UCLA
Department of Statistics Papers

Title
Kendall's tau for autocorrelation

Permalink
https://escholarship.org/uc/item/7jt8s827

Authors
Ferguson, Thomas
Genest, Christian
Hallin, Marc

Publication Date
1999-06-01

eScholarship.org Powered by the California Digital Library
University of California

The Canadian Journal of Statistics
SUBMISSION/PROPOSITION NO 62-JUNE'99
La revue canadienne de statistique

Kendall's tau for autocorrelation

Thomas S. FERGUSON, Christian GENEST and Marc HALLIN
University of California at Los Angeles, Université Laval and
Université libre de Bruxelles

Key words and phrases: Asymptotic relative efficiency; autocorrelation; Kendall's tau; rank procedures; Spearman's rho; tests of independence; time series.
Mathematics subject classification codes (1991): primary 62M10, 90A20; secondary 62G10, 60F05.

ABSTRACT

The authors show how Kendall's tau can be adapted to test against serial dependence in a univariate time series context. They provide formulas for the mean and variance of circular and non-circular versions of this statistic, and they prove its asymptotic normality. They present also a Monte Carlo study comparing the power and size of a test based on Kendall's tau to that of competing procedures based on alternative parametric and nonparametric measures of serial dependence. In particular, their simulations indicate that Kendall's tau outperforms Spearman's rho in detecting first-order autoregressive dependence, despite the fact that these two statistics are asymptotically equivalent.

RÉSUMÉ

Les auteurs montrent comment le tau de Kendall peut être adapté pour tester la présence de dépendance sérielle dans une série chronologique univariée. Ils déterminent l'espérance et la variance de versions circulaire et non-circulaire de cette statistique et en démontrent la normalité asymptotique. Une étude de Monte-Carlo leur permet aussi de comparer le seuil et la puissance d'un test fondé sur cette statistique à celle de tests concurrents s'appuyant sur d'autres mesures paramétriques et non paramétriques de dépendance sérielle. Leurs simulations indiquent entre autres que le tau de Kendall détecte plus facilement la présence de dépendance autorégressive de premier ordre que le rho de Spearman, bien que ces deux statistiques soient asymptotiquement équivalentes.

1. INTRODUCTION

Testing for randomness against serial dependence is a fundamental problem in time series analysis. To determine whether stock prices or exchange rates form a random walk, for instance, statistical procedures must be used to see whether successive changes (or log-changes) are mutually independent. Correlograms are the traditional inference tools for such problems. Though they remain valid under fairly general distributional assumptions, correlogram-based methods remain a Gaussian tool, as they typically do not allow for locally and asymptotically optimal inference beyond Gaussian linear processes (cf., e.g., Hallin and Werker 1999).

In the absence of any information on the distribution of the series under study, rank-based methods offer an obvious alternative to traditional correlograms. Their robustness and excellent performance in small and large samples have been recognized by a number of authors; the surveys by Hallin and Puri (1992) and Hallin and Werker (1999) provide fairly complete introductions to the subject.

Given a sequence $X_1, \ldots, X_n$ of $n \geq 3$ continuous random variables and their associated ranks $R_1, \ldots, R_n$, nonparametric measures of first-order serial dependence are typically based on the pairs

$$(R_1, R_2), (R_2, R_3), \ldots, (R_{n-1}, R_n), \tag{1}$$

possibly augmented with (R_n, R_1), in which case the statistic is termed *circular*. Variants of the Spearman-Wald-Wolfowitz autocorrelation statistic, for example, involve the sample correlation of these pairs (Wald and Wolfowitz 1943). Such is also the case, among others, for the van der Waerden, the Wilcoxon, and the Laplace or median test-score autocorrelation coefficients (Hallin and Puri 1988).

The purpose of this note is to investigate a serial version of Kendall's tau as an alternative to these rank-based measures of serial dependence. For the sake of simplicity, the presentation concentrates on the first-order case or lag-one serial dependence, though higher-order versions are considered in the final section.

Taking the subscripts to be written modulo n, so that $R_{n+1} \equiv R_1$, a serial version of Kendall's tau may be defined, *in the circular case*, as

$$\tau_n = 1 - 2N \Big/ \binom{n}{2} = 1 - \frac{4N}{n(n-1)}, \tag{2}$$

where N is the number of discordances, that is, the number of pairs (R_i, R_{i+1}) and (R_j, R_{j+1}) that satisfy either $R_i < R_j$ and $R_{i+1} > R_{j+1}$, or $R_i > R_j$ and $R_{i+1} < R_{j+1}$. More specifically, one has

$$\begin{aligned} N &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \{I(R_i < R_j, R_{i+1} > R_{j+1}) + I(R_i > R_j, R_{i+1} < R_{j+1})\} && (3) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} I(R_i < R_j, R_{i+1} > R_{j+1}), && (4) \end{aligned}$$

where $I(A)$ represents the indicator function of the set A. Substituting $n-1$ for n in expressions (2) and (4) yields *the non-circular version* of τ and N.

A test of first-order serial dependence based on (2) is introduced in Section 2, where finite-sample and asymptotic expressions for the mean and variance of the circular and non-circular versions of τ_n are given under the null hypothesis of randomness. Sections 3 and 4 contain derivations of these results, and the quality of the limiting approximation is investigated in Section 5. The simulation results reported in Section 6 illustrate the excellent performance of the new test, both in terms of size and power. Though they are shown in Section 7 to be asymptotically equivalent, Kendall's tau is seen to outperform Spearman's rho in small samples for the selected alternatives. The final section briefly describes higher-order extensions.

2. A TEST OF RANDOMNESS BASED ON KENDALL'S TAU

Consider a series $X_1, \ldots, X_n$ of $n \geq 3$ observations, and suppose that one wished to test for randomness against first-order serial dependence, using the nonparametric statistic τ_n defined in (2). As is traditional in the time series literature, the term *randomness* refers here to mutual independence between the X_i's, which are also assumed throughout to arise from the same continuous distribution, so that the probability of tied ranks is zero. It should be noted, however, that the present developments would remain equally valid under the more general assumption of exchangeability.

If positive dependence is suspected, an exact one-sided test based on the serial version of Kendall's tau would reject the null hypothesis of randomness whenever

$$T_n = \frac{\tau_n - E(\tau_n)}{\sqrt{\mathrm{Var}(\tau_n)}}$$

is larger than some critical value $t_{n,\alpha}$ such that $P(T_n > t_{n,\alpha}) = \alpha$, a predetermined level. If the series were sufficiently long, or if one had confidence that the asymptotic distribution of T_n is an appropriate approximation, one could also reject the null hypothesis for $T_n > t_\alpha$, where t_α is such that $\lim_{n\to\infty} P(T_n > t_\alpha) = \alpha$. The case of an alternative of negative serial dependence can be treated *mutatis mutandis*.

