Technical Report No. 107
March, 1967

ROBUST ESTIMATES OF LINEAR TEND
IN MULTIVARIATE TIME SERIES

by
G. K. Bhattacharyya
University of Wisconsin, Madison

This work was supported in part by the Office of Naval Research under Contract N00014-74-A-0170, Project No. 042-222.
Theorem 3.1 The test statistics $M_n$ and $L_n$ defined in (2.1) and (2.2) both satisfy (A) and (B) with $\xi = 0$.

Proof. From (2.1) we have

$$M_n(\mathbf{y} + b\mathbf{y}) = \frac{1}{\binom{n}{2}} \sum_{i < j} S(\sum_{j \neq i}(y_j - y_i) + b(j-i)).$$

For every $i < j$ the function $(y_j - y_i) + b(j-i)$ is strictly increasing in $b$ and since the sign function $S(\cdot)$ is nondecreasing $M_n$ satisfies (A). When $\theta = 0$ the components of $\mathbf{y}$ are independent and identically distributed (i.i.d.). Consequently all $(Y_j - Y_i)$ are symmetric around 0 and so are their sign functions. $M_n$ being an average of such symmetrically

relative weights

$$W_i = \frac{i}{\sum_{j=1}^{n} w_j} \cdot$$

(3.4)
\{W_i\} is strictly increasing, \( W_a = 1 \) and \( W_1 \leq 6/n \) \((n+1) < \frac{1}{2} \) whenever \( n > 3 \). Letting the integer \( \ell \) be such that \( W_{\ell-1} < \frac{1}{2} \) and \( W_{\ell} \geq \frac{1}{2} \) the following inequalities hold.

\[
L_n (X - b) = \begin{cases} 
> 0 & \text{if } b < X_{\ell} \\
= 0 & \text{if } X_{\ell} < b < X_{\ell+1} \text{ and } W_{\ell} = \frac{1}{2} \\
< 0 & \text{if } X_{\ell} < b < X_{\ell+1} \text{ and } W_{\ell} > \frac{1}{2} \\
< 0 & \text{if } X_{\ell+1} \leq b 
\end{cases}
\]

The estimate \( \hat{\theta}_L \) based on \( L_n \) test therefore turns out to be

\[
\hat{\theta}_L = \begin{cases} 
X_{\ell} & \text{if } W_{\ell-1} < \frac{1}{2} \text{ and } W_{\ell} > \frac{1}{2} \\
= \frac{1}{2} (X_{\ell} + X_{\ell+1}) & \text{if } W_{\ell-1} < \frac{1}{2} \text{ and } W_{\ell} = \frac{1}{2} \ .
\end{cases}
\]

Comparing this with (3.3), we note that \( \hat{\theta}_L \) may be viewed as a weighted median estimate from the terms \((Y_j - Y_i)/(j-i)\), \( 1 \leq i < j \leq n \) with the associated weights \((j-i)\). From the point of view of computation \( \hat{\theta}_M \) is easier to handle than \( \hat{\theta}_L \). It is shown in section 4 that they are equally efficient in large samples. The classical least square estimate \( \hat{\theta} \) of \( \theta \) can be expressed as a weighted mean of the quantities \((Y_j - Y_i)/(j-i)\) with associated weights \( w_i \), that is

\[
\hat{\theta} = \sum_{1 < j \leq n} (Y_j - Y_i) / \sum_{1 < j} (j-i) .
\]

Thus \( \theta \) and \( \hat{\theta}_L \) have structural similarity in the sense that one is weighted mean and the other is weighted median of the same set of quantities with the same system of weights. Incidentally, observe that the foregoing procedure is readily adaptable to estimating trend parameter in a model.
slightly more general than a linear one. If \( Y_1, Y_2, \ldots, Y_n \) are independent having c.d.f. \( F(y - \mu - \theta g(i)) \) \( i = 1, 2, \ldots, n \), where the sequence \( \{g(i)\} \) is known and is strictly monotone, the procedure leads to the median and weighted median estimates of \( \theta \).

