z_N}\middle|\Omega\right)\right] = \\ &= E\left[E\left(e^{it\sigma_N^{-1}\{(T_N-Z_N)-E(T_N-Z_N|\Omega)\}}\middle|\Omega\right)e^{it\sigma_N^{-1}E(T_N-Z_N|\Omega)}E\left(e^{it\sigma_N^{-1}Z_N}\middle|\Omega\right)\right].\end{aligned} \tag{4.7}$$

We note that conditionally on $\Omega = \omega$, $T_N - Z_N = \sum_{j=m+1}^{N-m} c_{D_j} J(j/(N+1))$ is distributed as a simple linear rank statistic for sample size N-2m based on a set of regression constants $\{c_1,c_2,\ldots,c_N\}\backslash\{c_{d_j}:j\in I\}$ and having a scores generating function

$$(4.8) \qquad J_N(t) = J\left(\frac{m+(N-2m+1)t}{N+1}\right) \quad \text{for } t \in (0,1).$$

We write this simple linear rank statistic as

$$(4.9) \qquad T_{\omega N} = \sum_{j=1}^{M} b_j J_N\left(\frac{Q_j}{M+1}\right),$$

where $M = N-2m$, $\{b_1,b_2,\ldots,b_M\} = \{c_1,c_2,\ldots,c_N\}\backslash\{c_{d_j}:j\in I\}$, $Q_1,Q_2,\ldots,Q_M$ are the ranks of $V_1,V_2,\ldots,V_M$, which are independent and uniformly distributed random variables on (0,1).

Define for $j = 1,2,\ldots,M$

$$(4.10) \qquad \hat{V}_j = E\left(\frac{Q_j}{M+1}\Big| V_j\right) = \frac{1}{M+1} + \frac{M-1}{M+1} V_j$$

and let $S_{\omega N}$ be a three-term Taylor expansion of $T_{\omega N}$, viz.

$$S_{\omega N} = \sum_{j=1}^{M} b_j\left\{J_N(\hat{V}_j) + J_N'(\hat{V}_j)\left(\frac{Q_j}{M+1} - \hat{V}_j\right) + \tfrac{1}{2}J_N''(\hat{V}_j)\left(\frac{Q_j}{M+1} - \hat{V}_j\right)^2\right\}.$$

We shall approximate $(T_{\omega N}-ET_{\omega N})$ by $(S_{\omega N}-ES_{\omega N})$ and for this we need

<u>LEMMA 4.1</u>. *Under the Assumptions* (A) *and* (B) *we have, uniformly in* ω,

$$(4.12) \qquad \sigma^2(T_{\omega N}-S_{\omega N}) = \left(1 + \left(\sum_{j\in I} c_{d_j}\right)^2\right)O(N^{-2-14\delta/15}).$$

<u>PROOF</u>. Let, for $j = 1,2,\ldots,M$,

$$Y_j = J_N\left(\frac{Q_j}{M+1}\right) - \left\{J_N(\hat{V}_j) + J_N'(\hat{V}_j)\left(\frac{Q_j}{M+1} - \hat{V}_j\right) + \tfrac{1}{2}J_N''(\hat{V}_j)\left(\frac{Q_j}{M+1} - \hat{V}_j\right)^2\right\}.$$

Because $\sum_{j=1}^{M} b_j^2 \le 1$ and

$$\left|\sum\sum_{(j,k)\neq} b_j b_k\right| = \left|\left(\sum_{j=1}^{M} b_j\right)^2 - \sum_{j=1}^{M} b_j^2\right| = \left|\left(\sum_{j\in I} c_{d_j}\right)^2 - \sum_{j=1}^{M} b_j^2\right| \le$$

$$\le 1 + \left(\sum_{j\in I} c_{d_j}\right)^2,$$

the Cauchy-Schwarz inequality yields

$$\sigma^2(T_{\omega N}-S_{\omega N}) \le E(T_{\omega N}-S_{\omega N})^2 = E\left(\sum_{j=1}^{M} b_j Y_j\right)^2 = \sum_{j=1}^{M} b_j^2 EY_1^2 +$$

$$+ \sum\sum_{(j,k)\neq} b_j b_k EY_1Y_2 \le EY_1^2 + \left(1 + \left(\sum_{j\in I} c_{d_j}\right)^2\right) E|Y_1Y_2| \le$$

$$\le \left(2 + \left(\sum_{j\in I} c_{d_j}\right)^2\right) EY_1^2.$$

Here $\sum\sum_{(j,k)\neq}$ denotes summation over all non-negative distinct integers j,k satisfying $1 \le j,k \le M$. Define $r(t) = \{t(1-t)\}^{-1}$. By Taylor's theorem, (4.8), (2.12) and the convexity of the function $r(t)$ we see that

$$EY_1^2 \le \frac{1}{36} E\left(\frac{Q_1}{M+1} - \hat{V}_1\right)^6 \sup_{0\le\eta\le 1} \left\{J_N'''\left(\eta \frac{Q_1}{M+1} + (1-\eta)\hat{V}_1\right)\right\}^2 \le$$

$$\le \frac{\Gamma^2}{36} E\left(\frac{Q_1}{M+1} - \hat{V}_1\right)^6 \left\{r^{6+1/7-2\delta}\left(\frac{m+Q_1}{N+1}\right) + r^{6+1/7-2\delta}\left(\frac{m+(M+1)\hat{V}_1}{N+1}\right)\right\}.$$

The independence of the vector of ranks $(Q_1,Q_2,\ldots,Q_M)$ and the vector of order statistics $(V_{1:M},V_{2:M},\ldots,V_{M:M})$ and Lemma A.2.3 of ALBERS, BICKEL & VAN ZWET (1976) imply

$$E\left(\frac{Q_1}{M+1} - \hat{V}_1\right)^6 r^{6+1/7-2\delta}\left(\frac{m+Q_1}{M+1}\right) = \left(\frac{M-1}{M+1}\right)^6 E\left(\frac{Q_1-1}{M-1} - V_1\right)^6 r^{6+1/7-2\delta}\left(\frac{m+Q_1}{N+1}\right) \le$$

$$\le \frac{1}{M} \sum_{j=1}^{M} E\left(V_{j:M} - \frac{j-1}{M-1}\right)^6 r^{6+1/7-2\delta}\left(\frac{m+j}{N+1}\right) =$$

$$= \mathcal{O}\left(\frac{1}{M^4} \sum_{j=1}^{M} r^{-3}\left(\frac{j}{M+1}\right) r^{6+1/7-2\delta}\left(\frac{m+j}{N+1}\right)\right) = \mathcal{O}(N^{-2-14\delta/15}).$$

Furthermore, the conditional distribution of Q_1-1 given V_1 is binomial with parameters M-1 and V_1 and by application of a recursion formula for the central moments of this distribution (cf. JOHNSON & KOTZ (1969, p.52) we find

$$E(\{Q_1 - E(Q_1|V_1)\}^6|V_1) = \mathcal{O}(\{MV_1(1-V_1)\}^3 + MV_1(1-V_1)).$$

Hence,

$$\text{(4.14)}\quad \begin{aligned} &E\left(\frac{Q_1}{M+1} - \hat{V}_1\right)^6 r^{6+1/7-2\delta}\left(\frac{m+(M+1)\hat{V}_1}{N+1}\right) = \mathcal{O}\left(E\left(\left\{\frac{V_1(1-V_1)}{M}\right\}^3 + \frac{V_1(1-V_1)}{M}\right)\cdot\right.\\ &\left.\cdot\, r^{6+1/7-2\delta}\left(\frac{m+(M+1)\hat{V}_1}{N+1}\right)\right) = \mathcal{O}(N^{-2-14\delta/15}).\end{aligned}$$

Combining (4.13) and (4.14) we find that $EY_1^2 = \mathcal{O}(N^{-2-14\delta/15})$. This proves the lemma. □

It follows from Lemma 4.1, (2.5) and (3.8) that

$$\text{(4.15)}\quad \begin{aligned} &\left|Ee^{it\sigma_N^{-1}(T_{\omega N}-ET_{\omega N})} - Ee^{it\sigma_N^{-1}(S_{\omega N}-ES_{\omega N})}\right| \leq \\ &\leq |t|\,\sigma_N^{-1}\, E|T_{\omega N}-ET_{\omega N}-S_{\omega N}+ES_{\omega N}| = \mathcal{O}\left(|t|N^{-1-7\delta/15}\left\{1+\left(\sum_{j\in I} c_{d_j}\right)^2\right\}^{\frac{1}{2}}\right),\end{aligned}$$

uniformly in t and ω.