To carry out the above procedure, it is necessary to derive the mean and the variance of τ_n under the null hypothesis of randomness, and to determine its asymptotic distribution. This information is summarized in the following propositions, whose proofs are given in Sections 3 and 4, respectively. Although the results cover both the circular and the non-circular versions of τ_n, time series applications of the test would typically be based on the latter statistic.

PROPOSITION 1. *The circular and non-circular versions of* τ_n *based on a random sample of size* $n \geq 3$ *have the same mean, viz.,*

$$E(\tau_n) = -\frac{2}{3(n-1)} = O(1/n),$$

but different variances. In the circular *case, Var*$(\tau_n) = 0$ *when* $n = 3$ *and*

$$\mathrm{Var}(\tau_n) = \frac{20n^3 - 14n^2 - 98}{45n^2(n-1)^2} = \frac{4}{9n} + o(1/n),$$

for $n \geq 4$. *In the* non-circular *case, Var*$(\tau_n) = 8/9$ *when* $n = 3$ *and*

$$\mathrm{Var}(\tau_n) = \frac{20n^3 - 74n^2 + 54n + 148}{45(n-1)^2(n-2)^2} = \frac{4}{9n} + o(1/n),$$

for $n \geq 4$.

PROPOSITION 2. *Under the null hypothesis of randomness, the circular and non-circular versions of* $\sqrt{n}\tau_n$ *are asymptotically distributed as normal random variables with mean zero and variance* $4/9$.

As shown in Section 4, the quality of this approximation is excellent for all samples of size $n > 10$, so that the asymptotic critical values t_α of the one-sided described above can be taken as the $(1-\alpha)$-quantile of the standard normal distribution.

3. PROOF OF PROPOSITION 1 ON THE FIRST TWO MOMENTS OF τ_n

Proposition 1 is a direct consequence of the following result, which pertains to the number N of discordances in the set (1), augmented with the pair (R_n, R_1) in the circular case. The statement of the lemma excludes the case $n = 3$, as in that case, the circular version of N is identically equal to 2, while the non-circular version is distributed as a Bernoulli random variable with parameter $P(N = 1) = 2/3$.

LEMMA. *Under the null hypothesis of randomness, the mean and variance of the* circular *version of N are*

$$E(N) = \frac{n(3n-1)}{12} \quad \textit{and} \quad \mathrm{Var}(N) = \frac{10n^3 - 7n^2 - 49}{360}, \quad n \geq 4.$$

For the non-circular *version of N, one has*

$$E(N) = \frac{(n-2)(3n-1)}{12} \quad \textit{and} \quad \mathrm{Var}(N) = \frac{10n^3 - 37n^2 + 27n + 74}{360}, \quad n \geq 4.$$

To check the formula for $E(N)$ in the circular case, it suffices to observe that

$$\begin{aligned} E(N) &= \sum_{i=1}^{n}\sum_{j=1}^{n} P\left(R_i < R_j, R_{i+1} > R_{j+1}\right) \\ &= \sum_{i=1}^{n} \left\{2P(R_1 < R_2 > R_3) + (n-3)P\left(R_1 < R_2, R_3 > R_4\right)\right\} \\ &= n\{2/3 + (n-3)/4\} = n(3n-1)/12. \end{aligned} \tag{5}$$

In the non-circular case, one has

$$\begin{aligned} E(N) &= \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1} \left\{P(R_i < R_j, R_{i+1} > R_{j+1}) + P(R_i > R_j, R_{i+1} < R_{j+1})\right\} \\ &= \sum_{i=1}^{n-2} \left\{2P(R_1 < R_2 > R_3) + (n-3)P\left(R_1 < R_2, R_3 > R_4\right)\right\} \\ &= (n-2)\{2/3 + (n-3)/4\} = (n-2)(3n-1)/12. \end{aligned} \tag{6}$$

Note that these formulas give the right answer even in the case $n = 3$.

The computation of the second moment of N, viz.,

$$E\left(N^2\right) = \sum_i \sum_j \sum_k \sum_\ell P\left(R_i < R_j, R_{i+1} > R_{j+1}, R_k < R_\ell, R_{k+1} > R_{\ell+1}\right)$$

for $n \geq 4$ is somewhat more involved. It is presented here in the circular case, but the changes required to handle the non-circular case are indicated along the way.

The above formula may be decomposed as the sum of four terms, the first when both $i = k$ and $j = \ell$, the second when exactly one of $i = k$ and $j = \ell$, the third when $i \neq k$ and $j \neq \ell$ but exactly one of $i = \ell$ and $k = j$, and the fourth when all of i, j, k, and ℓ are distinct. (The term with both $i = \ell$ and $k = j$ is zero.) This yields

$$E\left(N^2\right) = \sum_i \sum_j P\left(R_i < R_j, R_{i+1} > R_{j+1}\right) \tag{7}$$

$$+ \quad 2\sum_i \sum_j \sum_{\ell \neq j} P\left(R_i < R_j, R_{i+1} > R_{j+1}, R_i < R_\ell, R_{i+1} > R_{\ell+1}\right) \tag{8}$$

$$+ \quad 2\sum_i \sum_{k\neq i} \sum_{j \neq k} P\left(R_i < R_j, R_{i+1} > R_{j+1}, R_k < R_i, R_{k+1} > R_{i+1}\right) \tag{9}$$

$$+ \quad \sum\sum\sum\sum_{i,j,k,\ell \text{ distinct}} P\left(R_i < R_j, R_{i+1} > R_{j+1}, R_k < R_\ell, R_{k+1} > R_{\ell+1}\right). \tag{10}$$

Term (7) is exactly $E(N)$, which was already computed in (5) for the circular case and in (6) for the non-circular case. The evaluation of the three other terms is treated separately.

Evaluation of term (8). Fix i and consider five cases, according as (a) j and ℓ are next to i with i on the end; (b) j and ℓ are next to i with i in the middle; (c) one of j and ℓ is next to i and the other is at least two away; (d) j and ℓ are next to each other and i at least two away; or (e) all of i, j and ℓ at least two away from each other.

Term (8) may then be written as

$$\begin{aligned}2\sum_i \Big\{&4P\left(R_1 < R_2, R_2 > R_3, R_1 < R_3, R_2 > R_4\right)\\
&+ \quad 2P\left(R_2 < R_1, R_3 > R_2, R_2 < R_3, R_3 > R_4\right)\\
&+ \quad 4(n-4)P\left(R_1 < R_2, R_2 > R_3, R_1 < R_4, R_2 > R_5\right)\\
&+ \quad 2(n-4)P\left(R_1 < R_3, R_2 > R_4, R_1 < R_4, R_2 > R_5\right)\\
&+ \quad (n-5)(n-4)P\left(R_1 < R_3, R_2 > R_4, R_1 < R_5, R_2 > R_6\right)\Big\}\end{aligned}$$

with the sum over i running from 1 to n in the circular case, and from 1 to $n-3$ in the non-circular case. Therefore, term (8) reduces to

$$2n\left\{4\,\frac{3}{24} + 2\,\frac{5}{24} + 4(n-4)\,\frac{18}{120} + 2(n-4)\,\frac{11}{120} + (n-5)(n-4)\,\frac{1}{9}\right\}$$

$$= n\left\{\frac{11}{6} + \frac{47}{30}(n-4) + \frac{2}{9}(n-4)(n-5)\right\} \tag{11}$$

in the circular case, the front factor of n being replaced by $(n-3)$ in the non-circular case.

