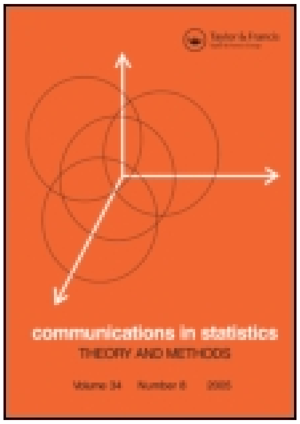


Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

# On Kendall's Autocorrelations

Miroslav Šiman [a]

[a] Institute of Information Theory and Automation of the ASCR, Prague, Czech Republic

Version of record first published: 13 Apr 2012

*Communications in Statistics—Theory and Methods*, 41: 1733–1738, 2012

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610926.2010.549991



# On Kendall's Autocorrelations

MIROSLAV ŠIMAN

Institute of Information Theory and Automation of the ASCR, Prague, Czech Republic

*This brief article extends the theory of sample Kendall's autocorrelations by providing their exact variances at lags higher than one under the null hypothesis of randomness, by introducing and investigating their weighted modifications, and by numerical demonstration of these results and their usefulness.*

**Keywords** Autocorrelation; Kendall's tau; Serial rank coefficient.

**Mathematics Subject Classification** Primary 62M10; Secondary 62G10.



## 1. Introduction

Let $R_1, R_2, \ldots, R_T$ be the ranks associated with a given time series $Y_1, Y_2, \ldots, Y_T$. These ranks are often used for building tests and estimators that do not require any restrictive distributional assumptions. Rank measures of autocorrelation are no exceptions to this rule and usually work well even in case of some distributional uncertainty or data distortion. They are well represented by the Kendall autocorrelation coefficients introduced in Ferguson et al. (2000). Let us recall their definition.

**Definition 1.1.** The Kendall autocorrelations $\hat{r}_K(k)$'s, $k = 1, 2, \ldots,$ are defined by

$$\hat{r}_K(k) = 1 - \frac{4N_{k,T}}{(T-k)(T-k-1)} \quad \text{where } N_{k,T} = \sum_{i=1}^{T-k}\sum_{j=1}^{T-k} \mathrm{I}(R_i < R_j, R_{i+k} > R_{j+k}).$$

All rank statistics have the same distribution whenever the null hypothesis $H_0^E$ considered here holds. It assumes that $Y_t$'s are exchangeable random variables with a continuous distribution, which also covers any white noise formed by independent and identically distributed continuous random variables. Expectations and variances computed with respect to this null hypothesis will be denoted by

Received August 30, 2010; Accepted December 17, 2010
Address correspondence to Miroslav Šiman, Institute of Information Theory and Automation of the ASCR, Pod Vodárenskou věží 4, CZ-182 08, Prague 8, Czech Republic; E-mail: siman@utia.cas.cz

$E_0$ and $var_0$, respectively. They can be used for computing standardized Kendall's autocorrelations $\tilde{r}_K(k)$'s:

$$\tilde{r}_K(k) = \frac{\hat{r}_K(k) - E_0(\hat{r}_K(k))}{\sqrt{var_0(\hat{r}_K(k))}}, \quad k = 1, 2, \ldots$$

The following theorem summarizes some relevant results from Ferguson et al. (2000) regarding $\hat{r}_K(k)$'s and their behavior under $H_0^E$.

**Theorem 1.1.** *If $H_0^E$ holds and $k \in \mathbb{N}$, then*

$$E_0(N_{k,T}) = \frac{(3T - 3k - 1)(T - k)}{12} - \frac{k}{6} \quad \textit{for } 1 \leq k < \frac{T}{2},$$
$$= \frac{(T - k)(T - k - 1)}{4} \quad \textit{for } \frac{T}{2} \leq k < T - 1,$$
$$var_0(N_{1,T}) = \frac{10T^3 - 37T^2 + 27T + 74}{360} \quad \textit{for } T \geq 4,$$
$$\sqrt{T}\hat{r}_K(k) \sim_{asympt.} N\left(0, \frac{4}{9}\right) \quad \textit{for } k \geq 1.$$

*Besides, it also holds asymptotically that $\tilde{r}_K(k)$'s are then jointly standard normal and independent.*

In the simulation study in Ferguson et al. (2000), $\tilde{r}_K(1)$ was shown superior to the ordinary (Spearman) rank autocorrelation at lag one in many cases (and despite their established asymptotic equivalence under $H_0^E$ and local alternatives). This empirical finding could be very likely generalized to any lag if the exact variances of $\hat{r}_K(k)$'s under $H_0^E$ were known.