Our next task is to evaluate $E\exp\{it\sigma_N^{-1}(S_{\omega N}-ES_{\omega N})\}$. The technique for doing this resembles that in HELMERS (1980). Let χ be the indicator function of $(0,\infty)$ and define

$$\text{(4.16)}\quad \begin{aligned} S_1 &= \sum_{j=1}^{M} b_j(J_N(\hat{V}_j) - EJ_N(\hat{V}_j)) = \sum_{j=1}^{M} b_j\tilde{J}_N(\hat{V}_j),\\ S_2 &= \frac{1}{M+1}\sum\sum_{(j,k)\neq} b_j J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j),\\ S_3 &= \frac{1}{2(M+1)^2}\sum\sum_{(j,k)\neq} b_j\left\{J_N''(\hat{V}_j)(\chi(V_j-V_k)-V_j)^2 - EJ_N''(\hat{V}_j)(\chi(V_j-V_k)-V_j)^2\right\},\\ S_4 &= \frac{1}{2(M+1)^2}\sum\sum\sum_{(j,k,\ell)\neq} b_j J_N''(\hat{V}_j)(\chi(V_j-V_k)-V_j)(\chi(V_j-V_\ell)-V_j).\end{aligned}$$

It is easy to see that $S_{\omega N} - ES_{\omega N} = \sum_{\nu=1}^{4} S_\nu$ and $ES_\nu = 0$ for $\nu = 1,\ldots,4$. First of all we compute a number of moments.

LEMMA 4.2. *Under the Assumptions* (A) *and* (B) *we have, uniformly in* ω,

$$E|S_2|^3 = O(N^{-13/10-7\delta/5}), \quad ES_3^2 = O(N^{-22/15-14\delta/15}),$$
$$ES_4^2 = O(N^{-7/5-14\delta/15}). \tag{4.17}$$

PROOF. By applying Hölder's inequality we obtain $E|S_2|^3 \le \{ES_2^4\}^{3/4}$. Let, for distinct j and k, $h(V_j,V_k) = J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)$. Define $h(x,x) = 0$ for all $0 < x < 1$. Direct computation of ES_2^4 shows that

$$ES_2^4 = \frac{1}{(M+1)^4} E\left\{\sum_{j=1}^{M} b_j \sum_{k=1}^{M} h(V_j,V_k)\right\}^4 =$$

$$= \frac{1}{(M+1)^4}\Bigg[\sum_{j=1}^{M} b_j^4 \left\{\sum_{r=1}^{M}\sum_{s=1}^{M}\sum_{t=1}^{M}\sum_{u=1}^{M} Eh(V_1,V_r)h(V_1,V_s)h(V_1,V_t)h(V_1,V_u)\right\} +$$
$$+ 4 \sum\sum_{(j,k)\neq} b_j^3 b_k \left\{\sum_{r=1}^{M}\sum_{s=1}^{M}\sum_{t=1}^{M}\sum_{u=1}^{M} Eh(V_1,V_r)h(V_1,V_s)h(V_1,V_t)h(V_2,V_u)\right\} +$$
$$+ 3 \sum\sum_{(j,k)\neq} b_j^2 b_k^2 \left\{\sum_{r=1}^{M}\sum_{s=1}^{M}\sum_{t=1}^{M}\sum_{u=1}^{M} Eh(V_1,V_r)h(V_1,V_s)h(V_2,V_t)h(V_2,V_u)\right\} + \tag{4.18}$$
$$+ 6 \sum\sum\sum_{(j,k,\ell)\neq} b_j^2 b_k b_\ell \left\{\sum_{r=1}^{M}\sum_{s=1}^{M}\sum_{t=1}^{M}\sum_{u=1}^{M} Eh(V_1,V_r)h(V_1,V_s)h(V_2,V_t)h(V_3,V_u)\right\} +$$
$$+ \sum\sum\sum\sum_{(j,k,\ell,n)\neq} b_j b_k b_\ell b_n \left\{\sum_{r=1}^{M}\sum_{s=1}^{M}\sum_{t=1}^{M}\sum_{u=1}^{M} Eh(V_1,V_r)h(V_2,V_s)h(V_3,V_t)h(V_4,V_u)\right\}\Bigg].$$

To bound the right-hand side of (4.18) we note that an expectation in (4.18) equals zero if at least one of the indices (r,s,t,u) occurs only once. With the aid of the Cauchy-Schwarz inequality the non-zero expectations may be bounded by either $Eh^4(V_1,V_2)$ or $\{Eh^4(V_1,V_2)\}^{\frac{1}{2}}Eh^2(V_1,V_2)$ or $\{Eh^2(V_1,V_2)\}^2$ and we obtain

$$ES_2^4 = O\Big(\Big\{\sum_{j=1}^{M} b_j^4\Big\} N^{-2}Eh^4(V_1,V_2) +$$

$$+ \Big\{\sum\sum_{(j,k)\neq} b_j^3 b_k\Big\}\Big\{N^{-2}\{Eh^4(V_1,V_2)\}^{\frac{1}{2}}Eh^2(V_1,V_2) + N^{-3}Eh^4(V_1,V_2)\Big\} +$$

$$\text{(4.19)} \qquad + \Big\{\sum\sum_{(j,k)\neq} b_j^2 b_k^2 + \sum\sum\sum_{(j,k,\ell)\neq} b_j^2 b_k b_\ell + \sum\sum\sum\sum_{(j,k,\ell,n)\neq} b_j b_k b_\ell b_n\Big\} \cdot$$

$$\cdot \{N^{-2}\{Eh^2(V_1,V_2)\}^2 + N^{-3}Eh^4(V_1,V_2)\}\Big).$$

In view of (4.8) and Assumption (B) we have, for $1 \le k \le 4$,

$$E|h^k(V_1,V_2)| = E|J_N'(\hat{V}_1)|^k|\chi(V_1-V_2)-V_1|^k \le$$

$$\le E|J_N'(\hat{V}_1)|^k\, V_1(1-V_1) \le E\Big|J'\Big(\frac{m+1+(M-1)V_1}{N+1}\Big)\Big|^k\, V_1(1-V_1) =$$

$$\text{(4.20)} \qquad = O\Big(\int_{\frac{m+1}{N+1}}^{1-\frac{m+1}{N+1}} \{t(1-t)\}^{-15k/14+k\delta}\Big\{\frac{(N+1)t-(m+1)}{M-1}\Big\}\Big\{\frac{(M+m)-(N+1)t}{M-1}\Big\}dt\Big) =$$

$$= O\Big(\int_{\frac{m}{N}}^{1-\frac{m}{N}} \{t(1-t)\}^{1-15k/14+k\delta}dt\Big) = O(N^{k/2-14/15-7k\delta/15}).$$

According to Assumption (A) and the fact that $\{b_1,b_2,\ldots,b_M\} = \{c_1,c_2,\ldots,c_N\}\setminus$ $\setminus\{c_{d_j}: j\in I\}$, we have

$$\Big|\sum_{j=1}^{M} b_j\Big| = \Big|\sum_{j\in I} c_{d_j}\Big| = O\Big(\frac{m}{\sqrt{N}}\Big) = O(N^{1/30})$$

and similarly

$$\sum_{j=1}^{M} b_j^4 = O(N^{-1}), \quad \Big|\sum\sum_{(j,k)\neq} b_j^3 b_k\Big| = O(N^{-7/15}),$$

$$\text{(4.21)} \qquad \sum\sum_{(j,k)\neq} b_j^2 b_k^2 = 1 + O(N^{-7/15}), \quad \Big|\sum\sum\sum_{(j,k,\ell)\neq} b_j^2 b_k b_\ell\Big| = O(N^{1/15}),$$

$$\Big|\sum\sum\sum\sum_{(j,k,\ell,n)\neq} b_j b_k b_\ell b_n\Big| = O(N^{2/15}).$$

Combining (4.19) through (4.21) we find that $ES_2^4 = \mathcal{O}(N^{-26/15-28\delta/15})$ and hence $E|S_2|^3 = \mathcal{O}(N^{-13/10-7\delta/5})$. In the same way one can obtain the other two assertions in (4.17). □

Define, for real t and N ≥ 2,

$$(4.22)\qquad \rho_N(t) = Ee^{it\sigma_N^{-1}(S_{\omega N}-ES_{\omega N})}$$

and

$$(4.23)\qquad \rho_{1N}(t) = Ee^{it\sigma_N^{-1}S_1}\left\{1 + \frac{it}{\sigma_N}(S_2+S_3+S_4) + \frac{(it)^2}{2\sigma_N^2} S_2^2\right\}.$$

The next lemma shows that ρ_N can be approximated by ρ_{1N}.