Evaluation of term (9). Of the six cases in which i, j and k are consecutive (in some order), there are exactly four nonempty cases, which yield

$$\begin{aligned}
P\left(R_1 < R_2, R_2 > R_3, R_2 < R_4, R_3 > R_1\right) &= 1/24,\\
P\left(R_1 < R_3, R_2 > R_4, R_3 < R_2, R_4 > R_3\right) &= 1/24,\\
P\left(R_1 < R_3, R_2 > R_4, R_3 < R_4, R_4 > R_1\right) &= 1/24,\\
P\left(R_1 < R_4, R_2 > R_1, R_4 < R_2, R_1 > R_3\right) &= 1/24.
\end{aligned}$$

Term (9) is the sum of this plus the three cases with exactly two next to each other, plus the case of all separated. This is

$$\begin{aligned}
2\sum_i \Big\{ & 4/24 + 2(n-4)P\left(R_1 < R_4, R_2 > R_5, R_2 < R_1, R_3 > R_2\right) \\
& + \quad 2(n-4)P\left(R_1 < R_2, R_2 > R_3, R_4 < R_1, R_5 > R_2\right) \\
& + \quad (n-4)P\left(R_1 < R_3, R_2 > R_4, R_4 < R_1, R_5 > R_2\right) \\
& + \quad (n-4)P\left(R_1 < R_4, R_2 > R_5, R_3 < R_1, R_4 > R_2\right) \\
& + \quad (n-4)(n-5)P\left(R_1 < R_3, R_2 > R_4, R_5 < R_1, R_6 > R_2\right) \Big\} \qquad (12)
\end{aligned}$$

with the sum on i running from 1 to n in the circular case, and from 1 to $n-3$ in the non-circular case. Therefore, term (9) reduces to

$$\begin{aligned}
2n & \left\{ \frac{1}{6} + 2(n-4)\,\frac{3}{120} + 2(n-4)\,\frac{3}{120} + 2(n-4)\,\frac{6}{120} + (n-4)(n-5)\,\frac{1}{36} \right\} \\
& = n\left\{ \frac{1}{3} + (n-4)\,\frac{2}{5} + (n-4)(n-5)\,\frac{1}{18} \right\}
\end{aligned}$$

in the circular case, the front factor of n being replaced by $(n-3)$ in the non-circular case.

Evaluation of term (10). First suppose $n = 4$. When all of i, j, k and ℓ are distinct, one must consider three cases circularly arranged in the orders $ik\ell j$, $ijk\ell$, and $i\ell jk$. These lead to

$$\begin{aligned}
P\left(R_1 < R_4, R_2 > R_1, R_2 < R_3, R_3 > R_4\right) &= 2/24, \\
P\left(R_1 < R_2, R_2 > R_3, R_3 < R_4, R_4 > R_1\right) &= 4/24, \\
P\left(R_1 < R_3, R_2 > R_4, R_4 < R_2, R_1 > R_3\right) &= 0.
\end{aligned}$$

The contribution of term (10) is the sum multiplied by 8, namely 2. This combined with 11/3 from term (7), 22/3 from term (8) and 4/3 from term (9) gives $E(N^2) = 43/3$ in the circular case when $n = 4$. In the non-circular case, the contributions are 11/6 for term (7), 1/3 for term (8), 1/3 for term (9), and there is no contribution from term (10), so that $E(N^2) = 4$ in the non-circular case when $n = 4$.

To evaluate term (10) for $n \geq 5$, one must first determine the contribution of those terms with i, j, k and ℓ next to each other. This may be done by evaluating those terms with i to the left of both j and k. There are six such terms, namely $ijk\ell$, $ij\ell k$, $ikj\ell$, $i\ell jk$, $ik\ell j$, and $i\ell kj$. This leads in order to

$$\begin{aligned}
P\left(R_1 < R_2, R_2 > R_3, R_3 < R_4, R_4 > R_5\right) &= 16/120, \\
P\left(R_1 < R_2, R_2 > R_3, R_4 < R_3, R_5 > R_4\right) &= 11/120, \\
P\left(R_1 < R_3, R_2 > R_4, R_2 < R_4, R_3 > R_5\right) &= 0, \\
P\left(R_1 < R_3, R_2 > R_4, R_4 < R_2, R_5 > R_3\right) &= 10/120, \\
P\left(R_1 < R_4, R_2 > R_5, R_2 < R_3, R_3 > R_4\right) &= 6/120, \\
P\left(R_1 < R_4, R_2 > R_5, R_3 < R_2, R_4 > R_3\right) &= 16/120,
\end{aligned}$$

whose sum is 59/120. This contribution must then be multiplied by $4n$ in the circular case, and by $4(n-4)$ in the non-circular case. In both cases, the factor 4 coming from the fact that the value is the same if one simultaneously interchanges

i with k and j with ℓ, and also if one interchanges i with j and k with ℓ. This yields $59n/30$ in the circular case and $59(n-4)/30$ in the non-circular case. For $n = 5$, this contribution — call it piece I — is $59/6$ in the circular case, which combined with $35/6$ from term (7), 17 from term (8) and $11/3$ from term (9), gives $E(N^2) = 109/3$ as the final answer when $n = 5$ in the circular case. Similarly, $E(N^2) = 206/15$ when $n = 5$ in the non-circular case.

When $n \geq 6$, there are further contributions. One of them comes from those terms with exactly three of i, j, k and ℓ in neighbouring positions. This may be done by evaluating those terms with i, j and k in positions 1, 2 and 3 and ℓ in position 5, and by multiplying the result by $4n(n-5)$ in the circular case, or by $4(n-4)(n-5)$ in the non-circular case. The evaluation of the six terms ijk, ikj, jik, kij, jki and kji yields in order to

$$\begin{aligned}
P\left(R_1 < R_2, R_2 > R_3, R_3 < R_5, R_4 > R_6\right) &= 5/48,\\
P\left(R_1 < R_3, R_2 > R_4, R_2 < R_5, R_3 > R_6\right) &= 1/18,\\
P\left(R_2 < R_1, R_3 > R_2, R_3 < R_5, R_4 > R_6\right) &= 1/16,\\
P\left(R_2 < R_3, R_3 > R_4, R_1 < R_5, R_2 > R_6\right) &= 1/16,\\
P\left(R_3 < R_1, R_4 > R_2, R_2 < R_5, R_3 > R_6\right) &= 1/18,\\
P\left(R_3 < R_2, R_4 > R_3, R_1 < R_5, R_2 > R_6\right) &= 5/48.
\end{aligned}$$

The sum is $4/9$, and hence piece II equals $16n(n-5)/9$ in the circular case, and $16(n-4)(n-5)/9$ in the non-circular case.