Next, we not only compute these variances and show them useful but also introduce weighted modifications of the Kendall autocorrelations, investigate their properties and empirically demonstrate their potential to increase the speed and power of the tests based on them.[1]

## 2. Theoretical Results

Unfortunately, Ferguson et al. (2000) does not provide any formulae for $var_0(\hat{r}_K(k))$'s at lags $k > 1$ because of their extremely tedious computation that would require to separately consider all the many cases when some of the eight indices at $R$ in

$$E_0(N_{k,T}^2) = \sum_{i}^{T-k} \sum_{j}^{T-k} \sum_{p}^{T-k} \sum_{q}^{T-k} P(R_i < R_j, R_{i+k} > R_{j+k}, R_p < R_q, R_{p+k} > R_{q+k})$$

mutually coincide. However, such formulae are badly needed for the accurate standardization of $\hat{r}_K(k)$'s and that is why we laboriously derived them in Maple:

[1]All the results are taken from Chapters 5 and 6 of the author's PhD thesis; see Šiman (2006).

**Theorem 2.1.** *If $H_0^E$ holds, $T-1>k>0$, and $T \geq 8$, then*

$$\begin{aligned}\operatorname{var}_0(\hat{r}_K(k)) &= \frac{2[10T^3+(-30k+15)T^2+(30k^2-30k-25)T+(-10k^3+15k^2+25k)]}{45(T-k)^2(T-k-1)^2}, \\ &\quad k+1<T<2k, \\ &= \frac{2[10T^3+(-30k+13)T^2+(30k^2-34k-21)T+(-10k^3+31k^2+17k)]}{45(T-k)^2(T-k-1)^2}, \\ &\quad 2k \leq T < 3k, \\ &= \frac{2[10T^3+(-30k-7)T^2+(30k^2+46k-37)T+(-10k^3-29k^2+65k)]}{45(T-k)^2(T-k-1)^2}, \\ &\quad 3k \leq T < 4k, \\ &= \frac{2[10T^3+(-30k-7)T^2+(30k^2+46k-49)T+(-10k^3-29k^2+113k)]}{45(T-k)^2(T-k-1)^2}, \\ &\quad T \geq 4k.\end{aligned}$$

*Note.* These formulae agree with the results obtained for $k=1$ in Ferguson et al. (2000) and with the partial result obtained for $T \geq 8k+1$ in Šiman (2005).

Next, we introduce weighted Kendall's autocorrelations that can be obtained easily as the most natural generalization of both those serial unweighted and non serial weighted Kendall's coefficients; see Ferguson et al. (2000) and Shieh (1998), respectively. For simplicity, we introduce only their standardized versions.

**Definition 2.1.** The standardized weighted Kendall autocorrelations $\tilde{r}_{K,w}(k)$'s, $k = 1, 2, \ldots$, are defined in the following way:

$$\tilde{r}_{K,w}(k) = -\frac{N_{k,w,T} - \mathrm{E}_0(N_{k,w,T})}{\sqrt{\operatorname{var}_0(N_{k,w,T})}}, \quad \text{where}$$

$$N_{k,w,T} = \sum_{i=1}^{T-k}\sum_{j=1}^{T-k} w(i,j)\mathrm{I}(R_i < R_j, R_{i+k} > R_{j+k})$$

and $w$ is a real function symmetric in its arguments and possibly dependent on $k$.