For the purpose of estimating \( \mu \) in the linear trend model we assume that the distribution \( F(y) \) is symmetric around 0, so that \( \mu \) is the location parameter of the trend-deflected time series. If \( \theta \) were known, the observable random variables \( Z_i = Y_i - \theta i, i = 1, 2, \ldots, n \) would be i.i.d. \( F(y - \mu) \) and hence a robust estimate of \( \mu \) could be obtained using a rank test of \( H: \mu = 0 \). In particular for the Wilcoxon signed rank test statistic it is well known that the estimate is the median of the average of all pairs of \( Z_i \). Since \( \theta \) is unknown the random variables \( Z_i \) are not observable and an estimate of \( \mu \) cannot be based on them. However when a robust estimate \( \hat{\theta} \) of \( \theta \) is available, a natural modification would be to consider the observable random variables \( \hat{Y}_i = Y_i - \hat{\theta} i \) instead of \( Z_i \), the modified Wilcoxon test statistic

\[
\hat{Q}_n (\mathcal{Y}) = \frac{2}{n(n+1)} \sum_{1 \leq i < j} I_{[Y_i + Y_j - \hat{\theta}(i+j) > 0]}
\]

(3.7)

and the resulting estimate

\[
\hat{\mu}(\mathcal{Y}) = \text{median}_{1 \leq i \leq j \leq n} \left[ \frac{Y_i + Y_j}{2} - \hat{\theta}(i+j) \right],
\]

(3.8)

where \( I \) is the indicator function. Asymptotic properties of this type of estimate were studied by Adichie [1] in a linear regression model under
the assumption of boundedness of the nonstochastic variate. For \( \hat{\theta} \) we will take the estimates \( \hat{\theta}_M \) and \( \hat{\theta}_L \) introduced earlier. In the following theorem we list a few important small sample properties of the estimates \( \hat{\theta} \) and \( \hat{\mu} \). The proofs being simple are omitted.

**Theorem 3.2**

(i) **For any constants** \( b \) **and** \( c \) **and the unit** \( \eta \)-**vector** \( \underline{1} = (1, 1, \ldots, 1) \)

\[
\hat{\theta} (Y + c \underline{1} + b Y) = \hat{\theta} (Y) + b, \quad \hat{\mu} (Y + c \underline{1} + b Y) = \hat{\mu} (Y) + c
\]

\[
\hat{\theta} (bY) = b\hat{\theta} (Y), \quad \hat{\mu} (bY) = b\hat{\mu} (Y)
\]

(ii) **The distribution of** \( \hat{\theta} (Y) \) **is symmetric around** \( \theta \) **and is independent of** \( \mu \). **Also** \( \hat{\theta} (Y) \) **has symmetric distribution around** \( \mu \) **independently of** \( \theta \).

The asymptotic normality and the efficiency properties of these estimates are discussed in the next section under a more general multivariate set up.

4. **Estimates in multivariate trend model and asymptotic relative efficiency.**

As a linear trend model in a \( p \)-**variate time series** we consider a set of \( n \) independent random \( p \)-vectors \( Y = (Y_{11}, Y_{21}, \ldots, Y_{n1}) \) representing the observations at time \( i = 1, 2, \ldots, n \), \( Y \) having a continuous \( p \)-**variate c.d.f.** \( \Psi (Y - \mu - \Theta') \), where \( \Theta = (\theta_1, \theta_2, \ldots, \theta_p) \) is the vector trend parameter.
where \( A = \mathbf{D}_h^{-1} \mathbf{R} \mathbf{D}_h^{-1} \), \( \mathbf{R}_{p \times p} = (\gamma_{\alpha \beta}) \) is the grade correlation matrix of the c.d.f. \( \Psi \), that is,

\[
\gamma_{\alpha \beta} = 3 \int \int_{-\infty}^{\infty} \left[ 2 \Psi_\alpha(x)^{-1} \right] \left[ 2 \Psi_\beta(y)^{-1} \right] d\Psi_{\alpha \beta}(x, y) , \tag{4.6}
\]

and \( \mathbf{D}_h \) is a \( p \times p \) diagonal matrix with the diagonal vector \( \mathbf{h} = (h_1, h_2, \ldots, h_p) \) given by \( h_\alpha = \int_{-\infty}^{\infty} \psi_\alpha(y) dy \).