Another contribution which arises when $n \geq 6$ comes from the terms with two pairs of i, j, k and ℓ in neighbouring positions separated by at least one space. It requires the evaluation of the six terms with i in position 1 and j, k and ℓ in positions 2, 4 and 6, whose sum must then be multiplied by $2n(n-5)$ in the circular case, and by $2(n-4)(n-5)$ in the non-circular case. Using the order ij-$k\ell$, ij-ℓk, ik-$j\ell$, ik-ℓj, $i\ell$-kj and $i\ell$-jk, one gets

$$\begin{aligned}
P\left(R_1 < R_2, R_2 > R_3, R_4 < R_5, R_5 > R_6\right) &= 1/9,\\
P\left(R_1 < R_2, R_2 > R_3, R_5 < R_4, R_6 > R_5\right) &= 1/9,\\
P\left(R_1 < R_4, R_2 > R_5, R_2 < R_5, R_3 > R_6\right) &= 0,\\
P\left(R_1 < R_5, R_2 > R_6, R_2 < R_4, R_3 > R_5\right) &= 1/36,\\
P\left(R_1 < R_5, R_2 > R_6, R_4 < R_2, R_5 > R_3\right) &= 1/9,\\
P\left(R_1 < R_4, R_2 > R_5, R_5 < R_2, R_6 > R_3\right) &= 1/8.
\end{aligned}$$

Accordingly, the contribution of piece III is $35n(n-5)/36$ in the circular case and $35(n-4)(n-5)/36$ in the non-circular case.

Together, pieces I, II and III total $59n/30 + 11n(n-5)/4$ in the circular case and $59(n-4)/30 + 11(n-4)(n-5)/4$ in the non-circular case. For $n = 6$, this is $59/5 + 33/2$ in the circular case, which combined with $17/2$ from term (7), $29 + 4/5 + 4/9$ from term (8) and $6 + 4/5 + 2/3$ from term (9), gives $E(N^2) = 76 + 11/15$ as the final answer in the circular case when $n = 6$. In the non-circular case, the answer works out to $E(N^2) = 526/15$.

In situations where $n \geq 7$, additional contributions must still be accounted for. One of them, say piece IV, corresponds to those terms with exactly two of i, j, k and ℓ in neighbouring positions. In the circular case, this amounts to

$$
\begin{aligned}
4n(n-5)(n-6)\Big\{ & P\left(R_1 < R_2, R_2 > R_3, R_4 < R_6, R_5 > R_7\right) \\
+ & \ P\left(R_1 < R_6, R_2 > R_7, R_2 < R_4, R_3 > R_5\right) \\
+ & \ P\left(R_1 < R_6, R_2 > R_7, R_4 < R_2, R_5 > R_3\right)\Big\} \\
= & \ 4n(n-5)(n-6)\left(\frac{1}{12}+\frac{1}{24}+\frac{1}{12}\right) = \frac{5}{6}n(n-5)(n-6),
\end{aligned}
$$

while in the non-circular case one gets $5(n-4)(n-5)(n-6)/6$ for piece IV.

The final contribution, piece V, comes from terms with no two of i, j, k or ℓ are next to each other. In the circular case, there are $n(n-5)(n-6)(n-7)$ such terms, each having probability $P(R_1 < R_3, R_2 > R_4, R_5 < R_7, R_6 > R_7) = 1/16$. In the non-circular case, however, the contribution of piece V is reduced to $(n-4)(n-5)(n-6)(n-7)/16$.

When $n \geq 7$, therefore, the total contribution of term (10) is

$$
\frac{59}{30}n + \frac{11}{4}n(n-5) + \frac{5}{6}n(n-5)(n-6) + \frac{1}{16}n(n-5)(n-6)(n-7) \tag{13}
$$

in the circular case, with the common factor of n being replaced by $(n-4)$ in the non-circular case.

Finally, $E(N^2)$ is then the sum of (5), (11), (12) and (13) in the circular case, which reduces to

$$
E(N^2) = \frac{1}{16}n^4 - \frac{1}{72}n^3 - \frac{1}{80}n^2 - \frac{49}{360}n,
$$

whence

$$
\mathrm{Var}(N) = \frac{1}{36}n^3 - \frac{7}{360}n^2 - \frac{49}{360}n.
$$

Similar expressions are available in the non-circular case, as given in the statement of Lemma 1.

4. PROOF OF PROPOSITION 2 ON THE ASYMPTOTIC DISTRIBUTION OF τ_n

Proposition 2 will be established if one can show that the limiting distributions, as $n \to \infty$, of the standardized version of the circular and non-circular versions of N are Gaussian. For this, one needs only check the conditions of a theorem of Sen (1972) on the asymptotic normality of U-statistics. In the non-circular case, the key observation is that since $R_i < R_j$ if and only if $X_i < X_j$, the expression (3) for N based on a series of length $n+1$ may be written as

$$
N = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{ I(X_i < X_j, X_{i+1} > X_{j+1}) + I(X_i > X_j, X_{i+1} < X_{j+1}) \right\}
$$

with X_{n+1} taken as independent of $X_1, \ldots, X_n$, rather than equal to X_1. Since the null distribution of N does not depend on the common continuous distribution of the mutually independent X_i's, the latter may be assumed to arise from the uniform density on the interval $(-1/2, 1/2)$.

Now let $\mathbf{Y}_i = (X_i, X_{i+1})$ for $1 \leq i \leq n$ and define U_n as

$$
U_n = N \Big/ \binom{n}{2} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} g(\mathbf{Y}_i, \mathbf{Y}_j) \Big/ \binom{n}{2},
$$

where g is the symmetric function

$$g(\mathbf{Y}_i, \mathbf{Y}_j) = I(X_i < X_j, X_{i+1} > X_{j+1}) + I(X_i > X_j, X_{i+1} < X_{j+1}).$$

Then U_n is a U-statistic for the sequence $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ which is one-dependent and hence *-mixing. Since g is an indicator random variable, its moments of all orders exist and one may conclude from Theorem 1 of Sen (1972) that

$$\sqrt{n}(U_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\sigma^2),$$

where

$$\theta = E\{g(\mathbf{Y}, \mathbf{Y}')\}$$

with $\mathbf{Y}$ and $\mathbf{Y}'$ two independent copies of $\mathbf{Y}_1$, and

$$\sigma^2 = \text{Var}\{g_1(\mathbf{Y}_1)\} + 2\text{Cov}\{g_1(\mathbf{Y}_1), g_1(\mathbf{Y}_2)\},$$

where, writing $\mathbf{y}$ as (x_1, x_2),

$$\begin{aligned} g_1(\mathbf{y}) &= E\{g(\mathbf{y}, \mathbf{Y}_1)\} = P(x_1 < X_1, x_2 > X_2) + P(x_1 > X_1, x_2 < X_2) \\ &= 1/2 - 2x_1x_2. \end{aligned}$$