For simplicity, we will hereinafter focus on the trimmed weighting functions satisfying $w(i,j)=0$ for $|i-j|>m$, $m<T$, that have the obvious advantage of speeding up the computation of the weighted autocorrelation coefficients. In our concluding simulation study, we consider

$$w_1(i,j) = \mathrm{I}(|i-j| \leq m) \quad \text{and} \quad w_2(i,j) = |i-j|\mathrm{I}(|i-j| \leq m)$$

that proved interesting in our Monte Carlo experiments. In these cases and under $H_0^E$, the means and variances of $N_{k,w,T}$'s can still be obtained with the aid of Maple and the asymptotic normality of $\tilde{r}_{K,w}(k)$'s can be established easily:

**Theorem 2.2.** *If the null hypothesis $H_0^E$ holds, $w = w_1$ or $w_2$, $T \geq 5k + m$, $m \geq 3k$, and $k, m \in \mathbb{N}$, then the means and variances of $N_{k,w,T}$'s are as follows*:

$$\mathrm{E}_0(N_{k,w_1,T}) = \left(\frac{1}{2}m + \frac{1}{6}\right)T - \frac{1}{4}m^2 - \frac{1}{2}km - \frac{1}{4}m - \frac{1}{3}k,$$

$$\mathrm{var}_0(N_{k,w_1,T}) = \left(\frac{1}{9}m^2 + \frac{13}{180}m - \frac{1}{15}k - \frac{23}{180}\right)T - \frac{5}{54}m^3 - \frac{43}{360}m^2 - \frac{1}{9}km^2 - \frac{29}{1080}m + \frac{1}{12}km + \frac{31}{180}k^2 + \frac{23}{60}k,$$

$$\mathrm{E}_0(N_{k,w_2,T}) = \left(\frac{1}{4}m^2 + \frac{1}{4}m + \frac{1}{6}k\right)T - \frac{1}{6}m^3 - \frac{1}{4}m^2 - \frac{1}{4}km^2 - \frac{1}{4}km - \frac{1}{12}m - \frac{1}{3}k^2,$$

$$\mathrm{var}_0(N_{k,w_2,T}) = \left(\frac{1}{36}m^4 + \frac{79}{540}m^3 + \frac{59}{360}m^2 - \frac{2}{9}km^2 - \frac{2}{9}km + \frac{49}{1080}m - \frac{19}{90}k^2 + \frac{7}{60}k^3 - \frac{1}{30}k\right)T - \frac{4}{135}m^5 - \frac{1}{36}km^4 - \frac{307}{2160}m^4 - \frac{23}{120}m^3 + \frac{1}{540}km^3 + \frac{16}{45}k^2m^2 - \frac{167}{2160}m^2 + \frac{7}{120}km^2 + \frac{16}{45}k^2m + \frac{31}{1080}km + \frac{1}{540}m - \frac{217}{1080}k^4 + \frac{4}{5}k^3 + \frac{37}{1080}k^2.$$

*Note.* Theorem 2.2 discusses only the situation when $T \geq 5k + m$ and $m \geq 3k$, which is usually the most relevant case in the time series context. Nevertheless, the other possibilities could be treated analogously. Besides, the exact formulae for $\mathrm{E}_0(N_{k,w,T})$ and $\mathrm{var}_0(N_{k,w,T})$ could be derived in the same way even for any other weighting function in a polynomial or other similar form.

**Theorem 2.3.** *Let us assume that $H_0^E$ holds and that $w(\cdot,\cdot)$ is an arbitrary non trivial symmetric weighting function such that $w(i,j) = 0$ for $|i - j| > m$ and*

$$\text{(A)} \qquad \lim_{T\to\infty} \frac{\max_{1\leq i,j\leq T-k} |w(i,j)|}{\sqrt{\mathrm{var}_0(N_{k,w,T})}} = 0,$$

$$\text{(B)} \qquad \frac{\sum_{i=1}^{T-k} \sum_{j=i+1}^{i+m} w^2(i,j)}{\mathrm{var}_0(N_{k,w,T})} = O(1).$$

*Then the individual coefficient $\tilde{r}_{K,w}(k)$ is asymptotically standard normal for any $k \geq 1$.*

*Proof.* The asymptotic distribution of $\tilde{r}_{K,w}(k)$ under $H_0^E$ will be the same as if $Y_t$'s were independent and uniformly distributed on [0, 1]. Then one can rewrite $N_{k,w,T}$ as a sum of $(m+k)$-dependent random variables, employ Orey (1958) and apply its central limit theorem whose assumptions are satisfied thanks to the conditions (A) and (B). □

*Note.* Both the weights $w_1$ and $w_2$ considered here meet the assumptions of Theorem 2.3. Besides, the joint asymptotic zero mean normal distribution of the first $\tilde{r}_{K,w}(k)$'s (possibly together with other such coefficients at the same lags but with different trimmed weighting functions) could be proved analogously.