LEMMA 4.3. *If the Assumptions* (A) *and* (B) *are satisfied, then*

$$(4.24)\qquad |\rho_N(t)-\rho_{1N}(t)| = \mathcal{O}(t^2N^{-17/15-14\delta/15})$$

uniformly for $|t| \leq \log N$ *and* ω.

PROOF. Repeated use of Lemma XV 4.1 of FELLER (1971) yields

$$|\rho_N(t)-\rho_{1N}(t)| = \mathcal{O}(t^2\sigma_N^{-2}E|S_2||S_3+S_4| + t^2\sigma_N^{-2}E(S_3^2+S_4^2) + $$
$$+ |t|^3\sigma_N^{-3}E|S_2|^3).$$

From (2.5) and (3.8) it follows that for all sufficiently large N there exist positive numbers $\varepsilon_1 \leq \varepsilon_2$ such that $\varepsilon_1 \leq \sigma_N^2 \leq \varepsilon_2$. Lemma 4.2 produces the desired result. □

Clearly our next task is to evaluate the right-hand side of (4.23) and we start with the leading term. According to (4.16) $S_1 = \sum_{j=1}^M b_j\tilde{J}_N(\hat{V}_j)$. We have $E\tilde{J}_N(\hat{V}_1) = 0$ and for all sufficiently large N, there exist positive numbers $\gamma_1 \leq \gamma_2$ such that $\gamma_1 \leq E\tilde{J}_N^2(\hat{V}_1) \leq \gamma_2$ (cf. (2.3)). In the sequel we shall assume

(4.25) $$\sum_{j\in I} c_{d_j}^2 < 1-\gamma$$

for some $\gamma \in (0,1)$, to guarantee that

(4.26) $$\gamma\gamma_1 \le \sigma^2(S_1) \le \gamma_2.$$

Finally we note that Assumptions (A) and (B) imply that $\sum_{j=1}^{M} b_j^4 = \mathcal{O}(N^{-1})$ and that the random variable $\tilde{J}_N(\hat{V}_1)$ has a finite 14-th absolute moment. It follows from the classical theory of Edgeworth expansions for sums of independent and non-identically distributed random variables (see e.g. Lemma VI 4.11 in PETROV (1972)) that

(4.27)
$$\Big|E\exp\{it\,S_1/\sigma(S_1)\} - e^{-\frac{1}{2}t^2}\Big\{1 - \frac{it^3}{6\sigma^3(S_1)}\sum_{j=1}^{M} b_j^3\, E\tilde{J}_N^3(\hat{V}_1) + \frac{t^4}{24\sigma^4(S_1)}\sum_{j=1}^{M} b_j^4\{E\tilde{J}_N^4(\hat{V}_1) - 3[E\tilde{J}_N^2(\hat{V}_1)]^2\} - \frac{t^6}{72\sigma^6(S_1)}\Big\{\sum_{j=1}^{M} b_j^3\, E\tilde{J}_N^3(\hat{V}_1)\Big\}^2\Big\}\Big| = \mathcal{O}(N^{-1}(t^4+|t|^9)e^{-\frac{1}{2}t^2})$$

uniformly for $|t| \le \log N$ and ω for which (4.25) is satisfied. Replacing t by $t_N = t\sigma(S_1)/\sigma_N$ and expanding $\exp\{-\frac{1}{2}t_N^2\}$ we find that uniformly for $|t| \le \log N$ and ω for which (4.25) is satisfied

(4.28)
$$\Big|E\, e^{it\sigma_N^{-1}S_1} - e^{-\frac{1}{2}t^2}\Big\{1 - \frac{it^3}{6\sigma_N^3}\sum_{j=1}^{M} b_j^3\, E\tilde{J}_N^3(\hat{V}_1) + \frac{t^4}{24\sigma_N^4}\sum_{j=1}^{M} b_j^4 \cdot \{E\tilde{J}_N^4(\hat{V}_1) - 3[E\tilde{J}_N^2(\hat{V}_1)]^2\} - \frac{t^6}{72\sigma_N^6}\Big\{\sum_{j=1}^{M} b_j^3 E\tilde{J}_N^3(\hat{V}_1)\Big\}^2 + \frac{t^2}{2\sigma_N^2}(\sigma_N^2-\sigma^2(S_1)) + \frac{t^4}{8\sigma_N^4}(\sigma_N^2-\sigma^2(S_1))^2 - \frac{it^5}{12\sigma_N^5}(\sigma_N^2-\sigma^2(S_1))\sum_{j=1}^{M} b_j^3 E\tilde{J}_N^3(\hat{V}_1)\Big\}\Big| =$$

$$= o(N^{-1}(t^4+|t|^9)e^{-\frac{1}{2}t^2}) + O(|\sigma_N^2-\sigma^2(S_1)|^3|t|P_1(t)e^{-\theta t^2}) +$$

$$+ O(N^{-1}|\sigma_N^2-\sigma^2(S_1)||t|P_2(t)e^{-\theta t^2}),$$

where $0 < \theta < \frac{1}{2}$ and P_1 and P_2 are fixed polynomials.

We now turn to the remaining terms on the right in (4.23). Let

$$\mu_N(t) = E\, e^{it\tilde{J}_N(\hat{V}_1)} \tag{4.29}$$

denote the characteristic function of $\tilde{J}_N(\hat{V}_1)$, so that

$$E\, e^{it\sigma_N^{-1}S_1} = \prod_{j=1}^{M} \mu_N\left(\frac{b_j t}{\sigma_N}\right). \tag{4.30}$$

From the Assumptions (A) and (B) it follows by Taylor expansion that for distinct integers $\ell_1,\ldots,\ell_n$ where $1 \le n \le 4$

$$\prod_{\nu=1}^{n} \mu_N\left(\frac{b_{\ell_\nu} t}{\sigma_N}\right) = 1 - \frac{t^2}{2\sigma_N^2}\left\{\sum_{\nu=1}^{n} b_{\ell_\nu}^2\right\}E\tilde{J}_N^2(\hat{V}_1) + O(N^{-3/2}|t|^3), \tag{4.31}$$

uniformly for $|t| \le \log N$ and ω for which (4.25) is satisfied.

In the last two lemmas we summarize the results we need.

<u>LEMMA 4.4</u>. *If the Assumptions* (A) *and* (B) *are satisfied then, uniformly for* $|t| \le \log N$ *and* ω *for which* (4.25) *is satisfied*

$$\begin{aligned}&\Bigg|E\left(e^{it\sigma_N^{-1}S_1} S_2\right) - E\, e^{it\sigma_N^{-1}S_1}\left\{\frac{it}{\sigma_N} ES_1S_2 + \frac{(it)^2}{2\sigma_N^2} ES_1^2S_2 - \right.\\ &\left.- \frac{(it)^3}{4N\sigma_N^3}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2]\right\}\Bigg| = O\left(N^{-1-\varepsilon_1}|t|P_1(t)e^{-\theta t^2}\right),\end{aligned} \tag{4.32}$$

$$\left|E\left(e^{it\sigma_N^{-1}S_1}S_3\right) - E\, e^{it\sigma_N^{-1}S_1}\left\{\frac{it}{\sigma_N} ES_1S_3\right\}\right| = O\left(N^{-1-\varepsilon_2}|t|P_2(t)e^{-\theta t^2}\right), \tag{4.33}$$

$$\left|E\left(e^{it\sigma_N^{-1}S_1}S_4\right)\right| = O\left(N^{-1-\varepsilon_3}|t|P_3(t)e^{-\theta t^2}\right), \tag{4.34}$$

where $0 < \theta < \frac{1}{2}$, $\varepsilon_j > 0$ *and* P_j *is a fixed polynomial,* $j = 1,2,3$.