Proof. Denote by \( M_n \) and \( L_n \) the vectors of the test statistics (2.1) and (2.2) computed from the marginals of the \( p \)-variate observations \( \chi = (\chi_1, \chi_2, \ldots, \chi_n) \) and consider their limit distributions under a sequence of linear trend parameters \( \theta_n \) converging to \( \theta \) at an appropriate rate \( n^{-3/2} \).

More specifically, we define a sequence of distributions

\( \chi(n) = (\chi^{(1n)}, \chi^{(2n)}, \ldots, \chi^{(nn)}) \) where \( \chi^{(in)}(\chi) = \Psi(\chi - \mu - \theta_n i) \), with
\[ U_{an} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j} \phi_0^\alpha (Y_i, Y_j) \]  \hspace{1cm} (4.9)

it is evident that \( Q_n \) is asymptotically equivalent to the vector \( \hat{U}_{an} \) of the corresponding modified U-statistics (cf. Sukhatme [9]) given by

\[ \hat{U}_{an} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j} \phi_{\hat{\theta}}^\alpha (Y_i, Y_j) \]  \hspace{1cm} (4.10)

**Theorem 4.2** For \( \delta \neq 0 \) let \( K^{(n)} (\chi) = \Psi (\chi - \delta / \sqrt{n}) \) be a sequence of distributions for the i.i.d. random vectors \( Y_1, \ldots, Y_n \), where \( \Psi \) has
marginal densities which are bounded, symmetric and continuous in some neighborhood of 0, then

$$\lim_{n \to \infty} \int \left[ \sqrt{n} \left( \hat{U}_n \left( \hat{\mu}_n - \mu \right) \right) \right] \text{d}K^{(n)} = \Phi_p \left( 2 \mu, \frac{4}{3} \frac{\sigma^2}{\nu} \right). \quad (4.11)$$

Before giving the proof we remark that the lemma 4.1 applied to (4.11) at once yields

**Corollary 4.1** Under the conditions on $\Psi$ mentioned above

$$\lim_{n \to \infty} \int \left[ \sqrt{n} \left( \hat{\mu}_n - \mu \right) \right] = \Phi_p \left( \frac{\sigma}{\nu}, \frac{1}{3} \frac{\sigma^2}{\nu} \right). \quad (4.12)$$

The proof of theorem 4.2 is accomplished through the following two lemmas.

Associated with the kernel $\phi_t$ of the modified U-statistic $\hat{U}_n$ define

$$A_{ij}^{(\alpha)}(t) = E \phi_t^{(\alpha)}(Y_i, Y_j), A^{(\alpha)}(t) = \frac{1}{p^2} \sum_{i < j} A_{ij}^{(\alpha)}(t) \quad (4.13)$$

$$A(t) = (A^{(1)}(t_1), \ldots, A^{(p)}(t_p))$$

and

$$I_n = \hat{U}_n - A(\hat{\theta}_n). \quad (4.14)$$
Lemma 4.2  Under the conditions of theorem 4.2

\[
\lim_{n \to \infty} \sqrt{n} \left[ \frac{1}{n} - \left( \overline{U}_n - E(U_n) \right) \right] = \mathcal{Q}
\]  \hspace{1cm} (4.15)

\[
\lim_{n \to \infty} \mathcal{E} \left( \sqrt{n} \frac{1}{n} | K^{(n)} \right) = \Phi_p \left( \mathcal{Q}, \frac{1}{3} \mathcal{I} \right) \]  \hspace{1cm} (4.16)

Proof. The first result follows by the use of theorem 4.1 and by adaptation of the conditions and the proof of theorem 3.1 of Sukhatme [9] on the limit distribution of modified U-statistic to the present time series model. The details are tedious but straightforward and therefore are omitted. The second result then follows from the fact that under $K^{(n)}$ the limit distribution of $\sqrt{n} \left[ \overline{U}_n - E(U_n) \right]$ is $\Phi_p \left( \mathcal{Q}, \frac{1}{3} \mathcal{I} \right)$, c.f. Bickel [4].