Simple calculations yield

$$\theta = E\{g_1(\mathbf{Y})\} = E(1/2 - 2X_1X_2) = 1/2,$$

$$\text{Var}\{g_1(\mathbf{Y})\} = \text{Var}(1/2 - 2X_1X_2) = 4E\left(X_1^2X_2^2\right) = 1/36$$

and

$$\begin{aligned} \text{Cov}\{g_1(\mathbf{Y}_1), g_1(\mathbf{Y}_2)\} &= \text{Cov}(1/2 - 2X_1X_2, 1/2 - 2X_2X_3) \\ &= 4\text{Cov}(X_1X_2, X_2X_3) = 0. \end{aligned}$$

As a result,

$$\sqrt{n}(U_n - 1/2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/9)$$

which implies that

$$\sqrt{n}\left\{\frac{N}{n(n-1)} - 1/4\right\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/36)$$

or

$$\sqrt{n}\,\tau_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4/9).$$

The asymptotic normality of the circular version of N can be obtained from this. For, if $\tilde{U}_n$ denotes the circular version of U_n, then $\tilde{U}_n$ is actually obtained from U_n by running the summation over i from one through n, and substituting $\tilde{\mathbf{Y}}_n = (X_n, X_1)$ for $\mathbf{Y}_n$. Then

$$|U_n - \tilde{U}_n| \leq \sum_{i=1}^{n-1} |g(\mathbf{Y}_j, \mathbf{Y}_n) - g(\mathbf{Y}_j, \tilde{\mathbf{Y}}_n)| \Big/ \binom{n}{2} \leq 2/n.$$

Therefore, $\sqrt{n}(U_n - 1/2)$ and $\sqrt{n}(\tilde{U}_n - 1/2)$ are asymptotically equivalent and so have the same limiting distribution.

5. DISTRIBUTION OF τ_n FOR SMALL n AND QUALITY OF THE ASYMPTOTIC APPROXIMATION

It is of interest to know when a series is sufficiently long that the asymptotic null distribution of τ_n or N may be used in testing against first-order serial dependence. To investigate this issue, the exact distribution of the number N of discordances was computed for series of length n ranging from 3 to 11 under the assumption of mutual independence between the observations. The results are given in Tables 1 and 2 for the non-circular and circular versions of N, respectively.

Table 1. Table of distribution of the number N of discordances for non-circular autocorrelation. Tabled is $P_n(N \leq x)$. The last column provides a normal approximation to $P_{10}(N \leq x)$ using the mean and variance given in Proposition 1, with continuity correction.

$x \backslash n$	3	4	5	6	7	8	9	10	approx.
0	0.333	0.083	0.017	0.003	0.000	0.000	0.000	0.000	0.000
1	1.000	0.250	0.050	0.008	0.001	0.000	0.000	0.000	0.000
2		0.833	0.167	0.031	0.005	0.001	0.000	0.000	0.000
3		1.000	0.193	0.086	0.013	0.002	0.000	0.000	0.000
4			0.767	0.275	0.040	0.006	0.001	0.000	0.000
5			0.967	0.447	0.116	0.015	0.002	0.000	0.001
6			1.000	0.697	0.212	0.043	0.006	0.001	0.001
7				0.831	0.375	0.083	0.796	0.002	0.003
8				0.961	0.512	0.162	0.903	0.004	0.006
9				0.994	0.702	0.244	0.963	0.009	0.011
10				1.000	0.813	0.380	0.096	0.018	0.020
11					0.917	0.490	0.161	0.032	0.034
12					0.966	0.641	0.229	0.057	0.056
13					0.994	0.742	0.332	0.087	0.087
14					0.999	0.848	0.424	0.137	0.130
15					1.000	0.907	0.543	0.188	0.186
16						0.960	0.636	0.265	0.255
17						0.982	0.742	0.336	0.335
18						0.996	0.814	0.432	0.423
19						0.999	0.885	0.511	0.515
20						1.000	0.927	0.610	0.607
21						1.000	0.963	0.684	0.693
22							0.981	0.768	0.769
23							0.993	0.826	0.834
24							0.997	0.884	0.885
25							1.000	0.920	0.924
26							1.000	0.953	0.952
27							1.000	0.971	0.971
28							1.000	0.986	0.984
29								0.993	0.991
30								0.997	0.995
31								0.999	0.998
32								1.000	0.999
33								1.000	1.000
34								1.000	1.000
35								1.000	1.000
36								1.000	1.000

Both tables include a column which shows that for $n > 10$, the probabilities derived from the asymptotic distribution with continuity correction are sufficiently precise for practical purposes. Reliance on Tables 1 and 2 is recommended for series of length 10 or less, however.

Note that no information is lost in Table 1 by reporting $P_n(N \leq x + n - 1)$ for even values of x only. This is because it may be verified that in the circular case,

(i) N assumes only even values if n is odd, and vice versa;

(ii) $n-1 \leq N \leq [n^2/2] - n + 1$, where $[x]$ denotes the integer part of x.

It may actually be seen that the lower bound in (ii) is attained when the X_i's are monotone increasing (i.e., $R_i = i$ for $1 \leq i \leq n$), while the upper bound obtains when $R_1 = 1$, $R_2 = n$, $R_3 = 2$, $R_4 = n-1$, …, $R_n = [n/2] + 1$. Compare with Hallin *et al.* (1992).

TABLE 2. Table of distribution of the number N of discordances for circular autocorrelation. Tabled is $P_n(N \leq x + n - 1)$. The last column provides a normal approximation to $P_{11}(N \leq x+n-1)$ using the mean and variance given in Proposition 1, with continuity correction.

$x\backslash n$	3	4	5	6	7	8	9	10	11	approx.
0	1.000	0.667	0.333	0.133	0.044	0.012	0.003	0.001	0.000	0.001
2		1.000	0.750	0.400	0.167	0.057	0.017	0.004	0.001	0.002
4			1.000	0.767	0.411	0.169	0.057	0.016	0.004	0.006
6				0.950	0.692	0.360	0.144	0.047	0.013	0.016
8				1.000	0.878	0.579	0.283	0.108	0.034	0.036
10					0.975	0.775	0.457	0.206	0.074	0.074
12					1.000	0.907	0.637	0.336	0.139	0.136
14						0.976	0.796	0.489	0.231	0.226
16						0.996	0.903	0.644	0.348	0.343
18						1.000	0.963	0.777	0.480	0.477
20							0.990	0.877	0.614	0.614
22							0.999	0.942	0.735	0.738
24							1.000	0.977	0.834	0.838
26								0.993	0.906	0.909
28								0.998	0.953	0.953
30								1.000	0.980	0.979
32								1.000	0.993	0.991
34									0.998	0.997
36									1.000	0.999
38									1.000	1.000
40									1.000	1.000

6. SIMULATION STUDY OF τ_n'S PERFORMANCE IN SMALL SAMPLES

Monte Carlo experiments comparing the performance of several parametric and nonparametric tests of first-order serial dependence have already been reported by Hallin and Mélard (1988). The same protocol was used here to compare, at the $\alpha = 5\,\%$ nominal level, the power of the τ_n-based one-sided test of independence to that of

a) four alternative rank-based procedures, namely the nonrandomized van der Waerden, Wilcoxon, Laplace and Spearman-Wald-Wolfowitz (or ρ_n) tests;

b) three versions of the traditional parametric test based on the classical first-order autocorrelation coefficient, namely those of Moran (1948), Ljung and Box (1978), and Dufour and Roy (1985).