PROOF. Because the statements (4.32) through (4.34) are all proved in essentially the same manner, we shall only prove the first statement, by way of an example. An application of Lemma XV 4.1 of FELLER (1971) shows

$$\Big|\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))\} - 1 - \frac{it}{\sigma_N}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k)) -$$
$$- \frac{(it)^2}{2\sigma_N^2}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))^2 - \frac{(it)^3}{6\sigma_N^3}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))^3\Big| \le$$
$$\le \frac{t^4}{\sigma_N^4}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))^4.$$

It follows that

$$E\,\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))\}J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j) =$$
$$= E[J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)][\frac{it}{\sigma_N}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k)) +$$
$$+ \frac{(it)^2}{2\sigma_N^2}(b_j^2\tilde{J}_N^2(\hat{V}_j) + 2b_jb_k\tilde{J}_N(\hat{V}_j)\tilde{J}_N(\hat{V}_k) + b_k^2\tilde{J}_N^2(\hat{V}_k)) + \tag{4.35}$$
$$+ \frac{(it)^3}{6\sigma_N^3}(b_j^3\tilde{J}_N^3(\hat{V}_j)+3b_j^2b_k\tilde{J}_N^2(\hat{V}_j)\tilde{J}_N(\hat{V}_k) + 3b_jb_k^2\tilde{J}_N(\hat{V}_j)\tilde{J}_N^2(\hat{V}_k) + b_k^3\tilde{J}_N^3(\hat{V}_k))] +$$
$$+ \mathcal{O}(t^4\ E|J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)|\{b_j^4\tilde{J}_N^4(\hat{V}_j) + b_k^4\tilde{J}_N^4(\hat{V}_k)\}).$$

We note that it is easy to check that

$$E[J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)]\Big[\sum_{\ell\neq j,k}\Big(\frac{it}{\sigma_N}\,b_\ell\tilde{J}_N(\hat{V}_\ell) + \frac{(it)^2}{2\sigma_N^2}\,b_\ell^2\tilde{J}_N^2(\hat{V}_\ell) +$$
$$+ \frac{(it)^2}{\sigma_N^2}\,b_\ell\tilde{J}_N(\hat{V}_\ell)\Big\{b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k) + \tfrac{1}{2}\sum_{n\neq j,k,\ell} b_n\tilde{J}_N(\hat{V}_n)\Big\}\Big) + \tag{4.36}$$
$$+ \frac{(it)^3}{6\sigma_N^3}\,b_j^3\tilde{J}_N^3(\hat{V}_j)\Big] = 0$$

and hence

$$E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_k\tilde{J}_N(\hat{V}_k))\}J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j) =$$

$$= E[J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)]\Big[\frac{it}{\sigma_N}S_1 + \frac{(it)^2}{2\sigma_N^2}S_1^2 + \frac{(it)^3}{6\sigma_N^3}\{3b_j^2b_k\tilde{J}_N^2(\hat{V}_j)\tilde{J}_N(\hat{V}_k) +$$

$$+ 3b_jb_k^2\tilde{J}_N(\hat{V}_j)\tilde{J}_N^2(\hat{V}_k) + b_k^3\tilde{J}_N^3(\hat{V}_k)\}\Big] + O(t^4\, E|J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)| \cdot$$

$$\cdot \{b_j^4\tilde{J}_N^4(\hat{V}_j) + b_k^4\tilde{J}_N^4(\hat{V}_k)\}).$$

From (4.31) it follows that for distinct integers $1 \le j,k \le M$ and $|t| \le \log N$

$$\prod_{\ell\neq j,k}\mu_N\Big(\frac{b_\ell t}{\sigma_N}\Big) = E\, e^{it\sigma_N^{-1}S_1}\Big\{1 + \frac{t^2}{2\sigma_N^2}(b_j^2+b_k^2)E\tilde{J}_N^2(\hat{V}_1) + O(N^{-3/2}|t|^3)\Big\}, \tag{4.37}$$

uniformly for $|t| \le \log N$ and ω for which (4.25) is satisfied. Hence, combining (4.35) through (4.37) and Assumption (A), we find after some algebra

$$E\Big(e^{it\sigma_N^{-1}S_1}\,S_2\Big) = \sum\sum_{(j,k)\neq}\frac{b_j}{M+1}\prod_{\ell\neq j,k}\mu_N\Big(\frac{b_\ell t}{\sigma_N}\Big)\cdot$$

$$\cdot E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))\}J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j) =$$

$$\begin{aligned} &= \Big[E\, e^{it\sigma_N^{-1}S_1}\Big]\Big[\frac{it}{\sigma_N}ES_1S_2 + \frac{(it)^2}{2\sigma_N^2}ES_1^2S_2 + \\ &+ \frac{(it)^3}{6\sigma_N^3}\sum\sum_{(j,k)\neq}E[J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)]\Big[\frac{3b_j^3b_k}{M+1}\tilde{J}_N^2(\hat{V}_j)\tilde{J}_N(\hat{V}_k) + \\ &+ \frac{3b_j^2b_k^2}{M+1}\tilde{J}_N(\hat{V}_j)\tilde{J}_N^2(\hat{V}_k) + \frac{b_jb_k^3}{M+1}\tilde{J}_N^3(\hat{V}_k)\Big] + \\ &+ \frac{(it)^3}{2\sigma_N^3}\sum\sum_{(j,k)\neq}\frac{b_jb_k(b_j^2+b_k^2)}{M+1}\{E\tilde{J}_N(\hat{V}_k)J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)\}\{E\tilde{J}_N^2(\hat{V}_1)\} + \\ &+ O(N^{-3/2}t^4e^{-\theta t^2}E\{|\tilde{J}_N(\hat{V}_1)| + \tilde{J}_N^4(\hat{V}_1)\}\{|J_N'(\hat{V}_1)| + |J_N'(\hat{V}_2)|\}), \end{aligned} \tag{4.38}$$

uniformly for $|t| \le \log N$ and ω for which (4.25) is satisfied. From

Assumption (B) and (4.8) it follows that (see also (4.19))

(4.39)
$$E|\tilde{J}_N^2(\hat{V}_1)\tilde{J}_N(\hat{V}_2)J_N'(\hat{V}_1)| = O(N^{1/10-7\delta/5});$$
$$E|\tilde{J}_N(\hat{V}_1)J_N'(\hat{V}_1)| = O(N^{1/15-14\delta/15});$$
$$E|\tilde{J}_N^4(\hat{V}_1)J_N'(\hat{V}_1)| = O(N^{1/6-7\delta/3});\ E\tilde{J}_N^2(\hat{V}_1) = O(1);$$
$$E(\tilde{J}_N^4(\hat{V}_1) + |\tilde{J}_N^3(\hat{V}_1)| + |\tilde{J}_N(\hat{V}_1)|)|J_N'(\hat{V}_2)| = O(N^{1/30-7\delta/15}).$$

Finally we obtain by partial integration

(4.40)
$$E\tilde{J}_N(\hat{V}_1)\tilde{J}_N^2(\hat{V}_2)J_N'(\hat{V}_1)(\chi(V_1-V_2)-V_1) =$$
$$= -\tfrac{1}{2}\left(\frac{M+1}{M-1}\right)^2 E\tilde{J}_N^4(\hat{V}_1) + \tfrac{1}{2}\left(\frac{M+1}{M-1}\right)^3\{E\tilde{J}_N^2(\hat{V}_1)\}^2.$$

Combining (4.38) through (4.40) and (4.21) we arrive at (4.32). □

LEMMA 4.5. *If the Assumptions* (A) *and* (B) *are satisfied then, uniformly for* $|t| \le \log N$ *and* ω *for which* (4.25) *is satisfied,*