## 3. Numerical Results

First, we compute $\sqrt{T\mathrm{var}_0(\hat{r}_K(k))}$ for several choices of $T$ and $k$, see Table 1. Apparently, this quantity cannot be accurately approximated by its asymptotic value 2/3 even for $T = 50$ or higher, especially for large values of $k$. This confirms that only the exact variances provided by Theorem 2.1 can make the Kendall autocorrelations at lags $k > 1$ successfully applicable in practice.

Second, we consider an artificial example and use $\tilde{r}_{K,w_1}(1)$ and $\tilde{r}_{K,w_2}(1)$ for detecting conditional heteroscedasticity generated by weakly stationary GARCH(1, 1) processes $\{Y_t\}$, see Berkes et al. (2004), based on the standard normal white noise $\{\varepsilon_t\}$,

$$Y_t = \varepsilon_t \sigma_t, \quad \sigma_t^2 = 1 + aY_{t-1}^2 + b\sigma_{t-1}^2,$$

with parameter vector $(a, b)$ equal to (0.15, 0.8), (0.2, 0.2), (0.3, 0.3), (0.4, 0.4), (0.2, 0.4), (0.2, 0.6), (0.4, 0.2), and (0.6, 0.2). The null hypothesis of our interest corresponds to $(a, b) = (0, 0)$. If we want to test it against volatile GARCH alternatives, we can do that by means of the weighted Kendall autocorrelation at lag one computed from $|Y_t|$'s. So we simulated $N = 10{,}000$ realizations of each GARCH time series with $T = 200$ observations and computed the empirical frequencies of rejection of the null hypothesis by each test at the nominal level $\alpha = 0.10$ for various values of the trimming parameter $m \in \{3, 4, \ldots, 66\}$. The results are displayed in Fig. 1 and illustrate two important facts, namely that the weighted Kendall autocorrelations may lead to virtually the same test power as their standard (non weighted) counterparts even for quite small values of the trimming parameter $m$, and that the uniform trimmed weighting function $w_1$ need not always be optimal in terms of test power, which makes the other weighting functions such as $w_2$ worth considering in some cases as well.

**Table 1**

This table shows $\sqrt{T\mathrm{var}_0(\hat{r}_K(k))}$ computed for the first few lags $k$ and for various time series lengths $T$. This quantity converges to 2/3 for any fixed $k$ and $T \to \infty$

| $T\backslash k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 10 | 0.7547 | 0.8384 | 0.9621 | 1.1008 | 1.2910 | 1.5516 | 2.0184 | 3.1623 |
| 15 | 0.7220 | 0.7612 | 0.8093 | 0.8699 | 0.9480 | 1.0180 | 1.1102 | 1.2280 |
| 20 | 0.7072 | 0.7323 | 0.7609 | 0.7940 | 0.8328 | 0.8786 | 0.9290 | 0.9778 |
| 25 | 0.6987 | 0.7171 | 0.7372 | 0.7596 | 0.7845 | 0.8127 | 0.8445 | 0.8811 |
| 30 | 0.6931 | 0.7076 | 0.7231 | 0.7400 | 0.7582 | 0.7782 | 0.8001 | 0.8243 |
| 35 | 0.6892 | 0.7011 | 0.7138 | 0.7272 | 0.7416 | 0.7570 | 0.7736 | 0.7915 |
| 40 | 0.6863 | 0.6964 | 0.7071 | 0.7183 | 0.7301 | 0.7427 | 0.7560 | 0.7701 |
| 45 | 0.6841 | 0.6929 | 0.7021 | 0.7117 | 0.7217 | 0.7323 | 0.7434 | 0.7551 |
| 50 | 0.6823 | 0.6901 | 0.6982 | 0.7066 | 0.7153 | 0.7244 | 0.7339 | 0.7439 |
| 500 | 0.6682 | 0.6689 | 0.6696 | 0.6702 | 0.6709 | 0.6716 | 0.6723 | 0.6730 |
| 5000 | 0.6668 | 0.6669 | 0.6670 | 0.6670 | 0.6671 | 0.6672 | 0.6672 | 0.6673 |