Lemma 4.3  Under the conditions of theorem 4.2

\[
\lim_{n \to \infty} \mathcal{E} \left[ \frac{n^{3/2}}{\hat{\theta}_n}, \sqrt{n} \frac{1}{n} | K^{(n)} \right] = \Phi_{2p} \left( \mathcal{Q}, \mathcal{B} \right) \]  \hspace{1cm} (4.17)

where

\[
B_{2p \times 2p} = \begin{pmatrix}
\hat{\Lambda} & \mathcal{Q} \\
\mathcal{Q} & \frac{1}{3} \mathcal{I} \\
\end{pmatrix}
\]

Proof. Consider first $\hat{\theta}_n = \hat{\theta}_{n \to \infty}$ where $\hat{\theta}_{n \to \infty}$ is defined in (4.2). Taking into account the limiting normality of $\hat{\theta}_n$ and of $\frac{1}{n}$ already established
and the fact that $\hat{\theta}_{M_n}$ is based on the parent test statistic $M_n$, it is sufficient to show that $\sqrt{n} M_n$ and $\sqrt{n} U_n$ are asymptotically independent. By Hoeffding's theorem these have asymptotically $2p$-variate normal distribution and hence it is enough to show that their limiting covariances are zero, i.e.

$$\lim_{n \to \infty} nE \left( \frac{M_n \ U_n}{\sim_n \sim_n} \right) = O_{p \times p}.$$  \hspace{1cm} (4.18)

From (2.1) and (4.9) we have, for any $1 \leq \alpha, \beta \leq p$

$$E \left( M_{\alpha n} \ U_{\beta n} \right) = \frac{1}{\binom{n}{2}} \sum_{{i,j}} \Sigma^* T_{ij(i')j'}^{(\alpha, \beta)} \hspace{1cm} (4.19)$$

where

$$T_{ij(i')j'}^{(\alpha, \beta)} = \mathbb{E} \left\{ S(Y_{\alpha j} - Y_{\alpha i}) Y_{\beta j} + Y_{\beta j'} > 0 \right\}$$

and $\Sigma^*$ denotes summation over all $1 \leq i, j, i', j' \leq n$ satisfying $i < j$ and $i' < j'$. We have clearly

$$T_{ij(i')j'}^{(\alpha, \beta)} = P(Y_{\alpha j} > Y_{\alpha i}, Y_{\beta j} + Y_{\beta j'} > 0) - P(Y_{\alpha j} < Y_{\alpha i}, Y_{\beta j} + Y_{\beta j'} > 0).$$

For $\alpha \neq \beta$ the vectors $(Y_{\alpha i}, Y_{\beta i})$, $i = 1, 2, \ldots, n$ are i.i.d. and further, each of the two components is symmetrically distributed around 0. Consequently for $i, j, i', j'$ all different $T_{ij(i')j'}^{(\alpha, \beta)} = 0$. symmetry and i.i.d. property also entail that for $i \neq i'$ and $j \neq j'$. 
\[ T^{(\alpha, \beta)} \left( \sum_{ij} \right) + T^{(\alpha, \beta)} \left( \sum_{ij} \right) = 0 \]

\[ T^{(\alpha, \beta)} \left( \sum_{ij} \right) + T^{(\alpha, \beta)} \left( \sum_{ij} \right) = 0 . \]

Such terms together account for all but \( \left( \frac{n}{2} \right) \) terms of the sum \( \sum^k \) in (4.19).