(4.41)
$$\left|E\left(e^{it\sigma_N^{-1}S_1}S_2^2\right) - \left\{E\, e^{it\sigma_N^{-1}S_1}\right\}\left\{ES_2^2 + \frac{it}{\sigma_N}ES_1S_2^2 + \right.\right.$$
$$\left.\left. + \frac{(it)^2}{4N\sigma_N^2}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2]\right\}\right| = O(N^{-1-\varepsilon}|t|P(t)e^{-\theta t^2}),$$

where $0 < \theta < \tfrac{1}{2}$, $\varepsilon > 0$ *and* P *is a fixed polynomial.*

PROOF. The proof of the statement (4.41) is similar to that of Lemma 4.4 and we shall only provide a sketch. Throughout, all order symbols will be uniform for $|t| \le \log N$ and ω for which (4.25) is satisfied. Let, for distinct j and k, $h(V_j,V_k) = J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)$. Direct computation of $E(\exp\{it\sigma_N^{-1}S_1\}S_2^2)$ shows

$$E\left(e^{it\sigma_N^{-1}S_1}S_2^2\right) = E\left(e^{it\sigma_N^{-1}S_1}\left\{\sum_{j=1}^{M}\frac{b_j}{M+1}\sum_{k\neq j}h(V_j,V_k)\right\}^2\right) =$$

$$= \frac{1}{(M+1)^2}\left[\sum_{j=1}^{M} b_j^2\left\{\sum_{r\neq j} E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_r\tilde{J}_N(\hat{V}_r))\}h^2(V_j,V_r)\,\cdot\right.\right.$$

$$\cdot \prod_{\ell\neq j,r}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right) + \sum_{r\neq j}\sum_{s\neq j,r} E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_r\tilde{J}_N(\hat{V}_r) +$$

$$\left.+ b_s\tilde{J}_N(\hat{V}_s))\}h(V_j,V_r)h(V_j,V_s)\prod_{\ell\neq j,r,s}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right)\right\} +$$

$$+ \sum\sum_{(j,k)\neq} b_j b_k\left\{E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_k\tilde{J}_N(\hat{V}_k))\}h(V_j,V_k)h(V_k,V_j)\,\cdot\right. \tag{4.42}$$

$$\cdot \prod_{\ell\neq j,k}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right) + \sum_{r\neq j,k} E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_k\tilde{J}_N(\hat{V}_k)+b_r\tilde{J}_N(\hat{V}_r))\}\,\cdot$$

$$\cdot\,[h(V_j,V_r)h(V_k,V_r) + 2h(V_j,V_r)h(V_k,V_j)]\prod_{\ell\neq j,k,r}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right) +$$

$$+ \sum_{r\neq j,k}\sum_{s\neq j,k,r} E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_r\tilde{J}_N(\hat{V}_r))\}h(V_j,V_r)\,\cdot$$

$$\left.\left.\cdot\, E\exp\{it\sigma_N^{-1}(b_k\tilde{J}_N(\hat{V}_k)+b_s\tilde{J}_N(\hat{V}_s))\}h(V_k,V_s)\prod_{\ell\neq j,k,r,s}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right)\right\}\right].$$

Using Lemma XV 4.1 of FELLER (1971), we expand all six exponentials in the right-hand side of (4.42) (cf. (4.35)). From (4.31) it follows that for distinct integers $\ell_1,\ldots,\ell_n$ where $1 \le n \le 4$ we have (cf. (4.37))

$$\prod_{\ell\neq\ell_1,\ldots,\ell_n}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right) = \tag{4.43}$$
$$= E\,e^{it\sigma_N^{-1}S_1}\left\{1 + \frac{t^2}{2\sigma_N^2}\left\{\sum_{\nu=1}^{n} b_{\ell_\nu}^2\right\}E\tilde{J}_N^2(\hat{V}_1) + O(N^{-3/2}|t|^3)\right\}.$$

With the aid of (4.43) and the expansions of the exponentials we proceed as in (4.38). For example, the term involving $h(V_j,V_r)h(V_k,V_r)$ on the right in (4.42) equals

$$\frac{1}{(M+1)^2}\sum\sum\sum_{(j,k,r)\neq} b_jb_k \prod_{\ell\neq j,k,r} \mu_N\left(\frac{b_\ell t}{\sigma_N}\right) E\ \exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_k\tilde{J}_N(\hat{V}_k)+b_r\tilde{J}_N(\hat{V}_r))\}\cdot$$

$$\cdot\ h(V_j,V_r)h(V_k,V_r) = \left\{E\ e^{it\sigma_N^{-1}S_1}\right\}\left\{\sum\sum\sum_{(j,k,r)\neq}\frac{b_jb_k}{(M+1)^2}\Big[Eh(V_j,V_r)h(V_k,V_r) +\right.$$

$$+\frac{it}{\sigma_N}ES_1h(V_j,V_r)h(V_k,V_r) + \frac{(it)^2}{2\sigma_N^2}E\Big(b_j^2\tilde{J}_N^2(\hat{V}_j)+b_k^2\tilde{J}_N^2(\hat{V}_k)+b_r^2\tilde{J}_N^2(\hat{V}_r)+2b_jb_k\tilde{J}_N(\hat{V}_j)\tilde{J}_N(\hat{V}_k) +$$

$$+ 2b_jb_r\tilde{J}_N(\hat{V}_j)\tilde{J}_N(\hat{V}_r) + 2b_kb_r\tilde{J}_N(V_k)\tilde{J}_N(\hat{V}_r)\Big)h(V_j,V_r)h(V_k,V_r)\Big] +$$

$$\left.+\frac{t^2}{2\sigma_N^2}\sum\sum\sum_{(j,k,r)\neq}\frac{b_jb_k}{(M+1)^2}(b_j^2+b_k^2+b_r^2)E\tilde{J}_N^2(\hat{V}_1)Eh(V_j,V_r)h(V_k,V_r)\right\} +$$

$$+ O\Big(N^{-3/2}|t|^3\ e^{-\theta t^2}\ E\{|\tilde{J}_N^3(\hat{V}_1)| + |\tilde{J}_N(\hat{V}_1)| + |\tilde{J}_N^3(\hat{V}_3)| +$$

$$+ |\tilde{J}_N(\hat{V}_3)| + 1\}|J_N'(\hat{V}_1)J_N'(\hat{V}_2)|\Big).$$

From the Assumptions (A) and (B) and (4.8) we are able to calculate these sums (cf. (4.21) and (4.39)). Note that by partial integration we have

$$E\tilde{J}_N(\hat{V}_1)\tilde{J}_N(\hat{V}_2)h(V_1,V_3)h(V_2,V_3) = \tfrac{1}{4}\left(\frac{M+1}{M-1}\right)^2[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2].$$

Following this program, we finally arrive at

$$\frac{1}{(M+1)^2}\sum\sum\sum_{(j,k,r)\neq} b_jb_k \prod_{\ell\neq j,k,r} \mu_N\left(\frac{b_\ell t}{\sigma_N}\right) E\ \exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_k\tilde{J}_N(\hat{V}_k)+b_r\tilde{J}_N(\hat{V}_r))\}\cdot$$

$$\cdot\ h(V_j,V_r)h(V_k,V_r) = \left\{E\ e^{it\sigma_N^{-1}S_1}\right\}\left\{\sum\sum\sum_{(j,k,r)\neq}\frac{b_jb_k}{(M+1)^2}\Big[Eh(V_j,V_r)h(V_k,V_r) +\right.$$

$$+\frac{it}{\sigma_N}ES_1h(V_j,V_r)h(V_k,V_r)\Big] + \frac{(it)^2}{4\sigma_N^2}\sum\sum_{(j,k)\neq}\frac{b_j^2b_k^2}{N}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2] +$$

$$+ O(N^{-1-\varepsilon}\ |t|P(t)e^{-\theta t^2}),$$

where $0 < \theta < \frac{1}{2}$, $\varepsilon > 0$ and P is a fixed polynomial. All other terms in the right-hand side of (4.42) can be handled in the same way. □