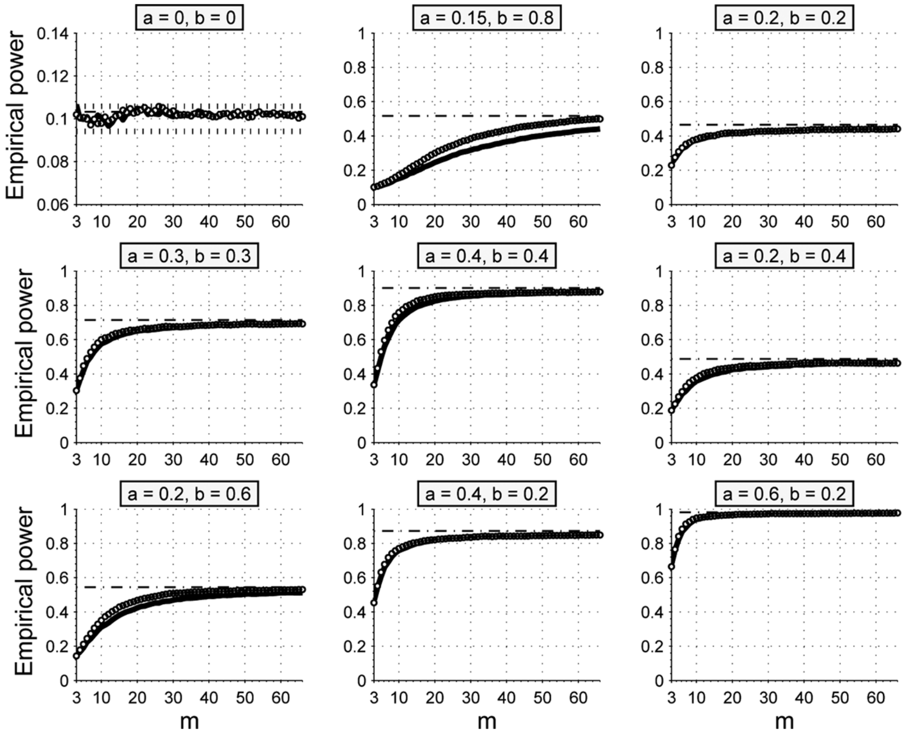


**Figure 1.** Empirical power of the tests based on $\tilde{r}_K(1)$ (-·-·-), $\tilde{r}_{K,w_1}(1)$ (▬▬▬), and $\tilde{r}_{K,w_2}(1)$ (ooooo) with trimming parameter $m$, when applied to absolute values of $T = 200$ consecutive observations from GARCH(1,1) models $Y_t = \varepsilon_t \sigma_t$, $\sigma_t^2 = 1 + \mathrm{a}Y_{t-1}^2 + \mathrm{b}\sigma_{t-1}^2$, with $N(0, 1)$ innovations $\{\varepsilon_t\}$. In the upper left subfigure, ıııııııı stands for the bounds of 95% confidence intervals for the empirical size.

## Acknowledgments

This research work of Miroslav Šiman was supported by Project 1M06047 of the Ministry of Education, Youth and Sports of the Czech Republic.

## References

Berkes, I., Horváth, L., Kokoszka, P. (2004). Probabilistic and statistical properties of GARCH processes. *Fields Instit. Commun.* 44:409–429.

Ferguson, T. S., Genest, C., Hallin, M. (2000). Kendall's tau for serial dependence. *Can. J. Statist.* 28:587–604.

Orey, S. (1958). A central limit theorem for *m*-dependent random variables. *Duke Mathemat. J.* 25:543–546.

Shieh, G. S. (1998). A weighted Kendall's tau statistic. *Statist. Probab. Lett.* 39:17–24.

Šiman, M. (2005). Portmanteau tests based on Kendall's autocorrelation coefficients. In: *Proc. 23rd Int. Conf. Mathemat. Meth. Econ.*, Hradec Králové, Czech Republic.

Šiman, M. (2006). On Portmanteau Tests of Randomness. Doctoral thesis, Faculty of Mathematics and Physics, Charles University in Prague.