Noting that \( |T| \leq 1 \), we then have \( E(M_{an} U_{bn}) = O(n^{-2}) \) for \( \alpha \neq \beta \). For \( \alpha = \beta \) it can be seen in the same way that (4.19) is identically 0 and hence (4.18) follows. For the case \( \hat{\theta}_n = \hat{\theta}_l_n \), the proof of (4.17) is exactly parallel.

To complete the proof of theorem 4.2 we write

\[ \sqrt{n} \left( \hat{\theta}_n - \frac{1}{2} \right) = \sqrt{n} \left( \hat{\theta}_n - \frac{1}{2} \right) + \sqrt{n} \left[ A(\hat{\theta}_n) - A(\theta) \right] \]

\[ + \sqrt{n} \left[ A(\theta) - \frac{1}{2} \right]. \tag{4.20} \]

Let \( \Psi^* = \Psi \ast \Psi \) denote the twofold convolution of \( \Psi \) and \( \Psi^* \) its density, so that by symmetry of \( \Psi \), we have \( \Psi^* (0) = \frac{1}{2} \) and \( \Psi^* (0) = \int_0^\infty \Psi^* (y) \, dy \neq 0 \) for \( \alpha < \infty \). Using (4.13) and the definition of \( K^{(n)} \) we have

\[ \lim_{n \to \infty} \sqrt{n} \left[ A(\alpha) (0) - \frac{1}{2} \right] = \lim_{n \to \infty} \sqrt{n} \left[ \frac{1}{2} - \Psi^* \left( -2 \delta \right) \right] = 2 \delta \alpha \]

and hence

\[ \lim_{n \to \infty} \sqrt{n} \left[ A(\theta) - \frac{1}{2} \right] = 2 \eta. \tag{4.21} \]
Also \( A_{ij}^{(\alpha)}(t) = 1 - \psi_{\alpha}^* \left[ t(i+j) - 2\delta_{\alpha} / \sqrt{n} \right] \) gives

\[
\lim_{n \to \infty} \frac{1}{n} A^{(\alpha)}(0) = - \lim_{n \to \infty} \psi_{\alpha}^* \left\{ - \frac{2\delta_{\alpha}}{\sqrt{n}} \right\} \frac{1}{n(\sigma^2)} \sum_{i < j} (i+j)
\]

\[
= - \psi_{\alpha}^*(0) = - h_{\alpha}, \quad (4.22)
\]

where the prime denotes the first derivative. By mean value theorem

\[
\sqrt{n} \left[ A^{(\hat{\alpha})}_{\alpha n} - A(Q) \right] = n^{3/2} \hat{\alpha}_{\alpha n} \left[ A^{(H_{\alpha n})} / n \right] \quad \text{where} \quad H_{\alpha n} = \Delta_{\alpha \alpha n}, \quad |\Delta_{\alpha}| < 1.
\]

Since \( \hat{\alpha}_{\alpha n} \to 0 \) in probability we have from (4.22)

\[
\text{Plim}_{n \to \infty} \left[ \frac{1}{n} A^{(H_{\alpha n})} \right] = - h_{\alpha}
\]

and hence by theorem 4.1 and Slutsky's theorem

\[
\lim_{n \to \infty} \int \left[ \sqrt{n} \left\{ A^{(\hat{\alpha}_{\alpha n})} - A(Q) \right\} \right| K^{(n)} \right] = \Phi_p (Q, \xi). \quad (4.23)
\]

(4.21), (4.23) and the two lemmas together complete the proof of the theorem.