From Lemmas 4.4 and 4.5 it follows that uniformly for $|t| \le \log N$ and ω for which (4.25) is satisfied (cf. (4.23)),

$$\rho_{1N}(t) = \left\{E\, e^{it\sigma_N^{-1}S_1}\right\}\left\{1 + \frac{(it)^2}{2\sigma_N^2}[2ES_1S_2 + 2ES_1S_3 + ES_2^2] + \right.$$
$$\left. + \frac{(it)^3}{2\sigma_N^3}[ES_1^2S_2 + ES_1S_2^2] - \frac{(it)^4}{8N\sigma_N^4}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2]\right\} +$$
$$+ O(N^{-1-\varepsilon}\ |t|P(t)e^{-\theta t^2}),$$

where $\varepsilon > 0$, $0 < \theta < \frac{1}{2}$ and P is a fixed polynomial. Using (4.26), Lemmas 4.1 and 4.2, as well as the fact that $ES_1S_4 = 0$, we obtain

(4.44)
$$2ES_1S_2 + 2ES_1S_3 + ES_2^2 = \sigma^2(S_{\omega N}) - \sigma^2(S_1) + O(N^{-17/15-14\delta/15}) =$$
$$= \sigma^2(T_{\omega N}) - \sigma^2(S_1) + \left(1 + \left(\sum_{j\in I} c_{d_j}\right)^2\right)^{\frac{1}{2}} O(N^{-1-7\delta/15}),$$

uniformly for ω satisfying (4.25). Writing $h(V_1,V_2) = J_N'(\hat{V}_1)(\chi(V_1-V_2)-V_1)$ as before, we find by repeated use of Assumptions (A) and (B) (cf. (4.20), (4.21) and (4.39)) that, uniformly for ω satisfying (4.25),

$$ES_1^2S_2 + ES_1S_2^2 = \frac{A_{1N}}{N} \sum\sum_{(j,k)\neq} b_j^2 b_k + O(N^{-1-\varepsilon}),$$

where $\varepsilon > 0$ and

(4.45)
$$A_{1N} = E\tilde{J}_N^2(\hat{V}_1)h(V_2,V_1) + 2E\tilde{J}_N(\hat{V}_1)\tilde{J}_N(\hat{V}_2)h(V_1,V_2) +$$
$$+ 2E\tilde{J}_N(\hat{V}_1)h(V_1,V_3)h(V_2,V_3).$$

It follows that uniformly for $|t| \le \log N$ and ω satisfying (4.25),

(4.46)
$$\rho_{1N}(t) = \left\{E\, e^{it\sigma_N^{-1}S_1}\right\}\left\{1 + \frac{(it)^2}{2\sigma_N^2}[\sigma^2(T_{\omega N}) - \sigma^2(S_1)] + \right.$$

$$+ \frac{(it)^3}{2\sigma_N^3}\frac{A_{1N}}{N} \sum\sum_{(j,k)\neq} b_j^2 b_k - \frac{(it)^4}{8N\sigma_N^4}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2]\Big\} +$$

$$+ \mathcal{O}\left(N^{-1-\varepsilon} |t|P(t)e^{-\theta t^2}\left(1 + \left(\sum_{j\in I} c_{d_j}\right)^2\right)^{\frac{1}{2}}\right),$$

where $\varepsilon > 0$, $0 < \theta < \frac{1}{2}$ and P is a fixed polynomial.

Let us turn back to our starting point (4.7). Choose $\gamma \in (0,1)$ and define the event $B = \{\sum_{j\in I} c_{D_j}^2 < 1-\gamma\}$ (cf. (4.25)). According to Lemma 3.4, $P(B^c) = \mathcal{O}(N^{-22/15})$, so

$$\psi_N(t) = E\ e^{itT_N^*} = E\Big[\chi(B)E\Big(e^{it\sigma_N^{-1}\{T_N-Z_N-E(T_N-Z_N|\Omega)\}}\Big|\Omega\Big)\ \cdot$$

$$\cdot\ e^{it\sigma_N^{-1}E(T_N-Z_N|\Omega)}\ E\Big(e^{it\sigma_N^{-1}Z_N}\Big|\Omega\Big)\Big] + \mathcal{O}(N^{-22/15}).$$

From Lemma 3.3 it follows that $E|Z_N|^5 = \mathcal{O}(N^{-1-7\delta/3})$ and $E(E(T_N-Z_N|\Omega))^2 = \mathcal{O}(N^{-4/3-14\delta/15})$. Hence by Taylor expansion we obtain

$$\psi_N(t) = E\ e^{itT_N^*} = E\Big[\chi(B)E\Big(e^{it\sigma_N^{-1}\{T_N-Z_N-E(T_N-Z_N|\Omega)\}}\Big|\Omega\Big)\ \cdot$$

$$\cdot \Big\{1 + \frac{it}{\sigma}\{E(Z_N|\Omega)+E(T_N-Z_N|\Omega)\} + \frac{(it)^2}{2\sigma_N^2}\{E(Z_N^2|\Omega) +$$

(4.47)

$$+ 2E(Z_N|\Omega)E(T_N-Z_N|\Omega)\} + \frac{(it)^3}{6\sigma_N^3}E(Z_N^3|\Omega) + \frac{(it)^4}{24\sigma_N^4}E(Z_N^4|\Omega)\Big\}\Big] +$$

$$+ \mathcal{O}(N^{-22/15}) + \mathcal{O}([t^2+|t|^5]N^{-1-7\delta/3}),$$

uniformly for $|t| \leq \log N$. In view of (4.15), (4.22) and (4.24) we have, uniformly for $|t| \leq \log N$ and ω satisfying (4.25)

$$E\Big(e^{it\sigma_N^{-1}\{T_N-Z_N-E(T_N-Z_N|\Omega=\omega)\}}\Big|\Omega=\omega\Big) = E\ e^{it\sigma_N^{-1}(T_{\omega N}-E(T_{\omega N}))} =$$

(4.48)

$$= \rho_N(t) + \mathcal{O}\left(|t|N^{-1-7\delta/15}\left(1 + \left(\sum_{j\in I} c_{d_j}\right)^2\right)^{\frac{1}{2}}\right) = \rho_{1N}(t) +$$

$$+ \mathcal{O}\left(N^{-1-\varepsilon}|t|P(t)\left(1 + \left(\sum_{j\in I} c_{d_j}\right)^2\right)^{\frac{1}{2}}\right),$$

where $\varepsilon > 0$ and P is a fixed polynomial.

Before substituting this in (4.47) we shall provide uniform bounds for the quantities $\sigma_N^2 - \sigma^2(T_{\omega N})$ and $\sigma^2(T_{\omega N}) - \sigma^2(S_1)$. Theorem II 3.1.c of HÁJEK & ŠIDÁK (1967) and Assumption (A) imply that

$$\sigma^2(T_{\omega N}) = \frac{1}{M-1}\left(1 - \sum_{j\in I} c_{d_j}^2 - \frac{1}{M}\left(\sum_{j\in I} c_{d_j}\right)^2\right)\sum_{j=1}^{M}\left(J_N\left(\frac{j}{M+1}\right) - \bar{J}_N\right)^2,$$

where (cf. (4.8))

$$\bar{J}_N = \frac{1}{M}\sum_{j=1}^{M} J_N\left(\frac{j}{M+1}\right) = \frac{1}{M}\sum_{j=m+1}^{N-m} J\left(\frac{j}{N+1}\right).$$

It follows from (3.13) that $|\bar{J}_N| = O(N^{-13/30-7\delta/15})$ and from Assumption (A) that $|\sum_{j\in I} c_{d_j}| = O(N^{1/30})$, hence

$$\sigma^2(T_{\omega N}) = \frac{1}{M-1}\left(1 - \sum_{j\in I} c_{d_j}^2\right)\sum_{j=1}^{M} J_N^2\left(\frac{j}{M+1}\right) + O(N^{-13/15-14\delta/15}), \tag{4.49}$$

uniformly in ω. Furthermore we know from (3.4) that $|\bar{J}| = O(N^{-13/14-\delta})$, so in view of (2.5) and Assumption (B) we have

$$|\sigma_N^2 - \sigma^2(T_{\omega N})| = \left|\frac{1}{N-1}\sum_{j=1}^{N} J^2\left(\frac{j}{N+1}\right) - \frac{1}{M-1}\left(1 - \sum_{j\in I} c_{d_j}^2\right)\sum_{j=1}^{M} J_N^2\left(\frac{j}{M+1}\right)\right| +$$

$$+ O(N^{-13/15-14\delta/15}) = \left|\frac{1}{N-1}\sum_{j\in I} J^2\left(\frac{j}{N+1}\right) + \frac{1}{M-1}\left(\sum_{j\in I} c_{d_j}^2 - \frac{2m}{N}\right)\sum_{j=1}^{M} J_N^2\left(\frac{j}{M+1}\right)\right| + \tag{4.50}$$

$$+ O(N^{-13/15-14\delta/15}) = O(N^{-2/5-14\delta/15}),$$

uniformly in ω.