The readers may refer to the paper by Hallin and Mélard for a precise description of these procedures.

In total, 5,000 pseudo-random, white noise series of length $n = 20$, 50, and 100 were generated from the normal, the logistic, the Laplace (or double exponential), and the Cauchy distributions. Using these $5,000 \times 4 \times 3 = 60,000$ series of innovations ϵ_i, a corresponding number of AR(1) series were constructed by setting

$$X_1 = (1-\theta^2)^{-1/2}\epsilon_1, \quad X_i = \theta X_{i-1} + \epsilon_i, \quad i = 2, \ldots, n.$$

As in Hallin and Mélard (1988), the powers of the tests was compared under alternative hypotheses of the form $\theta = 1/2^{j}$, with $j = 2, \ldots, 5$. The results are summarized in Tables 3–5, whose last column gives the observed level of the tests under the null hypothesis $\theta = 0$. Except for the power calculations involving Kendall's statistic or the Cauchy density, which are new, the results closely match the figures already reported by Dufour and Roy (1985) and Hallin and Mélard (1988). As in the latter study, the standard error is no larger than 0.7% throughout.

A glaring observation is that while all rank-based tests hold their nominal level quite well under the various distributional scenarios, such is not the case for the parametric tests, even under normality. This difficulty, which persists for all distributions and sample sizes in the case of the Ljung-Box test, reiterates the already well documented reliability and robustness of rank-based tests, thereby providing strong motivation for favouring them over classical competitors.

Keeping in mind that the van der Waerden, Wilcoxon and Laplace statistics are locally and asymptotically optimal for the normal, logistic and Laplace densities, the omnibus test based on Kendall's statistic is seen to be close to best in most circumstances. Furthermore, it dominates systematically the Spearman-Wald-Wolfowitz test, often by a wide margin. This is more surprising, in view of the local asymptotic equivalence between these two statistics, as established in the following section. This phenomenon is due, most probably, to the non-local nature of the alternatives considered.

7. ASYMPTOTIC POWERS AND RELATIVE EFFICIENCIES

The following result is an extension to the serial case of a classical result due to Daniels (1944) on the asymptotic equivalence of τ_n and ρ_n.

Proposition 3. *Under the null hypothesis of randomness, the difference between $3\tau_n/2$ and ρ_n is $o_P(n^{-1/2})$.*

In view of Proposition 2 and earlier work (cf., e.g., Hallin *et al.* 1986), the marginal asymptotic distribution of the serial versions of Kendall's tau and Spearman's rho are actually known to be normal with the appropriate parameters. Since it was seen in Section 4 that (in the non-circular case) $\tau_n = 1 - 2U_n$ is a linear function of a U-statistic of order 2, Proposition 3 will follow from the Central Limit Theorem for *-mixing sequences and an application of the Cramér-Wold device if one can show that the non-circular version of ρ_n is asymptotically equivalent to the U-statistic of degree 3 defined by

$$W_n = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=j+1}^{n} h(\mathbf{Y}_i, \mathbf{Y}_j, \mathbf{Y}_k) \Big/ \binom{n}{3}$$

in terms of

$$h(\mathbf{Y}_i, \mathbf{Y}_j, \mathbf{Y}_k) = (a_{ijk} + a_{ikj} + a_{jik} + a_{jki} + a_{kij} + a_{kji})\,/6$$

and

$$a_{ijk} = I(X_i > X_j, X_{i+1} > X_{k+1}), \quad 1 \le i, j, k \le n. \tag{14}$$

Indeed, the asymptotic correlation between $\sqrt{n}\tau_n$ and $\sqrt{n}\rho_n$ will then be equal to $-\mathrm{Corr}\{g_1(\mathbf{Y}), h_1(\mathbf{Y})\}$, where $h_1(\mathbf{y}) = E\{h(\mathbf{y}, \mathbf{Y}_2, \mathbf{Y}_3)\}$. Due to distribution-freeness, the latter may be computed under the assumption that the components of all $\mathbf{Y}_i = (X_i, X_{i+1})$ are mutually independent observations from a uniform distribution on the interval $(-1/2, 1/2)$. Writing $\mathbf{y} = (x_1, x_2)$ as before, one finds

$$h_1(\mathbf{y}) = \frac{1}{3}\{P(x_1 > X_1, x_2 > X_4) + P(X_1 > x_1, X_2 > X_4) + P(X_3 > X_1, X_4 > x_2)\}$$

i.e., $h_1(\mathbf{y}) = x_1x_2/3 + 1/4$. As a result, $E\{h_1(\mathbf{Y})\} = 1/4$ and

$$\mathrm{Var}\{h_1(\mathbf{Y})\} = E\left(X_1^2X_2^2\right)/9 = 1/1296.$$

Since it is known that $E\{g_1(\mathbf{Y})\} = 1/2$ and $\mathrm{Var}\{g_1(\mathbf{Y})\} = 1/36$ from earlier calculations, one finds also

$$\mathrm{Cov}\{g_1(\mathbf{Y}), h_1(\mathbf{Y})\} = E\left[\{g_1(\mathbf{Y}) - 1/2\}h(\mathbf{Y})\right] = -2E\left(X_1^2X_2^2\right)/3 = -1/226,$$

and hence $\mathrm{Corr}\{g_1(\mathbf{Y}), h_1(\mathbf{Y})\} = -1$. Thus, the joint distribution of $\sqrt{n}\tau_n$ and $\sqrt{n}\rho_n$ is asymptotically normal with degenerate covariance matrix, so that the difference $\sqrt{n}(3\tau_n/2 - \rho_n)$ between their asymptotic standardized versions is $o_P(1)$.

Table 3. Percentage of rejection of the null hypothesis of randomness under first-order autoregressive dependence $X_i - \theta X_{i-1} = \epsilon_i$ for one-sided tests at the 5% level applied to series of length $n = 20$ when the innovations ϵ_i form a random sample from the normal, the logistic, the Laplace or the Cauchy distribution.