Finally we consider the asymptotic relative efficiency (ARE) of the median estimates of the trend and location parameters relative to the corresponding classical estimates (4.1). The asymptotic multivariate normality of \( \hat{\theta} \) and \( \hat{\mu} \) in (4.1) follows by application of Liapounoff's central limit theorem under the assumption that for some \( \delta > 0 \) the marginals of \( \Psi \).
have \((2+\delta)\) th absolute moment. More specifically

\[
\lim_{n \to \infty} \int_{\Theta, \mu} \left[ n^{3/2} (\hat{\theta}_n - \theta) \right] = \Phi_p (L, 12 \Sigma)
\]

\[
\lim_{n \to \infty} \int_{\Theta, \mu} \left[ n^{1/2} (\hat{\mu}_n - \mu) \right] = \Phi_p (L, 4 \Sigma)
\]

where \(\Sigma\) is the dispersion matrix of \(\Psi\). For the median estimates \(\hat{\Theta}_M, \hat{\Theta}_L\) and \(\hat{\Theta}\) the ARE defined through the asymptotic generalized variances (c.f. Bickel [4]) become

\[
e(\hat{\Theta}_M, \hat{\Theta}) = e(\hat{\Theta}_L, \hat{\Theta}) = (12)^{1/3} \left| \Sigma \Lambda^{-1} \right|^{1/3p} \tag{4.24}
\]

\[
e(\hat{\Theta}, \hat{\mu}) = 12 \left| \Sigma \Lambda^{-1} \right|^{1/p} \tag{4.25}
\]

5. **Final remarks** Asymptotic comparisons of the median estimates, introduced here for the trend and location parameters in multivariate time series model, with the classical ones can be studied by considering bounds of the ARE (4.24) and (4.25) for various classes of parent c.d.f. \(\Psi\). Referring to the expression (4.3) of Bickel [4] one recognizes at once that (4.25) is precisely the ARE of his \(W_n\) estimate relative to the classical estimate for the location parameter of a multivariate distribution and (4.24) is cube root of the same. Bounds of (4.25) for various classes of parent distributions are investigated by Bickel and from there the bounds of
(4.24) trivially follow. In particular, when \( \Psi \) is p-variate normal the ARE bounds of the median estimates of the trend parameter \( \theta \) come out as

\[
e(\hat{\theta}, \theta) = .98 \quad \text{for } p = 1 \quad (5.1)
\]

\[
.97 \leq e(\hat{\theta}, \theta) \leq 1 \quad \text{for } p = 2 \quad (5.2)
\]

\[
0 < e(\hat{\theta}, \theta) < 1 \quad \text{for } p > 2. \quad (5.3)
\]

The first two results show that in the case of univariate and bivariate normal parent distributions the estimates \( \hat{\theta}_M \) and \( \hat{\theta}_L \) behave almost as good as the classical ones. The upper bound of 1 in (5.3) obtained in Bhattacharyya [2], shows that in the normal case \( \hat{\theta}_M \) and \( \hat{\theta}_L \) are always less efficient than \( \hat{\theta} \) and the lower bound 0 implies that the loss of efficiency could be very high in particular cases when dimensionality exceeds two. In the univariate case, for all continuous parent c.d.f.'s \( \Psi \) Hodges-Lehmann's bound gives \( e(\hat{\theta}, \theta) \geq .95 \) and hence \( \hat{\theta}_M \) and \( \hat{\theta}_L \) are always as good as and in many cases much better than \( \hat{\theta} \). The robustness properties of \( \hat{\theta}_M \), \( \hat{\theta}_L \) and \( \hat{\theta} \) in multivariate situations having contamination from heavy tailed distributions follows from Bickel [4].

**ACKNOWLEDGEMENT**

The author wishes to thank Dr. J. H. Klotz and Dr. R. Johnson for many helpful discussions.
REFERENCES


3. ROBUST ESTIMATES OF LINEAR TREND IN MULTIVARIATE TIME SERIES

5. Bhattacharyya, Gouri K.

6. April, 1967

7a. 19pp.  7b. 9 references

8a. ONR
8b. Contract NONR 1202 (17) 042-222

9a. Technical Report No. 107
9b. --

10. Distribution of this document is unlimited

11. --

12. Office of Naval Research
    Washington, D.C.

13. Abstract: Median and weighted median estimates are obtained for the linear trend parameters of a univariate time series by applying the Hodges-Lehmann method to some well known nonparametric tests for trend. The estimation procedure is extended to the multivariate trend model and the asymptotic efficiency properties relative to the classical estimates are studied.