To obtain the second bound, we argue as in Lemma 3.1 with J and h(t) replaced by J_N and $h_N(t) = h((N+1)^{-1}(m+(M+1)t))$ to conclude that

$$\frac{1}{M}\sum_{j=1}^{M} J_N^2\left(\frac{j}{M+1}\right) = EJ_N^2(V_1) + O(N^{-14/15-14\delta/15}).$$

One easily verifies that $|EJ_N^2(V_1) - E\tilde{J}_N^2(\hat{V}_1)| = O(N^{-13/15-14\delta/15})$ and together with (4.49) and (4.16) this yields

$$(4.51) \qquad |\sigma^2(T_{\omega N})-\sigma^2(S_1)| = \mathcal{O}(N^{-13/15-14\delta/15}),$$

uniformly in ω.

We now substitute the random versions of (4.48), (4.46) and (4.28) in (4.47). Using (4.50) and (4.51) we find after straightforward computations that uniformly for $|t| \le \log N$

$$
\begin{aligned}
\psi_N(t) = E\Big[\chi(B)\Big\{ & e^{-\frac{1}{2}t^2}\Big(1 - \frac{it^3}{6\sigma_N^3}\Big(\sum_{j=1}^N c_j^3 - \sum_{j\in I} c_{D_j}^3\Big)E\tilde{J}_N^3(\hat{V}_1) + \\
& + \frac{t^4}{24\sigma_N^4}\Big[\sum_{j=1}^N c_j^4[E\tilde{J}_N^4(\hat{V}_1)-3\{E\tilde{J}_N^2(\hat{V}_1)\}^2] - \frac{3}{N}[E\tilde{J}_N^4(\hat{V}_1)-\{E\tilde{J}_N^2(\hat{V}_1)\}^2]\Big] - \\
& - \frac{t^6}{72\sigma_N^6}\Big(\sum_{j=1}^N c_j^3\Big)^2\{E\tilde{J}_N^3(\hat{V}_1)\}^2 + \frac{t^2}{2\sigma_N^2}[\sigma_N^2-\sigma^2(T_N-Z_N|\Omega)] + \\
& + \frac{t^4}{8\sigma_N^4}[\sigma_N^2-\sigma^2(T_N-Z_N|\Omega)]^2 - \frac{it^5}{12\sigma_N^5}\sum_{j=1}^N c_j^3\, E\tilde{J}_N^3(\hat{V}_1)[\sigma_N^2-\sigma^2(T_N-Z_N|\Omega)] + \\
& + \frac{it^3}{2\sigma_N^3}\frac{A_{1N}}{N}\sum_{j\in I} c_{D_j} + o(N^{-1}|t|P(t)e^{-\theta t^2}) + \\
& + \mathcal{O}\Big(N^{-1-\varepsilon}|t|P(t)\Big\{1+\Big(\sum_{j\in I} c_{D_j}\Big)^2\Big\}\Big)\Big\}\Big]\Big[1+\frac{it}{\sigma_N}[E(Z_N|\Omega)+E(T_N-Z_N|\Omega)] - \\
& - \frac{t^2}{2\sigma_N^2}[E(Z_N^2|\Omega) + 2E(Z_N|\Omega)E(T_N-Z_N|\Omega)] - \frac{it^3}{6\sigma_N^3}E(Z_N^3|\Omega) + \\
& + \frac{t^4}{24\sigma_N^4}E(Z_N^4|\Omega)\Big] + \mathcal{O}(N^{-22/15} + |t|P(t)N^{-1-7\delta/3}),
\end{aligned} \qquad (4.52)
$$

where $\varepsilon > 0$, $0 < \theta < \frac{1}{2}$ and P is a fixed polynomial.

A few more facts are needed to complete our calculation of $\psi_N(t)$. First we note that for $a = (m+1)(N+1)^{-1} = \mathcal{O}(N^{-7/15})$, Assumption (B) and (4.8) imply that

$$\int_0^a \{|J(t)|^k + |J(1-t)|^k\}dt = \mathcal{O}(N^{-7/15+k/30-7k\delta/15}),$$

for k = 1,...,4 and hence

$$
\begin{aligned}
&|EJ_N(\hat{V}_1)| = O(N^{-13/30-7\delta/15}),\\
&E\tilde{J}_N^2(\hat{V}_1) = \frac{N+1}{M-1}\int_a^{1-a} J^2(t)dt + O(N^{-13/15-14\delta/15}),\\
&E\tilde{J}_N^3(\hat{V}_1) = \frac{N+1}{M-1}\int_a^{1-a} J^3(t)dt - 3\left(\frac{N+1}{M-1}\right)^2\left\{\int_a^{1-a} J^2(t)dt\right\}\left\{\int_a^{1-a} J(t)dt\right\} +\\
&\qquad + O(N^{-13/10-7\delta/3}),\\
&E\tilde{J}_N^4(\hat{V}_1) = \frac{N+1}{M-1}\int_a^{1-a} J^4(t)dt + O(N^{-13/30-7\delta/15}).
\end{aligned}
\tag{4.53}
$$

Furthermore, Lemma 3.3 yields

$$
\begin{aligned}
E(\sigma_N^2-\sigma^2(T_N-Z_N|\Omega)) &= E(E(Z_N^2|\Omega)) + 2E(E(Z_N|\Omega)E(T_N-Z_N|\Omega)) +\\
&+ O(N^{-4/3-14\delta/15}).
\end{aligned}
\tag{4.54}
$$

Combining (4.53) and (4.54) with (4.52) it follows after some computations and repeated use of Assumptions (A) and (B) that, uniformly for $N^{-3/2} \leq |t| \leq \log N$,

$$
\begin{aligned}
\psi_N(t) &= e^{-\frac{1}{2}t^2}\left\{1 - \frac{it^3}{6\sigma_N^3}\kappa_{3N} + \frac{t^4}{24\sigma_N^4}\kappa_{4N} - \frac{t^6}{72\sigma_N^6}\kappa_{3N}^2\right\} +\\
&+ o(N^{-1}|t|P(t)e^{-\theta t^2}) + O(N^{-1-\varepsilon}|t|P(t)),
\end{aligned}
\tag{4.55}
$$

where $\varepsilon > 0$, $0 < \theta < \frac{1}{2}$, P is a fixed polynomial and κ_{3N} and κ_{4N} are given by (2.9) and (2.10).