Density	Statistic	$\theta = 1/4$	$\theta = 1/8$	$\theta = 1/16$	$\theta = 1/32$	$\theta = 0$
Normal						
	van der Waerden	23.9	11.7	7.8	6.0	4.7
	Wilcoxon	23.7	11.5	7.9	6.0	4.8
	Laplace	18.7	9.9	7.2	6.3	5.2
	ρ_n	19.1	10.4	6.9	5.6	4.6
	τ_n	23.5	12.0	7.9	6.2	5.1
	Ljung-Box	18.3	7.8	4.7	3.7	2.8
	Moran	26.1	12.8	8.2	6.3	4.9
	Dufour-Roy	23.3	11.1	6.7	5.2	4.1
Logistic						
	van der Waerden	25.0	12.0	8.0	6.1	4.7
	Wilcoxon	25.3	12.1	8.0	6.1	4.8
	Laplace	20.9	10.7	7.5	6.4	5.2
	ρ_n	19.9	10.7	6.9	5.6	4.6
	τ_n	24.5	12.5	8.0	6.2	5.1
	Ljung-Box	18.3	7.5	4.4	3.5	2.6
	Moran	25.8	12.5	7.6	6.0	4.6
	Dufour-Roy	22.9	10.9	6.5	4.9	3.8
Laplace						
	van der Waerden	27.2	13.2	8.5	6.3	4.7
	Wilcoxon	28.6	14.0	8.8	6.4	4.8
	Laplace	26.2	13.4	8.6	6.7	5.2
	ρ_n	21.5	11.6	7.3	5.8	4.6
	τ_n	27.7	13.8	8.7	6.5	5.1
	Ljung-Box	17.5	7.0	4.1	3.0	2.4
	Moran	25.4	11.7	7.2	5.6	4.4
	Dufour-Roy	22.6	9.9	5.9	4.6	3.5
Cauchy						
	van der Waerden	48.7	25.7	15.3	9.7	4.7
	Wilcoxon	53.3	28.3	16.5	10.7	4.7
	Laplace	47.6	26.2	15.9	10.2	5.2
	ρ_n	38.4	19.5	12.2	8.5	4.6
	τ_n	51.9	26.6	15.7	10.2	5.1
	Ljung-Box	10.1	3.4	2.1	1.6	1.4
	Moran	15.9	5.7	3.5	2.8	2.1
	Dufour-Roy	13.3	4.7	2.9	2.3	1.8

To show the relation between W_n and ρ_n, note that the latter is a normalized version of the statistic $\sum_{i=1}^{n} R_i R_{i+1}$ with $R_i = 1 + \sum_{j=1}^{n+1} I(X_i > X_j)$. Up to a change of location and scale, ρ_n is thus equivalent to the V-statistic

$$V_n = \frac{1}{n^3} \sum_{i=1}^{n} R_i R_{i+1} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk}$$

defined in terms of the indicators a_{ijk} defined in (14). But W_n is the U-statistic corresponding to V_n and standard arguments show that, as $n \to \infty$, $\sqrt{n}\,(W_n - V_n) \to 0$ in probability, thereby completing the proof of Proposition 3.

An immediate corollary of this proposition is that τ_n and ρ_n have the same relative asymptotic efficiency with respect to sample autocorrelations: $9/\pi^2 \approx 0.912$ for the normal, 1 for the logistic, $(9/8)^2 \approx 1.266$ for the Laplace, and so on.

Table 4. Percentage of rejection of the null hypothesis of randomness under first-order autoregressive dependence $X_i - \theta X_{i-1} = \epsilon_i$ for one-sided tests at the 5% level applied to series of length $n = 50$ when the innovations ϵ_i form a random sample from the normal, the logistic, the Laplace or the Cauchy distribution.

Density	Statistic	$\theta = 1/4$	$\theta = 1/8$	$\theta = 1/16$	$\theta = 1/32$	$\theta = 0$
Normal						
	van der Waerden	50.4	20.8	10.9	7.2	4.7
	Wilcoxon	49.7	20.0	10.7	7.2	4.7
	Laplace	37.0	16.3	9.2	6.4	4.6
	ρ_n	44.9	18.5	9.9	7.1	4.9
	τ_n	48.5	19.7	10.4	6.9	4.6
	Ljung-Box	46.1	17.4	8.4	5.3	3.6
	Moran	51.7	21.2	11.4	7.4	4.9
	Dufour-Roy	50.5	20.2	10.8	6.9	4.6
Logistic						
	van der Waerden	52.4	21.3	11.1	7.1	4.7
	Wilcoxon	53.7	21.8	11.1	7.3	4.7
	Laplace	42.5	18.2	10.0	6.8	4.6
	ρ_n	47.3	19.3	10.1	7.3	4.9
	τ_n	51.3	20.7	10.8	7.0	4.7
	Ljung-Box	46.4	17.1	8.2	5.1	3.5
	Moran	52.0	21.0	11.0	7.0	4.6
	Dufour-Roy	50.6	20.1	10.4	6.5	4.4
Laplace						
	van der Waerden	57.3	23.4	12.2	7.4	4.7
	Wilcoxon	61.0	25.3	12.4	7.8	4.7
	Laplace	55.9	25.7	13.0	8.1	4.6
	ρ_n	53.0	21.8	11.2	7.6	4.9
	τ_n	58.3	23.4	11.6	7.5	4.6
	Ljung-Box	46.2	16.5	7.7	5.0	3.2
	Moran	52.4	20.9	10.2	6.9	4.6
	Dufour-Roy	51.2	19.9	9.6	6.5	4.2
Cauchy						
	van der Waerden	88.1	54.4	28.2	15.5	4.7
	Wilcoxon	91.5	61.4	32.7	16.9	4.7
	Laplace	88.1	56.7	30.7	16.9	4.6
	ρ_n	83.7	48.2	23.6	13.1	4.9
	τ_n	88.8	53.9	26.5	13.8	4.6
	Ljung-Box	44.6	6.8	3.4	2.5	1.7
	Moran	62.9	8.7	4.3	3.2	2.3
	Dufour-Roy	60.0	8.2	4.1	3.1	2.1

TABLE 5. Percentage of rejection of the null hypothesis of randomness under first-order autoregressive dependence $X_i - \theta X_{i-1} = \epsilon_i$ for one-sided tests at the 5% level applied to series of length $n = 100$ when the innovations ϵ_i form a random sample from the normal, the logistic, the Laplace or the Cauchy distribution.