To conclude the proof of Theorem 2.1 we note that (3.1) implies

$$
\sigma_N^2 = 1 + O(N^{-6/7-2\delta}).
$$

Substituting this in (4.55) we obtain (4.5) with λ_N as in (4.2) and the proof of the theorem is complete. □

5. TWO-SAMPLE LINEAR RANK STATISTICS

In this section we compare our results with the expansions for the two-sample linear rank statistics in BICKEL & VAN ZWET (1978). Let $1 \le n \le N$, $\lambda = nN^{-1}$ and assume that $\varepsilon \le \lambda \le 1-\varepsilon$ for some fixed $\varepsilon \in (0,\frac{1}{2})$ and all N. Define $c_j = (1-\lambda)/\{N\lambda(1-\lambda)\}^{\frac{1}{2}}$, $j = 1,2,\ldots,n$ and $c_j = -\lambda/\{N\lambda(1-\lambda)\}^{\frac{1}{2}}$, $j = n+1,\ldots,N$. It is easy to check that in this case the c_j's satisfy Assumption (A) and

$$\sum_{j=1}^{N} c_j^3 = \frac{1-2\lambda}{\{N\lambda(1-\lambda)\}^{\frac{1}{2}}}, \qquad \sum_{j=1}^{N} c_j^4 = \frac{1-3\lambda+3\lambda^2}{N\lambda(1-\lambda)}.$$

Taking a scores generating function J which satisfies Assumption (B), we define the two-sample linear rank statistic as in (1.1). For the distribution function F_N^* of the standardized version of this statistic Theorem 2.1 provides an Edgeworth expansion with remainder $o(N^{-1})$:

if

$$\tilde{F}_N(x) = \Phi(x) - \phi(x)\Big\{\frac{1-2\lambda}{6\{N\lambda(1-\lambda)\}^{\frac{1}{2}}}\Big(\int_0^1 J^3(t)dt\Big)(x^2-1) +$$

$$(5.1) \qquad + \frac{1}{24N\lambda(1-\lambda)}\Big[(1-6\lambda+6\lambda^2)\int_0^1 J^4(t)dt - 3(1-2\lambda)^2\Big](x^3-3x) +$$

$$+ \frac{(1-2\lambda)^2}{72N\lambda(1-\lambda)}\Big(\int_0^1 J^3(t)dt\Big)^2(x^5-10x^3+15x)\Big\},$$

then

$$\sup_{x\in\mathbb{R}} |F_N^*(x)-\tilde{F}_N(x)| = o(N^{-1}), \qquad \text{as } N \to \infty.$$

BICKEL & VAN ZWET (1978) consider the two-sample linear rank statistic T_N' for an arbitrary vector of scores $a = (a_1,a_2,\ldots,a_N)$, i.e.

$$(5.2) \qquad T_N' = \sum_{j=1}^{N} a_j V_j,$$

where

$$V_j = \begin{cases} 1 & \text{if } 1 \le D_j \le n, \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 1,2,\ldots,N$ and where $D_1, D_2, \ldots, D_N$ denote the antiranks. In their paper they establish asymptotic expansions for the distribution function of T_N' under the null-hypothesis as well as under contiguous alternatives. A related paper is that of ROBINSON (1978) which deals only with the null-hypothesis.

In order to compare the results in BICKEL & VAN ZWET (1978) with Theorem 2.1 in the present paper we introduce the following assumption on the scores a_j.

ASSUMPTION (C). *Let* $a_j = J(j/(N+1))$ *for* $j=1,2,\ldots,N$. *This scores generating function* J *is twice continuously differentiable on* $(0,1)$ *and*

$$\text{(5.3)} \qquad \limsup_{t \to 0,1} t(1-t)\left|\frac{J''(t)}{J'(t)}\right| < 2;$$

there exist positive numbers $K > 0$ *and* $0 < \beta < 1/6$ *such that its first derivative* J' *satisfies*

$$\text{(5.4)} \qquad |J'(t)| \le K\{t(1-t)\}^{-7/6+\beta} \quad \textit{for } t \in (0,1).$$

Furthermore

$$\text{(5.5)} \qquad \int_0^1 J(t)dt = 0, \qquad \int_0^1 J^2(t)dt = 1.$$

LEMMA 5.1. *If* $\varepsilon \le \lambda \le 1-\varepsilon$ *for some fixed* $\varepsilon \in (0,\frac{1}{2})$ *and Assumption* (C) *are satisfied, then as* $N \to \infty$

$$\text{(5.6)} \qquad \sup_{x \in \mathbb{R}} \left| P\left(\frac{T_N' - ET_N'}{\sigma(T_N')} \le x\right) - \tilde{F}_N(x)\right| = o(N^{-1}),$$

where $\tilde{F}_N$ *is defined in* (5.1).

PROOF. The present lemma is almost an immediate consequence of Corollary 2.1 of BICKEL & VAN ZWET (1978). Assumption (C) guarantees that there exists a positive fraction of the scores which are at a distance of at least

$N^{-3/2}$ log N apart from each other. Furthermore, in view of Lemma 3.1 and Appendix 2 of ALBERS, BICKEL & VAN ZWET (1976), Assumption (C) yields that

$$\sum_{j=1}^{N} a_j = O(N^{1/6-\beta}), \qquad \sum_{j=1}^{N} a_j^2 = N + O(N^{1/3-2\beta}),$$

$$\sum_{j=1}^{N} a_j^3 = N \int_0^1 J^3(t)dt + O(N^{\frac{1}{2}-3\beta}),$$

$$\sum_{j=1}^{N} a_j^4 = N \int_0^1 J^4(t)dt + O(N^{2/3-4\beta}).$$

Substituting this in the expansion $\tilde{R}(x,\bar{\lambda})$ (cf. (2.56) in BICKEL & VAN ZWET (1978)) and standardizing T'_N with the exact variance $\sigma^2(T'_N)$ the result follows. □

For the two-sample case Lemma 5.1 is clearly a better result than Theorem 2.1, as was to be expected. Roughly speaking, Assumption (B) in Theorem 2.1 requires a bit more smoothness than Assumption (C) in Lemma 5.1; it also requires $\int |J|^{14+\varepsilon} < \infty$ instead of $\int |J|^{6+\varepsilon} < \infty$. For practical purposes, however, Assumption (B) is already quite satisfactory. It is gratifying to find that the expansions in the two results coincide. We note that some numerical examples are contained in BICKEL & VAN ZWET (1978).

ACKNOWLEDGEMENT. *It is a pleasure to acknowledge the many helpful discussions with Professor W.R. van Zwet during the preparation of this paper; I greatly appreciated his encouragement and constructive criticism.*

REFERENCES

[1] ALBERS, W., BICKEL, P.J. & W.R. VAN ZWET (1976), *Asymptotic expansions for the power of distributionfree tests in the one-sample problem*, Ann. Statist. 4, 108-156.

[2] BICKEL, P.J. & W.R. VAN ZWET (1978), *Asymptotic expansions for the power of distributionfree tests in the two-sample problem*, Ann. Statist. 6, 937-1004.

[3] DOES, R.J.M.M. (1981), *Berry-Esseen theorems for simple linear rank statistics under the null-hypothesis*, Mathematical Centre Report SW 71, Amsterdam.

[4] DUPAČ, V. & J. HÁJEK (1969), *Asymptotic normality of simple linear rank statistics under alternatives II*, Ann. Math. Statist. 40, 1992-2017.

[5] FELLER, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edition, Wiley, New York.

[6] HÁJEK, J. & Z. ŠIDÁK (1967), *Theory of Rank Tests*, Academic Press, New York.

[7] HÁJEK, J. (1968), *Asymptotic normality of simple linear rank statistics under alternatives*, Ann. Math. Statist. 39, 325-346.

[8] HELMERS, R. (1980), *Edgeworth expansions for linear combinations of order statistics with smooth weight functions*, Ann. Statist. 8, 1361-1374.

[9] HO, S.T. & L.H.Y. CHEN (1978), *An* L_p *bound for the remainder in a combinatorial central limit theorem*, Ann. Probability 6, 231-249.

[10] HUŠKOVÁ, M. (1977), *The rate of convergence of simple linear rank statistics under hypothesis and alternatives*, Ann. Statist. 5, 658-670.

[11] HUŠKOVÁ, M. (1979), *The Berry-Esseen theorem for rank statistics*, Comment. Math. Univ. Carolin. 20, 399-415.

[12] JOHNSON, N.L. & S. KOTZ (1969), *Distributions in Statistics: Discrete Distributions*, Wiley, New York.

[13] PETROV, V.V. (1972), *Sums of Independent Random Variables*, Springer Verlag, Berlin.

[14] ROBINSON, J. (1978), *An asymptotic expansion for samples from a finite population*, Ann. Statist. 6, 1005-1011.

[15] VAN ZWET, W.R. (1980), *On the Edgeworth expansion for the simple linear rank statistic*, Mathematical Centre Report SW 70, Amsterdam, to appear in the Proceedings of the Colloquium on Nonparametric Statistical Inference, I.Vincze editor, Budapest.