Density	Statistic	$\theta = 1/4$	$\theta = 1/8$	$\theta = 1/16$	$\theta = 1/32$	$\theta = 0$
Normal						
	van der Waerden	79.2	33.9	15.0	9.4	5.2
	Wilcoxon	77.0	33.3	15.2	9.2	5.4
	Laplace	61.5	25.8	12.6	8.5	5.6
	ρ_n	73.3	30.5	15.0	8.9	5.1
	τ_n	75.7	32.2	15.4	9.3	5.3
	Ljung-Box	76.7	31.0	13.1	7.8	4.4
	Moran	79.7	34.8	15.1	9.2	5.5
	Dufour-Roy	79.3	34.2	14.7	8.9	5.3
Logistic						
	van der Waerden	80.8	34.8	15.4	9.5	5.2
	Wilcoxon	81.6	35.8	16.1	9.5	5.4
	Laplace	69.3	29.4	13.8	9.1	5.6
	ρ_n	76.5	32.1	15.4	9.0	5.1
	τ_n	79.1	34.0	15.9	9.6	5.3
	Ljung-Box	77.1	30.8	12.8	7.7	4.5
	Moran	80.3	34.8	15.2	8.9	5.5
	Dufour-Roy	79.8	34.0	14.8	8.7	5.2
Laplace						
	van der Waerden	84.5	38.4	16.6	10.1	5.2
	Wilcoxon	88.0	42.4	18.5	10.3	5.4
	Laplace	84.3	42.6	19.9	11.1	5.6
	ρ_n	82.9	36.4	17.3	9.8	5.1
	τ_n	85.0	38.8	17.5	10.3	5.3
	Ljung-Box	77.4	29.7	12.4	7.2	4.1
	Moran	80.6	34.3	14.7	8.6	5.0
	Dufour-Roy	80.0	33.6	14.2	8.3	4.8
Cauchy						
	van der Waerden	99.1	82.2	47.7	24.7	5.2
	Wilcoxon	99.5	87.4	55.6	29.2	5.4
	Laplace	99.3	85.6	53.8	28.6	5.6
	ρ_n	98.6	76.9	41.0	21.1	5.1
	τ_n	99.1	79.9	43.7	22.3	5.3
	Ljung-Box	90.3	13.3	4.5	3.1	2.2
	Moran	91.8	16.6	5.4	3.4	2.5
	Dufour-Roy	91.6	15.7	5.3	3.4	2.4

8. HIGHER-ORDER EXTENSIONS

While the serial version of Kendall's tau considered in (2) provides an adequate tool for testing against first-order serial dependence, it is unfit in situations where a dependence of higher order is suspected. To test against dependence at lag $k = 2, \ldots, n$, an obvious extension of the circular statistic would be defined as in (2), but with

$$N_k = \sum_{i=1}^{n} \sum_{j=1}^{n} I\left(R_i < R_j, R_{i+k} > R_{j+k}\right).$$

In the non-circular case, one would have

$$\tau_{k,n} = 1 - 2N \Big/ \binom{n-k}{2} = 1 - \frac{4N_k}{(n-k)(n-k-1)},$$

but with n replaced by $n - k$ in the above formula for N_k.

As might be expected, the distribution of the resulting statistics depend on k. It is easy to check, for instance, that the expected value of the circular version of N_k is still given by

$$E(N_k) = \frac{n(3n-1)}{12},$$

except when $k = n/2$, in which case the appropriate formula is simply $n^2/4$. In the non-circular case, one finds

$$E(N_k) = \frac{(3n-3k-1)(n-k)}{12} - \frac{k}{6}$$

for $1 \leq k < n/2$ and

$$E(N_k) = \frac{(n-k)(n-k-1)}{4}$$

for $n/2 \leq k < n$.

An explicit value of the variance of $\tau_{k,n}$ for $k > 1$ could be obtained along similar lines; the derivation, however, would be extremely tedious. A straightforward modification of the argument considered in Section 4 leads to the same result that $3\sqrt{n}\tau_{k,n}/2$ is asymptotically standard normal for all fixed value of k. More precisely, it can be seen that any K-tuple of the form $3\sqrt{n}\,(\tau_{1,n} \ldots, \tau_{K,n})/2$ is asymptotically multinormal, with mean zero and unit covariance matrix, under the hypothesis of randomness, as $n \to \infty$. Such K-tuples thus have the same asymptotic behaviour, and admit the same intuitive interpretation as the traditional or the rank-based correlograms (van der Waerden, Wilcoxon, etc).

ACKNOWLEDGEMENTS

This project was completed while the first two authors were visiting the Institut de statistique et de recherche opérationnelle (ISRO) at the Université libre de Bruxelles, which graciously provided the research facilities. Financial assistance in support of this work was provided by the Francqui Foundation, the Natural Sciences and Engineering Research Council of Canada, le Fonds pour la formation et l'aide à la recherche du Gouvernement du Québec, and a Research in Brussels contract from the Région de Bruxelles–Capitale / Brussels Hoofdstedelijk Gewest.

REFERENCES

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, 33, 129–135.

Dufour, J.-M., and Roy, R. (1985). Some robust exact results on sample autocorrelations and tests for randomness. *J. Econom.*, 29, 257–273.

Hallin, M., and Mélard, G. (1988). Rank-based tests for randomness against first-order serial dependence. *J. Amer. Statist. Assoc.*, 83, 1117–1128.

Hallin, M., Mélard, G., and Milhaud, X. (1992). Permutational extreme values of autocorrelation coefficients and a Pitman test against serial dependence. *Ann. Statist.*, 20, 523-534.

Hallin, M., and Puri, M. L. (1988). Optimal rank-based procedures for time-series analysis: testing an ARMA model against another ARMA model. *Ann. Statist.*, 16, 402-432.

Hallin, M., and Puri, M. L. (1992). Rank tests for time series analysis: A survey. *New Directions in Time Series Analysis*, 111–154. (D. R. Brillinger, E. Parzen and M. Rosenblatt, *eds.*), Springer-Verlag, New York.

Hallin, M., and Werker, B. J. M. (1999). Optimal testing for semi-parametric AR models — From Gaussian Lagrange multipliers to autoregressive rank scores and adaptive tests. *Asymptotics, Nonparametrics, and Time Series*, 295-350 (S. Ghosh, *ed.*), Marcel Dekker, New York.

Ljung, G. M., and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297–303.

Moran, P. A. P. (1948). Some theorems on time series II: The significance test of the serial correlation coefficient. *Biometrika*, 35, 255–260.

Sen, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes. *Z. Wahrsch. verw. Geb.*, 25, 71–82.

Wald, A., and Wolfowitz, J. (1943). An exact test for randomness in the nonparametric case based on serial correlation. *Ann. Math. Statist.*, 14, 378-388.

Received 2 June 1999
Revised ???
Accepted ???

Mathematics Department
University of California
405 Hilgard Avenue
Los Angeles, CA 90095
USA

e-mail: tom@math.ucla.edu

Département de mathématiques et de statistique
Université Laval
Sainte-Foy (Québec)
Canada G1K 7P4

e-mail: genest@mat.ulaval.ca

Institut de statistique et de recherche opérationnelle
Université libre de Bruxelles
C. P. 200, Campus de la plaine
1050 Bruxelles
Belgium

e-mail: mhallin@ulb.ac.be